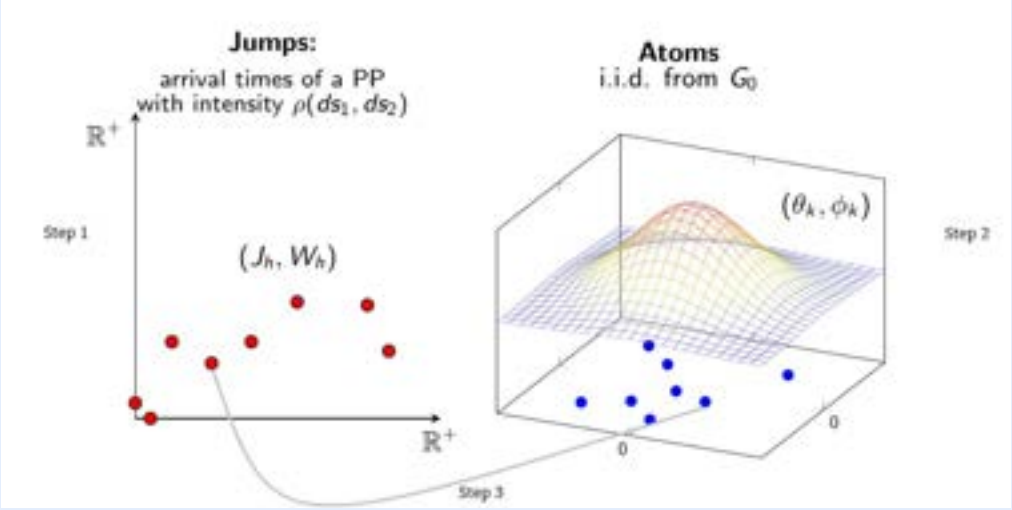
ON DEPENDENT PROCESSES IN BAYESIAN NONPARAMETRICS

THEORY, METHODS, AND APPLICATIONS



PhD Thesis by Beatrice **FRANZOLINI**
Advisors: Antonio **LIJOI** and Igor **PRÜNSTER**

PhD program in Statistics

Year 2022

UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"
PhD SCHOOL

PhD program in Statistics
Cycle: XXXIII
Disciplinary Field: SECS-S/01

On Dependent Processes in Bayesian Nonparametrics: Theory, Methods, and Applications

Advisor: Antonio LIJOI
Co-Advisor: Igor PRÜNSTER

PhD Thesis by
Beatrice FRANZOLINI
ID number: 1811371

Year 2022

This Page Intentionally Left Blank

Acknowledgements

I would like to thank my supervisors Antonio Lijoi and Igor Prünster for their guidance, encouragement, and dedication in mentoring me. I honestly believe I could not have had better supervisors for doing this PhD. My sincere and dutiful thanks also go to Filippo Ascolani and Giovanni Rebaudo, who together with us contributed to the contents of respectively Chapter 3 and Chapter 2. Furthermore, I would like to thank the rest of my PhD colleagues and in particular Marta, Francesco, Tommaso and Laura for always having time to give me advice and discuss with me about statistics.

This Page Intentionally Left Blank

Abstract

The main topics of the thesis are dependent processes and their uses in Bayesian nonparametric statistics. With the term *dependent processes*, we refer to two or more infinite dimensional random objects, i.e., random probability measures, completely random measures, and random partitions, whose joint probability law does not factorize and, thus, encodes non-trivial dependence. We investigate properties and limits of existing nonparametric dependent priors and propose new dependent processes that fill gaps in the existing literature. To do so, we first define a class of priors, namely *multivariate species sampling processes*, which encompasses many dependent processes used in Bayesian nonparametrics. We derive a series of theoretical results for the priors within this class, keeping as main focus the dependence induced between observations as well as between random probability measures. Then, in light of our theoretical findings, as well as considering specific motivating applications, we develop novel prior processes outside this class, enlarging the types of data structures and prior information that can be handled by the Bayesian nonparametric approach. We propose three new classes of dependent processes: *full-range borrowing of information priors*, *invariant dependent priors* (with a focus on symmetric hierarchical Dirichlet processes), and *dependent priors for panel count data*. Full-range borrowing of information priors are dependent random probability measures that may induce either positive or negative correlation across observations and, thus, they achieve high flexibility in the type of induced dependence. Moreover, they introduce an innovative idea of borrowing of information across samples which differs from classical shrinkage. Invariant dependent priors are instead dependent random probabilities that almost surely satisfy a specified invariance condition, e.g., symmetry. They may be employed both when a priori knowledge on the shape of the unknown distribution is available or, as we do, to flexibly model errors terms in complex models without losing identifiability of other parameters of interest. Finally, dependent priors for panel count data are flexible priors based on completely random measures, that take into account dependence between the observed counts and the frequency of observation in panel count data studies. We study a priori and a posteriori properties of all the proposed models, develop algorithms to derive inference, compare the performances of our proposals with existing methods, and apply these constructions to simulated and real datasets. Through all the thesis, we try to balance theoretical and methodological results with real-world applications.

This Page Intentionally Left Blank

Contents

This Page Intentionally Left Blank

Glossary of Symbols

$a.s.$	Almost surely
$\#A$	Cardinality of the set A
$X \perp Y$	Independence between X and Y
$(\theta)_n$	Rising factorial $(\theta)_n = \Gamma(\theta + n)/\Gamma(\theta)$
$\mathcal{B}(\mathcal{P}_{\mathbb{X}})$	Borel σ-algebra on $\mathcal{P}_{\mathbb{X}}$ with respect to the distance of weak convergence
$\mathcal{B}(\mathbb{R})$	Borel σ-algebra on $\mathbb{R}$
Beta(α, β)	Beta distribution with parameters α and β
Cov(X, Y)	Covariance between X and Y
Corr(X, Y)	Correlation between X and Y
$CRM(v)$	Law of a completely random measure with Lévy intensity v
$\stackrel{d}{=}$	Equal in distribution
δ_x	Dirac measure at x
Δ_k	k-dimensional simplex $\Delta_k = \{(p_1, \ldots, p_{k+1}) : p_i \geq 0$ and $\sum_{i=1}^{k+1} p_i = 1\}$
$D_{k-1}(\alpha_1, \ldots, \alpha_k)$	Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_k$
DP(θ, P_0)	Dirichlet process with concentration θ e base distribution P_0
$\mathbb{E}(X)$	Expected value of X
E	Euclidean space
$\mathcal{E}$	Borel σ-algebra on E
${}_pF_q$	generalized hypergeometric function
Γ	Gamma function $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\mathrm{d}x$, for $z > 0$
Gammma(α, β)	Gamma distribution with shape α and rate β
$\overset{iid}{\sim}$	Independently and identically distributed
$\overset{ind}{\sim}$	Independently distributed
InvGamma(α, β)	Inverse gamma distribution with shape α and scale β
$\tilde{\mu}$	Completely random measures
$\mathbb{M}_{\mathbb{X}}$	Space of boundedly finite measures
$\mathcal{M}_{\mathbb{X}}$	Borel σ-algebra on $\mathbb{M}_{\mathbb{X}}$
Multinomial(n, p)	Multinomial distribution for n trials and vector of event probability p
$[n]$	Set of the first n natural numbers $[n] = \{1, \ldots, n\}$
$\mathbb{N}$	Set of natural numbers, i.e. positive integers $\{1, 2, \ldots\}$
$\mathbb{N}_0$	Set of non-negative integers $\{0, 1, \ldots\}$
N (μ, σ^2)	Law of a normal distribution with mean μ and variance σ^2
N$_k$ $(\boldsymbol{\mu}, \Sigma)$	k-variate normal distribution with mean vector $\boldsymbol{\mu}$ and var-cov matrix Σ
NIG$(\mu, \tau, \alpha, \beta)$	Normal inverse gamma distribution
$NRMI(\rho, \theta, P_0)$	Normalized random measure with independent increments with Lévy intensity $v(\mathrm{d}s, \mathrm{d}x) = \theta\, \rho(\mathrm{d}s)\, P_0(\mathrm{d}x)$

$\tilde{p}$	Random probability measures
$\mathbb{P}$	Probability measure on $(\Omega, \mathcal{F})$
P_n	Empirical distribution of n observations
$\mathcal{P}([n])$	Space of partitions of $[n]$
$\mathcal{P}_{\mathbb{X}}$	Space of probability measures on $\mathbb{X}$
Poisson(λ)	Poisson distribution with rate λ
$PY(\sigma, \theta, P_0)$	Pitman-Yor process
$\mathbb{R}$	Real line
$\mathbb{R}^+$	Positive real line
$\sigma(\Theta)$	Borel σ-algebra on Θ
Θ	Finite–dimensional parameter space
U(a, b)	Continuous uniform distribution on (a, b)
Var(X)	Variance of X
$\mathbb{X}$	Polish Space
$\mathcal{X}$	Borel σ-algebra on $\mathbb{X}$
Ω	Measure space

Glossary of Acronyms and Abbreviations

ALCPO	Average logarithmic CPO
ANOVA	Analysis of Variance
BNP	Bayesian nonparametrics
c.d.f	Cumulative distribution function
CLS	Compatible latent structure
CPO	Conditional predictive ordinates
CRF	Chinese restaurant franchise
CRM	Completely random measures
CRV	Completely random vector
DGP	Data generating process
DDP	Dependent Dirichlet process
DP	Dirichlet process
DPM	Dirichlet process mixture
EPPF	Exchangeable partition probability function
FuRBI	Full–range borrowing of information
GM	Griffiths–Milne
HDP	Hierarchical Dirichlet process
IDP	Invariant Dirichlet process
MCMC	Markov chain Monte Carlo
MLCPO	Median logarithmic CPO
mPPF	Multivariate prediction probability function
mSSM	Multivariate species sampling model
mSSP	Multivariate species sampling process
mSSS	Multivariate species sampling sequence
NCoRM	Normalized compound random measures
NDP	Nested Dirichlet process
NRMI	Normalized random measure with independent increments
OU	Ornstein-Uhlenbeck
p.d.f.	Probability density function
pEPPF	Partially exchangeable partition probability function
PP	Poisson process
PRM	Poisson random measures
s-HDP	Symmetric hierarchial Dirichlet process
SSM	Species sampling model
SSP	Species sampling process

This Page Intentionally Left Blank

Introduction

The past three decades have seen an increased availability of high-dimensional and complex structured datasets in many fields and applications such as genetics, ecology, and natural language processing, to name just a few. Such amount of information has required new statistical models and methodologies. Flexibility, interpretability, reasonable computational time, and quantifiable uncertainty are among the most important features for a statistical model nowadays. Bayesian nonparametric statistics provides a solid, coherent, and principled framework that nicely fits this new scenario as it avoids strong assumptions on data generative processes as well as the black-box approach of algorithmic modeling.

Bayesian nonparametric statistics dates its roots back to 1937, when de Finetti derived the theorem that bears his name (de Finetti, 1937) and that contains the theoretical foundations of the Bayesian nonparametric approach. However, only more recently, it developed into the outright and flourishing field that it is today. In 1972, D. V. Lindley was writing *"Nonparametric statistics. This is a subject about which the Bayesian method is embarrassingly silent"*, (Lindley, 1972, p. 66). Nonetheless, just a year after, T. S. Ferguson published his work on the Dirichlet process (Ferguson, 1973), thanks to which nonparametric techniques have become an appreciable and effective approach within the Bayesian framework.

Standard parametric and nonparametric Bayesian models typically assume exchangeability of the observables, which is a homogeneity condition implying the existence of a random probability measure, conditionally on which data can be seen as independent and identically distributed (de Finetti, 1937). The literature on nonparametric priors in this setting is well-established. They consist in the law of a single random probability measure, which most often can be obtained as a transformed completely random measure (Kingman, 1967).

However, real data usually present a level of heterogeneity that makes exchangeability an unrealistic assumption and Bayesian models require many dependent random probabilities to be constructed. In this framework, the study of dependence between random probabilities and observations is a very interesting topic. It provides intuitions on how existing models behave and how to best construct tailored priors to model real data. Recently there has been a growing literature devoted to nonparametric models for non-exchangeable data (see, for reviews, Foti & Williamson, 2015; Müller et al., 2015; Quintana et al., 2020) and this thesis aims at bringing further advances to this research area.

The thesis is organized in six almost self-contained chapters and three appendixes. Chap-

ter 1 contains a review of the literature that is most relevant to the novel works presented in subsequent chapters. We introduce formally the concept of exchangeability and present some nonparametric models satisfying this assumption, with a focus on models based on the Dirichlet process and on completely random measures. We introduce then the notion of *partial exchangeability*, which is a natural generalization of exchangeability suited to deal with data that are grouped into distinct samples. Partial exchangeability would be the main assumption from Chapter 2 to Chapter 4. We provide also a brief review of the literature of Bayesian nonparametric models for partial exchangeable data.

Chapter 2 is devoted to the introduction and derivation of theoretical results of what we call *multivariate species sampling models*. They are a wide class of dependent nonparametric processes that can be used to model partially exchangeable data. They are a very natural generalization of the species sampling models introduced by Pitman (1996) to a multivariate setting. It appears that the vast majority of the almost-sure discrete prior currently used to model partially exchangeable data belong to this class. Therefore, we believe that the original results in this chapter are of great interest, because they constitute a nice and formal framework to understand *'where we are'* and *'what can be done further'* in the research field of nonparametric models for partially exchangeable data. Moreover, as explained in the chapter, many of the results and ideas underlying multivariate species sampling processes can be generalized even beyond partial exchangeability. One of the findings of the second chapter is that multivariate species sampling models imply a non-negative correlation between observations in different samples. However such property, which is strictly connected to the idea *borrowing of information* across populations, is neither implied by partial exchangeability nor always appropriate in some applications.

In Chapter 3, we extend the study of dependence to a wider class of nonparametric models also outside the class of multivariate species sampling processes, introducing the notion of hyper-ties. We show how hyper-ties play a crucial role in driving the correlation between observations in different samples and thus borrowing of information. We note that existing nonparametric priors either do not allow an explicit evaluation of the value of the correlation or, when they do, they are able to induce only non-negative correlation. Thus in this chapter we propose a novel class of dependent nonparametric priors, which may induce either positive or negative correlation across samples based on the value of a hyperparmeter. Our proposal not only fills a gap in the literature of partially exchangeable models, but also introduces a new and more flexible idea of borrowing of information. Moreover, many of the models in the literature can be obtained as specific cases of the one proposed in this chapter. We investigate prior and posterior theoretical properties of the model and develop algorithms to perform posterior inference. The merits of our proposal are further discussed through illustrative examples on simulated and real data, where our model outperforms competing ones.

Chapter 4 focus on dependent nonparametric priors that satisfy invariance conditions, the most obvious example being symmetry. In order to impose such conditions, we need to

develop a new prior that again lies outside the class of multivariate species sampling processes. The processes introduced in this chapter are useful both when prior information about the observable is avalaible or when the nonparametric construction is used to model latent error terms. Even though the proposed priors can be employed for different inferential goals, here they are used as main building block in a broader model. This chapter has a more applied flavor with respect to the previous ones. The main motivating application is the development of a flexible and interpretable model selection procedure to study the relationship between cardiac dysfunctions and hypertensive disorder of pregnancy. Hypertensive disorders of pregnancy are diseases that occur in about 10% of pregnant women around the world. Though there is evidence that hypertension impacts maternal cardiac functions, the relation between hypertension and cardiac dysfunctions is only partially understood. The study of this relationship can be framed as a joint inferential problem on multiple populations, each corresponding to a different hypertensive disorder diagnosis, that combines multivariate information provided by a collection of cardiac function indexes. A Bayesian nonparametric approach seems particularly suited for this setup and we demonstrate it on a dataset consisting of transthoracic echocardiography results of a cohort of Indian pregnant women. We are able to perform model selection, provide density estimates of cardiac function indexes and a latent clustering of patients: these readily interpretable inferential outputs allow to single out modified cardiac functions in hypertensive patients compared to healthy subjects and progressively increased alterations with the severity of the disorder. The analysis relies on a novel hierarchical structure, called symmetric hierarchical Dirichlet process, which is a specific example of invariant dependent process. This is suitably designed so that the mean parameters are identified and used for model selection across populations, a penalization for multiplicity is enforced, and the presence of unobserved relevant factors is investigated through a latent clustering of subjects. Posterior inference relies on a suitable Markov chain Monte Carlo algorithm and the model behaviour is also showcased on simulated data.

Chapter 5 deals with dependent processes for panel count data, where for each subject cumulative counts are recorded at discrete time points. Both the time points and the cumulative counts are realizations of point processes, namely the *observation process* and the *event process*. Anytime prior information about dependence between counts and observation times is available, independence assumptions should clearly be avoided. Here we use completely random measures to define nonparametric priors that may reflect positive associations between the event and the observational processes underlying the observations. Chapter 5 interestingly shows how dependent processes may be employed even on non-partially exchangeable data.

The thesis concludes with Chapter 6 which contains related ideas and further extensions of the works in previous chapters.

This Page Intentionally Left Blank

Chapter 1

A Guided Tour on the Basics of Bayesian Nonparametrics

This first introductory chapter contains a literature review of those topics that are most relevant for the following chapters. The chapter is structured in four main sections. In Section 1.1 we introduce the concept of exchangeability, which is the usual starting assumption of classical Bayesian models. Section 1.2 contains a review of the Dirichlet process and the Dirichlet process mixture model, which is undoubtedly the most famous and used Bayesian nonparametric model. Moreover, in Section 1.2, it can be found a review of the invariant Dirichlet process, which is not only a prior closely related to the Dirichlet process but is also specifically relevant for the contribution in Chapter 4, where this process is extended to model non-exchangeable data. Section 1.3 is devoted to the description of completely random measures and their uses in Bayesian nonparametric models for exchangeable data. Completely random measures and their multivariate extension will be of particular interest for the reader of this thesis in light both of Chapter 3 and Chapter 5. Finally, section 1.4 introduces the concept of partial exchangeability, which is a natural extension of exchangeability for data grouped into distinct but related samples and that would be a main reference framework for all the chapters in this thesis, except for Chapter 5. In this section, we also review what we consider the most significant Bayesian nonparametric models for partial exchangeability. All these models are particular cases of the novel general class of processes that will be introduce in Chapter 2.

1.1 Exchangeability and prior processes

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Polish space $\mathbb{X}$ equipped with the corresponding Borel σ-algebra $\mathcal{X}$. Denote with $(X_n)_{n\geq 1}$ a sequence of observable random variables each taking values in $(\mathbb{X}, \mathcal{X})$, such that observed data $(x_1, \ldots, x_n)$ are a realization of $(X_1, \ldots, X_n)$. A classical assumption of Bayesian models is that the sequence $(X_n)_{n\geq 1}$ is exchangeable.

Definition 1.1 (Exchangeability). *A sequence of random variables $(X_n)_{n\geq 1}$ such that, for any $n \geq 2$, the law of $(X_1, \ldots, X_n)$ is invariant with respect to permutations of its elements, i.e.*

$$(X_1, \ldots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)})$$

for any σ permutation of $[n] = \{1, 2, \ldots, n\}$, is said exchangeable.

From a modeling point of view, assuming exchangeability of the observables means that the order of the data is not informative and, therefore, it should not affect inference results. This is a typical situation in many cross-sectional studies, when a sample is drawn from a single population at a specific point in time. Firstly, notice that it follows directly from Definition 1.1 that random variables in an exchangeable sequence are marginally identically distributed. Moreover, thanks to B. de Finetti and his representation theorem, we also know that exchangeable random variables can be represented as conditionally independent and identically distributed. More formally, denote with $\mathcal{P}_{\mathbb{X}}$ the space of all probability measures on $\mathbb{X}$, which, if endowed with the distance of weak convergence, is a Polish space with respect to the Borel σ-algebra $\mathcal{B}(\mathcal{P}_{\mathbb{X}})$.

Theorem 1.1 (de Finetti, 1937). *A sequence of random variables $(X_n)_{n\geq 1}$ is exchangeable if and only if there exists a probability measure Q on $\mathcal{P}_{\mathbb{X}}$ such that, for any $n \geq 1$ and $A_1, A_2, \ldots, A_n$, with $A_i \in \mathcal{X}$ for $i = 1, \ldots, n$,*

$$\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} p(A_i)\, Q(dp).$$

where Q is said de Finetti measure.

The theorem can be equivalently restated as

$$\begin{aligned} X_i \mid \tilde{p} &\overset{iid}{\sim} \tilde{p} \qquad \text{for } i = 1, \ldots, n \\ \tilde{p} &\sim Q \end{aligned} \tag{1.1}$$

where $\tilde{p}$ is a measurable function from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathcal{P}_{\mathbb{X}}, \mathcal{B}(\mathcal{P}_{\mathbb{X}}))$, i.e., a random probability measure. Therefore, any exchangeable sequence of data can be represented through two elements: $\tilde{p}$, which is the conditional law of the data, and a prior distribution Q, which can be seen as a probability law reflecting information and uncertainty about $\tilde{p}$. Notice that $\tilde{p}$ can be interpreted as a stochastic process taking values in $[0, 1]$ with index set given by $\mathcal{X}$ and whose law is define by the prior Q; for this reason we refer to the random measure $\tilde{p}$ also with the term *process*. Consider also that to transform (1.1) into a working model for analyzing data, it is enough to choose the law Q appropriately and for this reason, when a specific prior Q is specified we refer to (1.1) with the term *model*.
In parametric models, the prior Q is chosen to be degenerate on a finite-dimensional sub-

space of $\mathscr{P}_{\mathbb{X}}$. For example, the normal-normal model

$$\begin{aligned} X_i \mid \theta &\overset{iid}{\sim} \mathrm{N}(\theta,\, 1) \qquad \text{for } i = 1, \ldots, n \\ \theta &\sim \mathrm{N}(0,\, 1) \end{aligned}$$

coincides with a prior Q such that $Q(\{\tilde{p} \neq \mathrm{N}(\theta,\, 1), \theta \in \mathbb{R}\}) = 0$. Even though choices of this type simplify the inferential procedure, they also usually coincide with a overly informative prior that rarely can be justified in real applications. Conversely, in nonparametric problems Q has an infinite-dimensional support. Enlarging the prior support, nonparametric models permit to relax unrealistic parametric assumption in favor of more flexible constructions. However, this usually comes with a cost in terms of model complexity, difficulties in the interpretation of results and heavier computational burden. These are some of the reasons why further investigation on Bayesian nonparametric models is still needed nowadays.

1.2 Dirichlet process models

The Dirichlet process (DP) was introduced by Ferguson (1973) and it is the most celebrated Bayesian nonparametric prior. To describe the value of the DP, S. Ghosal and A. Van der Vaart wrote "*the importance of the Dirichlet process in Bayesian nonparametrics is comparable to that of the normal distribution in probability and general statistics*", (Ghosal & Van der Vaart, 2017, p. 96). The DP laid the foundations for many Bayesian nonparametric models, the vast majority of which can be interpreted as extensions of the DP itself.

1.2.1 Dirichlet process

The Dirichlet process admits many equivalent definitions, however the first provided by T.S. Ferguson is the one that makes use of the Dirichlet distribution and that gives the name to the process. Thus, we firstly recall the definition of Dirichlet distribution.

Definition 1.2 (Dirichlet distribution)**.** *Consider the (k-1)-dimensional probability simplex* $\Delta_{k-1} = \{(p_1, \ldots, p_k) : p_i \geq 0 \text{ and } \sum_{i=1}^{k} p_i = 1\}$*, for some* $k \in \mathbb{N} \setminus \{1\}$*. A probability distribution on* Δ_{k-1} *is Dirichlet with parameter* $(\alpha_1, \ldots, \alpha_k)$*, if the corresponding density* f *(with respect to the Lebesgue measure on* $\mathbb{R}^{k-1}$*) is*

$$f_k(p_1, \ldots, p_k) = \begin{cases} \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} p_1^{\alpha_1} \ldots p_{k-1}^{\alpha_{k-1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{\alpha_k} & \text{for } (p_1, \ldots, p_k) \in \Delta_{k-1} \\ 0 & \text{otherwise} \end{cases}$$

where Γ *is the Gamma function.*

We use the notation $(p_1, \dots, p_k) \sim D_{k-1}(\alpha_1, \dots, \alpha_k)$ to indicate that $(p_1, \dots, p_k)$ is distributed according to a Dirichlet distribution. More details on the Dirichlet distribution can be found in Appendix A, along with a brief review of the two most typical uses of the Dirichlet distribution in Bayesian statistics, namely the Multinomial-Dirichlet model and finite mixture models.

Definition 1.3 (Dirichlet Process)**.** *Consider a non-null measure α on $(\mathbb{X}, \mathcal{X})$ with $\theta := \alpha(\mathbb{X}) \in \mathbb{R}^+$. $\tilde{p}$ is a Dirichlet process (DP) on $(\mathbb{X}, \mathcal{X})$ with parameter α if $\tilde{p}(\mathbb{X}) = 1$ a.s. and for any $k \in \mathbb{N} \setminus \{1\}$ and measurable partition $(A_1, \dots, A_k)$ of $\mathbb{X}$*

$$(\tilde{p}(A_1), \dots, \tilde{p}(A_k)) \sim D_{k-1}(\alpha(A_1), \dots, \alpha(A_k))$$

where θ is called concentration parameter (or total mass) and the probability measure $P_0(\cdot) = {}^{\alpha(\cdot)}/_{\alpha(\mathbb{X})}$ is called baseline probability measure.

We use the notation $\tilde{p} \sim \text{DP}(\, \theta,\, P_0\,)$ to indicate that the random measure $\tilde{p}$ is distributed accordingly to a DP with concentration parameter θ and baseline P_0. We refer to Ferguson (1973) for proof of existence of the process. From Definition 1.3, it is immediate to compute the marginal moments of a DP, in particular, if $\tilde{p} \sim \text{DP}(\, \theta,\, P_0\,)$, then for any $A \in \mathcal{X}$

$$\begin{aligned} \mathbb{E}[\,\tilde{p}(A)\,] &= P_0(A) \\ \text{Var}[\,\tilde{p}(A)\,] &= \frac{P_0(A)(1 - P_0(A))}{1 + \theta} \end{aligned}$$

Therefore the baseline P_0 is the mean measure of the DP, while the concentration parameter controls the variability, so that the higher θ, the lower the variability. By Theorem 1.1 by de Finetti we also have that, if

$$\begin{aligned} X_i \mid \tilde{p} &\sim \tilde{p} \qquad\qquad \text{for } \; i = 1, \dots, n \\ \tilde{p} &\sim \text{DP}(\, \theta,\, P_0\,) \end{aligned} \tag{1.2}$$

then $\mathbb{P}[X_i \in A] = \mathbb{E}[\,\tilde{p}(A)\,]$. Thus, when the DP is used as a prior in an exchangeable model, P_0 is the marginal law of one observation. The two most important features of the Dirichlet process are its full weak support and conjugacy, the former ensures flexibility, while the latter guarantee tractability of DP–based models. They are formally stated in the next two theorems.

Theorem 1.2 (Ferguson, 1973)**.** *Given the model in equation (1.2), it follows that*

$$\tilde{p} \mid X_1, \dots, X_n \sim DP\left(\theta + n,\, \frac{\theta}{\theta + n} P_0 + \frac{n}{\theta + n} P_n\right)$$

where P_n is the empirical distribution of $X_1, \dots, X_n$, i.e. $P_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}$, where δ_x denotes the Dirac measure giving mass 1 to the point x.

Theorem 1.3 (Majumdar, 1992). *Consider the weak support of* $Q = DP(\theta, P_0)$, *i.e.* $S_Q = \bigcap_{A \in \mathcal{A}} A$ *with* $\mathcal{A} = \{A \in \mathcal{B}(\mathcal{P}_{\mathbb{X}}) : A \text{ is closed}, Q(A^c) = 0\}$, *then*

$$S_Q = \{p \in \mathcal{P}_{\mathbb{X}} : S_p \subset S_{P_0}\}$$

where S_p *and* S_{P_0} *denote the support of* p *and* P_0 *with respect to* $\mathcal{X}$.

Theorem 1.3 was first proved by Ferguson (1973) for $\mathbb{X} = \mathbb{R}$ and then generalized to Polish spaces by Majumdar (1992).
The joint distribution of a random sample drawn from a DP can be described through the generalized Pólya urn scheme of Blackwell and MacQueen provided here below.

Theorem 1.4 (Blackwell & MacQueen, 1973). *Let* $X_i \mid \tilde{p} \overset{iid}{\sim} \tilde{p}$, *with* $\tilde{p} \sim DP(\theta, P_0)$, *then*

$$X_1 \sim P_0$$

$$X_{i+1} \mid X_1, \ldots, X_i \sim \frac{\theta}{\theta + i} P_0 + \frac{1}{\theta + i} \sum_{j=1}^{i} \delta_{X_j}$$

It is evident from the previous theorem that the DP induces ties with positive probability between the random variables in the sequence $\boldsymbol{X} = (X_n)_{n \geq 1}$, i.e. $\mathbb{P}[X_i = X_j] > 0$.
The generalized Pólya urn scheme for the DP provides an easy procedure to sample a sequence of exchangeable random variables $\boldsymbol{X} = (X_n)_{n \geq 1}$ whose de Finetti measure is the law of a DP. However, in many cases one may be interested in sampling the process itself. To do so there are different possible constructive representations, we report the probably most celebrated one, i.e. the stick breaking representation, in the next theorem. It is due to Sethuraman & Tiwari (1982) and Sethuraman (1994).

Theorem 1.5 (Sethuraman, 1994). *If* $\tilde{p} \sim DP(\theta, P_0)$, *then*

$$\tilde{p} \overset{a.s.}{=} \sum_{h=1}^{+\infty} \pi_h \, \delta_{\phi_h}$$

$$\text{with} \qquad \pi_h = \pi'_h \prod_{r=1}^{h-1} (1 - \pi'_r) \qquad \pi'_r \overset{iid}{\sim} Beta(1, \theta) \qquad \phi_h \overset{iid}{\sim} P_0.$$

Notice that the almost sure discreteness of the DP induces a random partition $\Pi_n = \{A_1, \ldots, A_k\}$ over the set $\{X_1, \ldots, X_n\}$ so that X_i and X_j are both in A_l if and only if $X_i = X_j$.

Theorem 1.6 (Antoniak, 1974). *Let* $X_i \mid \tilde{p} \overset{iid}{\sim} \tilde{p}$, *with* $\tilde{p} \sim DP(\theta, P_0)$, *then the probability of observing a specific partition* $\{A_1, \ldots, A_k\}$ *of the elements in* $\{X_1, \ldots, X_n\}$ *consisting of* $k \leq n$ *distinct values with respective frequencies* $n_1, \ldots, n_k$ *coincides with*

$$\mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\}) = \frac{\theta^k}{(\theta)_n} \prod_{i=1}^{k} (n_i - 1)!$$

where $(\theta)_n = \Gamma(\theta + n)/\Gamma(\theta)$.

More details on random partitions can be found in Chapter 2.

1.2.2 Dirichlet process mixture model

The almost sure discreteness of the DP makes it inappropriate as prior over densities. However, such limitation can be overcome through the so-called Dirichlet process mixture (DPM) model introduced by Ferguson (1983) and Lo (1984). According to a DPM model, the common density $f(x)$ generating the data is obtained as an infinite mixture with a Dirichlet process as mixing distribution.
To formally describe the model, we need to introduce the notion of probability kernel. Let $(\Theta, \sigma(\Theta))$ be a finite-dimensional parameter space endowed with the corresponding σ-algebra and k be a mapping from $\mathbb{X} \times \Theta$ into $[0,\, 1]$, such that, for every fixed $\theta \in \Theta$, the map $x \mapsto k(x;\theta)$ is a probability density function (p.d.f.) with respect to a given σ-finite measure v, and, for every fixed $x \in \mathbb{X}$, the map $\theta \mapsto k(x;\theta)$ is measurable.

Definition 1.4 (Dirichlet process mixture 1)**.** *A sequence of random variables* $(X_i)_{i=1}^n$ *follows a Dirichlet process mixture (DPM) model if*

$$\begin{aligned} X_i \mid \tilde{p} &\overset{iid}{\sim} f(x) = \int_\Theta k(x;\theta)\tilde{p}(\mathrm{d}\theta) \qquad \textit{for } i = 1, \dots, n \\ \tilde{p} &\sim DP(\,\alpha,\, P_0\,) \end{aligned}$$

Introducing a sequence of latent parameters $(\theta_i)_{i=1}^n$, the model can be conveniently rewritten as

$$\begin{aligned} X_i \mid \theta_i &\overset{ind}{\sim} k(x;\theta_i) \qquad \theta_i \mid \tilde{p} \overset{iid}{\sim} \tilde{p} \qquad \text{for } i = 1, \dots, n \\ \tilde{p} &\sim \mathrm{DP}(\,\alpha,\, P_0\,) \end{aligned}$$

DPM models are mainly used to estimate the density $f(x)$ or, even more often, for clustering observations. As for the former, it is important to notice that conjugacy does not hold anymore and the posterior distribution of $f(x)$ is not available in closed form. However, conditioning on the sequence of latent parameters and by the conjugacy of the DP (see Theorem 1.2) one has

$$\tilde{p} \mid X_1, \dots, X_n, \theta_1, \dots, \theta_n \sim DP\left(\alpha + n,\, \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{\theta_i}\right) \tag{1.3}$$

Since $f(x)$ is a deterministic linear transformation of $\tilde{p}$, its posterior estimate under a square

loss function (conditioning on the latent parameters) is

$$\mathbb{E}\left[f(x) \mid X_1, \dots, X_n, \theta_1, \dots, \theta_n\right] = \frac{\alpha}{\alpha + n} \int k(x;\theta) P_0(\mathrm{d}\theta) + \frac{1}{\alpha + n} \sum_{i=1}^{n} k(x;\theta_i)$$

Thanks to these results, posterior inference can be rather easily conducted through the Gibbs-sampling algorithms provided in Escobar (1994) and Escobar & West (1995) and refined in Neal (2000). An interesting result, that comes from (1.3), is obtained if one marginalizes out the latent parameters to get the posterior.

$$\tilde{p} \mid X_1, \dots, X_n \sim \int DP\left(\alpha + n,\ \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{\theta_i}\right) \mathrm{d}p\left((\theta_i)_{i=1}^n \mid (X_i)_{i=1}^n\right)$$

where $p\left((\theta_i)_{i=1}^n \mid (X_i)_{i=1}^n\right)$ is the posterior distribution of the latent parameters. The posterior distribution of the mixing measure $\tilde{p}$ turns out to be a mixture of DPs, which is a construction studied by Antoniak (1974).

As anticipated, DPM models are often used to perform model–based clustering. To clarify this point, notice that the prior $\tilde{p}$ gives positive probability to ties among the latent parameters, i.e. $\mathbb{P}[\theta_i = \theta_j] > 0$, and thus induces a random partition Π_n, that has been described at the end of Section 1.2.1. Using the partition on the latent parameters, we can define a natural clustering rule among observations and cluster together X_i and X_j if and only if $\theta_i = \theta_j$. Moreover, we denote with $\theta_1^*, \dots, \theta_k^*$ the distinct values (in order of appearance) in $(\theta_1, \dots, \theta_n)$, where $k \leq n$. For notation and computational convenience a set of cluster membership indicators is usually introduced, namely $c_1, c_2, \dots, c_n$, with $c_i \in [k]$, so that $c_i = c_j$ if and only if $\theta_i = \theta_j$ and $c_i = c$ if and only if $\theta_i = \theta_c^*$. In light of this, a DPM model can be equivalently restated as follows.

Definition 1.5 (Dirichlet process mixture 2)**.** *A sequence of random variables $(X_i)_{i=1}^n$ follows a Dirichlet process mixture (DPM) model if*

$$\begin{aligned} X_i \mid c_i, \theta_1^*, \dots, \theta_k^* &\overset{ind}{\sim} k(x;\theta_{c_i}^*) && \text{for } i = 1, \dots, n \\ (c_1, \dots, c_n) &\sim Q_\alpha \\ \theta_c^* &\overset{iid}{\sim} P_0 && \text{for } c = 1, \dots, k \end{aligned}$$

where Q_α denotes the distribution of the cluster membership indicators associated to the partition induced by a Dirichlet process with concentration parameter α.

Denoting with k_i the number of unique values in $\{\theta_1, \dots, \theta_i\}$ and defining $n_j^{(i)} = \#\{\theta_l : \theta_l = \theta_j^* \text{ and } l \in \{1, \dots, i\}\}$, for $j \in [k_i]$, it is immediate to show that Q_α coincides with

$$c_1 \sim \delta_1$$

$$c_i \mid c_1, \dots, c_{i-1} \sim \sum_{j=1}^{k_i} \frac{n_j^{(i)}}{\alpha + i} \delta_j + \frac{\alpha}{\alpha + i} \delta_{k_i+1}$$

From an interpretation point of view, two observations are clustered together if they come from the same parametric mixture component, this is the reason why DPM models are ascribed to model-based clustering techniques.

1.2.3 Invariant Dirichlet process

The invariant Dirichlet process (IDP) was introduce by Dalal (1979a). It will serve as building block for the original model presented in Chapter 4. After recalling the definition, we present two representations of the process: the former is the analogue of the stick-breaking representation, while the latter is an extension of the generalized Pólya urn scheme of Blackwell & MacQueen (1973). We will then conclude the section with a constructive definition for the symmetric-IDP and define the symmetric Dirichlet process mixture (symmetric-DPM). The symmetric-DPM will be the analogue of the Dirichlet process mixture of Lo (1984) for symmetric distributions.
Let $(E, \mathcal{E})$ be any p-dimensional measurable Euclidean space. Let $\mathcal{G} = \{g_1, \dots, g_L\}$ be a finite group of measurable transformations on $(E, \mathcal{E})$.

Definition 1.6 (Invariant Probability)**.** *A probability measure P_0 on $(E, \mathcal{E})$ is a $\mathcal{G}$-invariant probability distribution, if $P_0(A) = P_0(g_l(A))$, $\forall A \in \mathcal{E}$ and $\forall l = 1, \dots, L$.*

Definition 1.7 (Invariant Random Probability)**.** *A random probability $\tilde{p}$ on $(E, \mathcal{E})$ is said $\mathcal{G}$-invariant, if it is almost surely $\mathcal{G}$-invariant.*

Definition 1.8 (Invariant Partition)**.** *A measurable partition $A_1, A_2, \dots, A_K$ of E is $\mathcal{G}$-invariant partition, if $A_k = g_l(A_k)$, $\forall k = 1, \dots, K$ and $\forall l = 1, \dots, L$.*

Definition 1.9 (Invariant Dirichlet Process)**.** *A random probability $\tilde{p}$ is an invariant Dirichlet process with group of transformations $\mathcal{G}$, if*

1. *$\tilde{p}$ is almost surely $\mathcal{G}$-invariant*

$$\tilde{p}(A) = \tilde{p}(g_l(A)) \quad \text{for } l = 1, \dots, L \quad \text{a.s.}$$

2. *there exists a $\mathcal{G}$-invariant probability distribution P_0 on $(E, \mathcal{E})$ and $\alpha \in \mathbb{R}^+$, such for any $k \in \mathbb{N}$ and any $\mathcal{G}$-invariant measurable partition $A_1, \dots, A_k$ of E*

$$\big(\tilde{p}(A_1), \dots, \tilde{p}(A_k)\big) \sim D_{k-1}(\alpha P_0(A_1), \dots, \alpha P_0(A_k))$$

where α is called concentration parameter and P_0 is called baseline probability measure.

We use the notation $\tilde{p} \sim \text{IDP}(\alpha,\, P_0,\, \mathcal{G})$ to indicate that the random measure $\tilde{p}$ is distributed according to a IDP. Notice that if $\tilde{p} \sim \text{DP}(\alpha,\, P_0)$, then $\tilde{p}$ is not an IDP, since it is not an invariant random probability. And, vice versa, if $\tilde{p} \sim \text{IDP}(\alpha,\, P_0,\, \mathcal{G})$, then $\tilde{p}$ is not a DP, since in general its finite dimensional distributions over non $\mathcal{G}$-invariant partitions do not follow a Dirichlet distribution. However there exists a strong connection between the two processes, which is provided by Dalal (1979a) with the following theorem.

Theorem 1.7 (Dalal, 1979a)**.** *Let $\tilde{q} \sim DP(\alpha,\, P_0)$ and $\tilde{p} \sim IDP(\alpha,\, P_0,\, \mathcal{G})$. Define*

$$q^*(\cdot) = \frac{1}{L}\sum_{l=1}^{L} \tilde{q}(g_l(\cdots))$$

then

$$\tilde{p} \stackrel{d}{=} q^*.$$

Tiwari (1988) provided also a constructive definition for the IDP, which is the analogue of the stick-breaking representation of Sethuraman (1994) for the DP.

Proposition 1.1 (Tiwari, 1988)**.** *If $\tilde{p} \sim IDP(\alpha,\, P_0,\, \mathcal{G})$, then*

$$\tilde{p} = \sum_{h=1}^{\infty} \pi_h \sum_{l}^{L} \delta_{g_l(\phi_h)}$$

$$\textit{with} \qquad \pi_h = \frac{\pi'_h}{L}\prod_{r=1}^{h-1}(1-\pi'_r) \qquad \pi'_r \sim \textit{Beta}(1,\, \alpha) \qquad \phi_h \overset{iid}{\sim} P_0.$$

While given that

$$\phi_i \mid \tilde{p} \overset{iid}{\sim} \tilde{p} \qquad \tilde{p} \sim IDP(\alpha,\, P_0,\, \mathcal{G})$$

integrating out $\tilde{p}$, we get the correspondent generalized Pólya urn representation for the process, which is

$$\phi_1 \sim P_0$$

$$\phi_i \mid \phi_1, \ldots \phi_{i-1} \sim \sum_{j=1}^{i-1} \frac{1}{i-1+\alpha}\left(\frac{1}{L}\sum_{l=1}^{L}\delta_{g_l(\phi_j)}\right) + \frac{\alpha}{i-1+\alpha}P_0$$

For more details about IDPs we refer to Dalal (1979a), Dalal (1979b), Hannum & Hollander (1983), Doss (1984), Diaconis & Freedman (1986), Tiwari (1988), Ferguson et al. (1992) and Ghosal et al. (1999).

1.3 Completely random measures and their uses in BNP

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Polish space $(\mathbb{X}, \mathcal{X})$, a random element $\tilde{\mu}$ is a random measure, if $\tilde{\mu}$ is a function $\tilde{\mu} : \Omega \times \mathbb{X} \to [0, +\infty]$ such that

1. $\tilde{\mu}(\omega, \cdot)$ is a measure on $(\mathbb{X}, \mathcal{X})$, for every $\omega \in \Omega$
2. $\tilde{\mu}(\cdot, A)$ is a positive random variable, for every $A \in \mathcal{X}$.

In the following, we are dropping ω and denoting the random measure as $\tilde{\mu}$ and the corresponding random variables on measurable sets as $\tilde{\mu}(A)$.
When random measures are used in statistical models, the space $\mathbb{X}$ is usually the space where observations (or latent parameters of the model) take values. We restrict our attention to the cases in which $\tilde{\mu}$ is almost surely boundedly finite, i.e. for any bounded $A \in \mathcal{X}$, we have $\tilde{\mu}(A) < \infty$ with probability one. We denote with $\mathbb{M}_{\mathbb{X}}$ the space of boundedly finite measures on $\mathbb{X}$ and with $\mathcal{M}_{\mathbb{X}}$, the topology of weak$^{\#}$ converge (cf. Daley & Vere-Jones, 2003, pp.402–406).

Definition 1.10 (Completely random measure - CRM)**.** *Consider the Polish space $(\mathbb{X}, \mathcal{X})$ and the space $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random element $\tilde{\mu}$ from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ is a completely random measure (CRM) if, for every collection of pairwise disjoint sets $(A_i)_{i\geq 1}^n$ in $\mathcal{X}$, the random variables $\tilde{\mu}(A_1)$, $\tilde{\mu}(A_2), \ldots, \tilde{\mu}(A_n)$ are mutually independent.*

In words, a CRM is a boundedly finite random measure that, when evaluated on disjoint measurable sets, gives rise to mutually independent random variables. CRMs were first introduced in Kingman (1967) on spaces $\mathbb{X}$ more general than Polish spaces. The only assumption made by Kingman on $(\mathbb{X}, \mathcal{X})$ is that singletons are measurables (i.e. $\{x\} \in \mathcal{X}$, for all $x \in \mathbb{X}$).
The probably most important property of CRMs is their almost surely discreteness, which is stated in the following theorem.

Theorem 1.8 (Kingman, 1967)**.** *If $\tilde{\mu}$ is a CRM on $(\mathbb{X}, \mathcal{X})$, then*

$$\mu \overset{a.s.}{=} \mu_f + \sum_{i=1}^{M} W_i \delta_{T_i} + \sum_{j=1}^{\infty} J_j \delta_{X_j}$$

where μ_f is a purely deterministic measure, $M \in \{0, \ldots, \infty\}$; $T_1, \ldots, T_M$ are fixed; $(J_j)_{j\geq 1}$, $(X_j)_{j\geq 1}$ and $(W_i)_{i\geq 1}$ are sequences of independent random variables such that $(J_j, X_j)_{j\geq 1}$ is independent from $(W_i)_{i\geq 1}$.

In what follows, if not differently specified, we are going to assume $\mu_f = 0$ and $M = 0$, so that the CRM has no deterministic component and no fixed jumps. Such CRMs are characterized by the following Laplace functional transform for any measurable function $f : \mathbb{X} \to \mathbb{R}^+$

$$\mathbb{E}\left[e^{-\int_{\mathbb{X}} f(x)\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{X}} [1 - e^{-sf(x)}] v(\mathrm{d}s, \mathrm{d}y)\right\} \tag{1.4}$$

where v is a measure on $\mathbb{R}^+ \times \mathbb{X}$, called *Lévy intensity*, that satisfies

$$\int_{\mathbb{R}_+ \times B} \min\{1, s\}\, v(\mathrm{d}s, \mathrm{d}x) < \infty, \quad \text{for any } B \in \mathcal{X}.$$

Theorem 1.9. *Any CRM $\tilde{\mu}$ (with $\mu_f = 0$ and $M = 0$) can be represented as a linear functional of a Poisson random measure (PRM) N, i.e.*

$$\tilde{\mu}(\mathrm{d}x) = \int_0^{+\infty} s\, N(\mathrm{d}s, \mathrm{d}x)$$

where N is a PRM on $\mathbb{R}^+ \times \mathbb{X}$ with mean measure equal to the Lévy intensity of $\tilde{\mu}$.

Lastly, in Chapter 3 and 5, we will also assume that CRMs are homogeneous which means that jumps $(J_j)_{j\geq 1}$ and locations $(X_j)_{j\geq 1}$ are independent. In terms of Lévy intensity it reads

$$v(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\alpha(\mathrm{d}x)$$

where ρ is a measure on $\mathbb{R}^+$, named *Lévy density*, and α is a non-atomic measure on $\mathbb{X}$, usually called the centring measure. Typical examples of CRMs are given by the *gamma* process,

$$v(\mathrm{d}s, \mathrm{d}x) = s^{-1}e^{-s}\mathrm{d}s\, \alpha(\mathrm{d}x), \quad s > 0,$$

the *homogeneous beta* process,

$$v(\mathrm{d}s, \mathrm{d}x) = \theta s^{-1}(1-s)^{\theta-1}\mathrm{d}s\, \alpha(\mathrm{d}x), \quad 0 < s < 1,\, \theta > 0,$$

the *σ-stable* process,

$$v(\mathrm{d}s, \mathrm{d}x) = \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma}\mathrm{d}s\, \alpha(\mathrm{d}x), \quad s > 0,\, 0 < \sigma < 1,$$

and the *superposed gamma* process,

$$v(\mathrm{d}s, \mathrm{d}x) = \frac{1 - e^{-\gamma s}}{1 - e^{-s}} s^{-1}e^{-s}\mathrm{d}s\, \alpha(\mathrm{d}x), \quad s > 0, \gamma > 0.$$

For more details on CRMs, we refer to Kingman (1967) and Kingman (1993).

1.3.1 Normalized random measures with independent increments

CRMs are often normalized to obtain random probability measures, called normalized random measures with independent increments, introduced in Regazzini et al. (2003). Clearly, one needs $\mathbb{P}(0 < \mu(\mathbb{X}) < \infty) = 1$ for the normalization to make sense, that in terms of Lévy

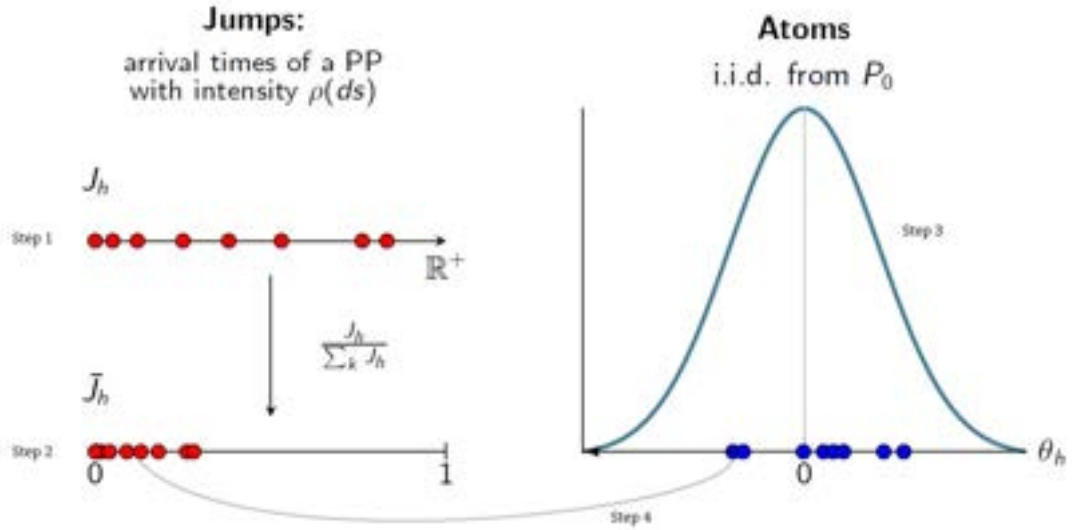


Figure 1.1: Simulation steps for a homogeneous NRMI with intensity $v(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\, P_0(\mathrm{d}x)$. At step 1 arrival times of a Poisson process (PP) on $\mathbb{R}^+$ with intensity $\rho(\mathrm{d}s)$ are sampled, at step 2 they are normalized to sum up to 1, at step 3 a atom from $P_0(\mathrm{d}x)$ is sampled for each jump, at step 4 the i.i.d. atoms are associated to the jumps.

intensity translates to $v(\mathbb{R}^+, \mathbb{X}) = \infty$ and

$$\psi(\lambda) := \int_{\mathbb{R}_+ \times \mathbb{X}} [1 - e^{-\lambda s}] v(\mathrm{d}s, \mathrm{d}x) < \infty \quad \text{for any } \lambda > 0,$$

where $\psi(\lambda)$ is called *Laplace exponent*. This explains the following definition.

Definition 1.11 (Normalized random measure with independent increments - NRMI)**.** *Let $\tilde{\mu}$ be a CRM on $\mathbb{X}$ with intensity v, such that $v(\mathbb{R}^+, \mathbb{X}) = \infty$ and $\psi(\lambda)$ is finite for any positive λ. Then the random probability measure*

$$\tilde{p}(\cdot) = \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}$$

is termed normalized random measure with independent increments (NRMI).

Notice that the definition above could be extended to any random measure $\tilde{\mu}$ such that $\mathbb{P}(0 < \mu(\mathbb{X}) < \infty) = 1$. However, the strength of CRMs lies in the representation in (1.4), that allows an unprecedented analytical tractability. Moreover, the class is fairly large and encompasses many interesting priors: for instance the well-known Dirichlet process is a normalized Gamma process. Figure 1.1 summarizes the steps to sample a NRMI starting from the correspondent intensity. NRMIs have been extensively studied to model exchangeable data (see, for instance, James et al., 2006, 2009, 2010; Lijoi & Prünster, 2010; Barrios et al., 2013; Nieto-Barajas et al., 2004; Favaro et al., 2016; Camerlenghi et al., 2018).

1.3.2 Hazard mixture model

CRMs have been effectively employed to model hazard functions, which are of particular interest in time-to-event analysis. The law of a random variable X, taking values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is typically described through the corresponding cumulative distribution function (c.d.f.), or, when its distribution is absolutely continuous with respect to the Lebesgue measure, through the corresponding p.d.f. $f(x)$. However, another equivalent instrument to describe the law of a real valued random variable is the hazard function $h(x)$, defined as

$$h(x) = \lim_{h \to 0^+} \frac{\mathbb{P}(x \leq X \leq x + h \mid X \geq x)}{h}$$

The hazard function be can interpreted as a conditional density and its linked to the p.d.f. through the following relation

$$f(x) = h(x) \exp\left\{ - \int_{-\infty}^{x} h(t) \mathrm{d}t \right\}$$

Hazard functions are the typical inferential goal in survival studies, where X is the time-to-event variable and takes values on $\mathbb{R}^+$. Therefore it is not surprising that models for survival data often target explicitly $h(x)$. In particular, hazard mixture models define the prior law over the hazard function defining $h(x)$ as mixture of positive double-measurable kernels with a random measure $\tilde{\mu}$ as mixing measure

$$h(x) = \int_{\mathbb{Y}} k(x \mid y) \tilde{\mu}(\mathrm{d}y)$$

This specification was introduced by Dykstra & Laud (1981) with a kernel $k(x \mid y) = \mathbb{1}_{\{0<y\leq x\}}\beta(x)$, where $\beta(x)$ is a measurable non-negative function and $\tilde{\mu}$ is a gamma CRM. Alternative common choices of kernel are

- Rectangular kernel : $k(x \mid y) = \mathbb{1}_{\{|x-y|\leq\tau\}}$, for $\tau > 0$;
- Bathtub or U-shaped kernel: $k(x \mid y) = \mathbb{1}_{\{0<y\leq|x-\tau|\}}$, for $\tau > 0$;
- Ornstein-Uhlenbeck (OU) kernel: $k(x \mid y) = \mathbb{1}_{\{0<y\leq x\}}\kappa \exp\{-\kappa(x-y)\}]$, for $\kappa > 0$;
- Exponential kernel: $k(x \mid y) = s^{-1} \exp\{-x/s\}$, for $s > 0$.

James (2005) provides a posterior characterization of hazard mixture model, when $\tilde{\mu}$ is a CRM. While asymptotic results for this model can be found in Peccati et al. (2008) and De Blasi et al. (2009). More recently, Lijoi & Nipoti (2014) and Camerlenghi et al. (2021) have introduce two generalizations of the mixture hazard model to treat multivariate data coming from many heterogeneous populations. The specification described here for hazard

rates can be used also to model intensities functions of counting processes as will be shown in Chapter 5.
Other well–established and interesting uses of CRMs are neutral to the right processes and priors for cumulative hazards, however we omit them from this review since they are not related with the content of this thesis. We refer the interested reader to Lijoi & Prünster (2010) and references therein.

1.4 Partial exchangeability and dependent priors processes

Even though exchangeability may appear as a relatively weak assumption, we have clarified that it is essentially a homogeneity condition that implies the existence of a common unknown distribution generating the data. Real phenomena often present a level of heterogeneity that makes exchangeability an unrealistic assumption. For instance, collected data may refer to different populations, or be collected under different experimental conditions, or covariate values may be available. Consider, for instance, the case of data referring to the same variable, but that have been collected under two alternative experimental conditions. These data can be conveniently grouped into two samples, such that observations corresponding to the same experimental condition are in the same sample. In this case, if overall exchangeability were assumed, the inherent heterogeinty across samples would be ignored. However, even assuming exchangeability within and independence across sample does not appear optimal, since it means to ignore any possibile connection between different samples. A more realistic assumption is instead partial exchangeability.

Definition 1.12 (Partial Exchangeability)**.** *Let $\boldsymbol{X} = (X_n)_{n\geq 1}$ and $\boldsymbol{Y} = (Y_n)_{n\geq 1}$ be two collections of random variables taking values respectively in the Polish spaces $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$. If $\forall\, n_1 \geq 2$ and $\forall\, n_2 \geq 2$ the law of $(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2})$ is invariant with respect to permutations within each group of random variables, i.e.,*

$$(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}) \stackrel{d}{=} (X_{\sigma_1(1)}, \ldots, X_{\sigma_1(n_1)}, Y_{\sigma_2(1)}, \ldots, Y_{\sigma_2(n_2)})$$

for any σ_1 and σ_2 permutations of respectively $[n_1]$ and $[n_2]$, then the two sequences are said partially exchangeable.

The definition of partial exchangeability can be extended also to more than two sequences (see Chapter 2 and, in particular, Section 2.1). See also Aldous (1985) for a complete discussion of exchangeability and its extensions. From an inferential point of view, partial exchangeability entails that the order of the observations within each sequence is non-informative (i.e. marginal exchangeability for each sequence), while the belonging to a specific sequence is relevant and has to be taken into account.
Also in this case an extension of de Finetti's representation holds true and gives us insights regarding how to model partial exchangeable data. Let $\mathcal{P}_{\mathbb{X}}$ be the space of all probability measures on $\mathbb{X}$ and $\mathcal{P}_{\mathbb{Y}}$ be the space of all probability measures on $\mathbb{Y}$.

Theorem 1.10 (de Finetti, 1938)**.** *Two sequences of random variables* $(X_n)_{n\geq 1}$ *and* $(Y_n)_{n\geq 1}$ *are partially exchangeable iff there exists a probability measure* Q *on* $\mathcal{P}_{\mathbb{X}} \times \mathcal{P}_{\mathbb{Y}}$ *such that,* $\forall\, n_1 \geq 1$ *and* $\forall\, n_2 \geq 1$ *and* $A_1, A_2, \ldots, A_{n_1}, B_1, B_2, \ldots, B_{n_2}$*, with* $A_i \in \mathcal{X}$ *for* $i = 1, \ldots, n_1$ *and* $B_i \in \mathcal{Y}$ *for* $i = 1, \ldots, n_2$

$$\mathbb{P}[X_1 \in A_1, \ldots, X_{n_1} \in A_{n_1}, Y_1 \in B_1, \ldots, Y_{n_2} \in B_{n_2}] = \int_{\mathcal{P}_{\mathbb{X}} \times \mathcal{P}_{\mathbb{Y}}} \prod_{i=1}^{n_1} p_1(A_i) \prod_{i=1}^{n_2} p_2(B_i) Q(\mathrm{d}p_1 \times \mathrm{d}p_2)$$

The theorem can be equivalently be written as

$$\begin{aligned} (X_i, Y_j) \mid \tilde{p}_1, \tilde{p}_2 &\overset{iid}{\sim} \tilde{p}_1 \times \tilde{p}_2 \qquad \text{for } i = 1, \ldots, n_1 \text{ and } j = 1, \ldots, n_2 \\ (\tilde{p}_1, \tilde{p}_2) &\sim Q \end{aligned}$$

where Q is a probability measure on $\mathcal{P}_{\mathbb{X}} \times \mathcal{P}_{\mathbb{Y}}$ that plays the role of the prior and encodes the dependence between $\tilde{p}_1$ and $\tilde{p}_2$ and between X_i and Y_j. When Q has an infinite dimensional support, it is said *dependent nonparametric prior* and $\tilde{p}_1$ and $\tilde{p}_2$ are *dependent processes* (i.e. the main topic of this work).

Starting from the pioneering works of Cifarelli & Regazzini (1978) and of MacEachern (1999, 2000), Bayesian nonparametric contributions for non-exchangeable data have grown substantially in the last two decades, see Dunson (2010), Foti & Williamson (2015), and Müller et al. (2015) for interesting reviews. A large class of dependent nonparametric priors admits an almost-sure discrete representation such that

$$\tilde{p}_i \overset{a.s.}{=} \sum_{k\geq 1} \bar{J}_{k,i} \delta_{\theta_{k,i}}, \qquad i = 1, 2 \tag{1.5}$$

Starting from 1.5 and imposing explicitly a dependence between weights and/or atoms of $(\tilde{p}_1, \tilde{p}_2)$ allows to model jointly the distribution of the two groups: this approach led to dependent Dirichlet processes (DDP) (MacEachern, 1999, 2000; Quintana et al., 2020), dependent stick-breaking processes, kernel stick-breaking processes (Dunson & Park, 2008), probit stick-breaking processes (Rodriguez & Dunson, 2011) and others. Despite their flexibility and the availability of suitable conditional Markov chain Monte Carlo (MCMC) schemes for posterior computations, it is very difficult to derive analytical results for these processes; it is often not clear how dependence of the series reflects at the level of the observables and therefore such methods may lack of transparency. A second popular strategy, often more manageable analytically, consists in working directly on the law of multidimensional vectors of CRMs (Epifani & Lijoi, 2010; Griffin & Leisen, 2017) or in combining (conditionally) independent CRMs, using either additive structures (Müller et al., 2004; Griffin et al., 2013; Lijoi & Nipoti, 2014; Lijoi et al., 2014a,b), nested structures (Rodriguez et al., 2008; Camerlenghi et al., 2019a), or hierarchical structures (Teh et al., 2006; Camerlenghi et al., 2019b). CRMs are then normalized in order to obtain NRMIs.

1.4.1 Completely random vectors based processes

The notion of completely random measure can be extended to the multivariate case through the concept of completely random vector. Here we provide a quick review of bivariate completely random vectors, however all the results can be straightforwardly extended to vectors with a number $d \geq 2$ of entries.

Definition 1.13 (Completely random vector - CRV)**.** *Let $\underline{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2)$ be a vector of CRMs on $\mathbb{X}$. We say that $\underline{\mu}$ is a completely random vector (CRV) if, for every collection of pairwise disjoint sets $(A_i)_{i\geq 1}^n$ in $\mathcal{X}$, the random vectors $(\tilde{\mu}_1(A_1), \tilde{\mu}_2(A_1)), \ldots, (\tilde{\mu}_1(A_n), \tilde{\mu}_2(A_n))$ are mutually independent.*

Considering again the case with no fixed atoms and no deterministic drift, we have a multivariate analogue of the Lévy-Khintchine representation, that reads

$$\mathbb{E}\left[e^{-\tilde{\mu}_1(f_1)-\tilde{\mu}_2(f_2)}\right] = \exp\left\{-\int_{\mathbb{R}_+^2\times\mathbb{X}} (1 - e^{-s_1 f_1(x) - s_2 f_2(x)})\, v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x)\right\} \tag{1.6}$$

for any $f_1, f_2 : \mathbb{X} \to \mathbb{R}^+$ measurable and almost surely integrable functions, where $\tilde{\mu}_j(f_j) = \int_{\mathbb{X}} f_j(x)\tilde{\mu}_j(dx)$ and

$$v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x) = \rho_x(\mathrm{d}s_1, \mathrm{d}s_2)\alpha(\mathrm{d}x)$$

is called *joint Lévy intensity*. As in the univariate case, one may obtain the CRV starting from an underlying PRM N on $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{X}$ with intensity $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x)$ as

$$\begin{pmatrix}\tilde{\mu}_1(\mathrm{d}x)\\ \tilde{\mu}_2(\mathrm{d}x)\end{pmatrix} = \int_{\mathbb{R}^+\times\mathbb{R}^+} \begin{pmatrix}s_1\\ s_2\end{pmatrix} N(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x)$$

In the following we are focusing on homogeneous CRVs, where jumps and locations are independent and the Lévy intensity reads $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x) = \rho(s_1, s_2)\mathrm{d}s_1\mathrm{d}s_2\alpha(\mathrm{d}x)$, where $\rho(s_1, s_2)$ is often named *joint Lévy density*. From the jointy Lévy density, we can retrieve the marginal Lévy density ρ_j of the j-th component of the vector as

$$\rho_1(s) = \int_0^{+\infty} \rho(s, s_2)\mathrm{d}s_2$$

$$\rho_2(s) = \int_0^{+\infty} \rho(s_1, s)\mathrm{d}s_1$$

Finally, the bivariate Laplace exponent reads

$$\psi_b(\lambda_1, \lambda_2) := \int_{\mathbb{R}_+^2\times\mathbb{X}} [1 - e^{-\lambda_1 s_1 - \lambda_2 s_2}] v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x).$$

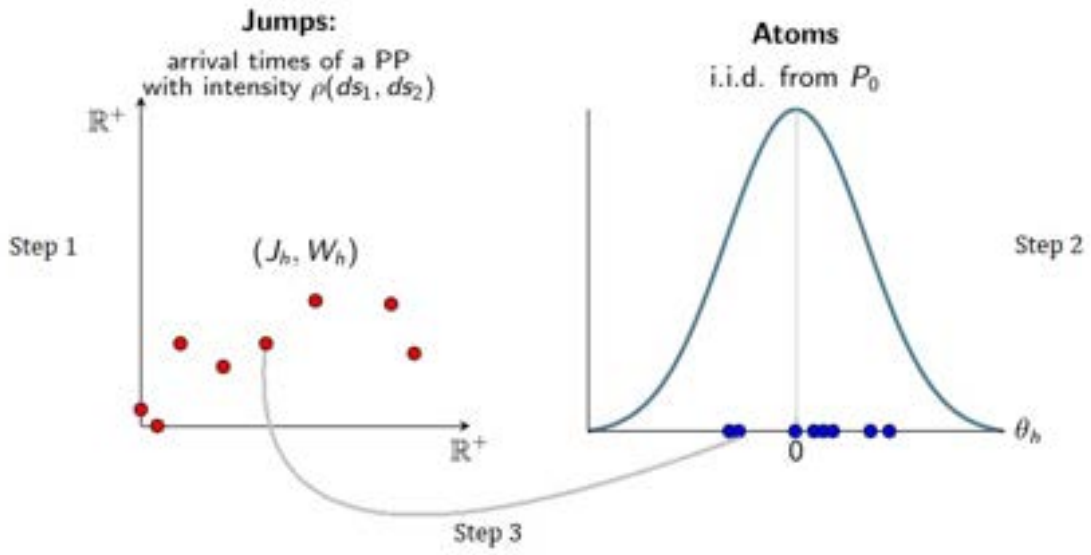


Figure 1.2: Simulation steps for a bivariate homogeneous CRV with intensity $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x) = \rho(\mathrm{d}s_1, \mathrm{d}s_2)\, P_0(\mathrm{d}x)$. At step 1 arrival times of a PP on $\mathbb{R}^+ \times \mathbb{R}^+$ with intensity $\rho(\mathrm{d}s_1, \mathrm{d}s_2)$ are sampled, at step 2 a atom from $P_0(\mathrm{d}x)$ is sampled for each couple of jumps, at step 3 the i.i.d. atoms are associated to the couples of jumps. To get the correspondent NRMIs, it is enough to normalize the two sequences of jumps.

Figure 1.2 summarizes the steps to sample a bivariate homogeneous CRV. For more details on completely random vectors and an interesting account of their dependence structure, we refer to Catalano et al. (2021).
In a way completely similar to that described in Section 1.3.1, the coordinates of a CRV may be normalized in order to obtain two random probability measures. Note that the correspondent normalized measures are as

$$\tilde{p}_1 = \frac{\tilde{\mu}_1(\cdot)}{\tilde{\mu}_1(\mathbb{X})} \overset{a.s.}{=} \sum_{k\geq 1} \bar{J}_k \delta_{\theta_k}, \quad \tilde{p}_2 = \frac{\tilde{\mu}_2(\cdot)}{\tilde{\mu}_2(\mathbb{X})} \overset{a.s.}{=} \sum_{k\geq 1} \bar{W}_k \delta_{\theta_k}.$$

GM-dependent completely random measures

GM-dependent completely random measures and their normalized version have been introduced and studied in Lijoi et al. (2014a); Lijoi & Nipoti (2014) and Lijoi et al. (2014b). The intuitive idea behind these dependent processes is to generate dependent CRMs as sum of common and idiosyncratic components. To construct such processes one may start considering the dependent PRMs proposed in Griffiths & Milne (1978) and then apply Theorem 1.9.

Definition 1.14 (GM-dependent CRMs). *Let $(\tilde{N}_1, \tilde{N}_2)$ be a vector of Griffiths–Milne (GM) de-*

pendent PRMs as in Griffiths & Milne (1978) and define the CRMs $\tilde{\mu}_l(dy) = \int_{\mathbb{R}^+} s\tilde{N}_l(ds, dy)$*, for* $l \in \{1, 2\}$*. Then* $(\tilde{\mu}_1, \tilde{\mu}_2)$ *is said to be a vector of GM-dependent CRMs and we write*

$$(\tilde{\mu}_1, \tilde{\mu}_2) \stackrel{d}{=} \textit{GM-dependent CRMs}$$

Proposition 1.2 (Lijoi et al. 2014a)**.** *If* $(\tilde{\mu}_1, \tilde{\mu}_2) \stackrel{d}{=}$ *GM-dependent CRMs, then there exist* μ_0*,* μ_1*, and* μ_2 *independent CRMs, such that*

$$\tilde{\mu}_1 = \mu_0 + \mu_1$$

and

$$\tilde{\mu}_2 = \mu_0 + \mu_2$$

Moreover if the Lévy intensities of μ_0*,* μ_1*, and* μ_2 *are respectively*

$$\begin{aligned} v_0(ds, dy) &= \theta(1 - z)P_0(dy)\rho(s)ds, \\ v_1(ds, dy) &= \theta\, z\, P_0(dy)\rho(s)ds, \\ v_2(ds, dy) &= \theta\, z\, P_0(dy)\rho(s)ds \end{aligned}$$

then $\tilde{\mu}_1$ *and* $\tilde{\mu}_2$ *are CRMs with marginal Lévy intensities given by* $v(ds, dy) = \theta\, P_0(dy)\rho(s)ds$*.*

Notice that the hyperparameter z controls the dependence between the two CRMs, since when $z = 0$, $\tilde{\mu}_1 \stackrel{a.s.}{=} \tilde{\mu}_2$, while $z = 1$ implies $\tilde{\mu}_1 \perp \tilde{\mu}_2$.

Proposition 1.3 (Lijoi et al. 2014a)**.** *If* $(\tilde{\mu}_1, \tilde{\mu}_2) \stackrel{d}{=}$ *GM-dependent CRMs, the joint Laplace functional transform of* $(\tilde{\mu}_1, \tilde{\mu}_2)$ *is given by*

$$\mathbb{E}[e^{-\tilde{\mu}_1(f_1)-\tilde{\mu}_2(f_2)}] = e^{-\theta\, z\, [\psi(f_1)+\psi(f_2)]-\theta\,(1-z)\,\psi(f_1+f_2)} =: e^{-\theta\,\psi_z(f_1,f_2)}$$

for any $f_1, f_2 : \mathbb{X} \to \mathbb{R}^+$ *measurable and almost surely integrable functions and where* $\theta\,\psi(f) = \int_{\mathbb{R}^+\times\mathbb{X}} [1 - e^{-sf(x)}]v(\mathrm{d}s, \mathrm{d}x)$*.*

Thus, it turns out that GM-dependent CRMs form a CRV, whose Lévy intensity is

$$v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x) = \{z\,[\rho(\mathrm{d}s_1)\delta_0(\mathrm{d}s_2) + \rho(\mathrm{d}s_2)\delta_0(\mathrm{d}s_1)] + (1 - z)\,\rho(\mathrm{d}s_1)\delta_{s_1}(\mathrm{d}s_2)\}\,\theta\,P_0(\mathrm{d}x)$$

and on which usually it is imposed an hyper-prior on z, e.g. a uniform prior on $[0,\, 1]$.

Dependence through Clayton Lévy copula

An elegant alternative strategy to define the joint law of a CRV is provided by Lévy copulae (see Cont & Tankov, 2004; Kallsen & Tankov, 2006; Tankov, 2016, for an extensive treatment of the topic).

Once the marginal behavior of the components of a CRV has been fixed, the dependence may be established using Lévy copulae, analogously to what can be done with copulae for real-valued random variables according to Sklar's theorem (Sklar, 1959).

Definition 1.15 (Lévy copulae). *A function $\mathcal{C} : [0, +\infty]^d \to [0, \infty]$ is a Lévy copula if satisfies the following conditions*

1. *$\mathcal{C}$ is d-non-decreasing, i.e. for every set $A = [l_1, u_1] \times \cdots [l_d, u_d] \subset [0, +\infty)^d$, with $l_j \leq u_j$, for $j = 1, \ldots, d$,*
$$\sum sign(\boldsymbol{v})\mathcal{C}(\boldsymbol{v}) \geq 0$$
where the sum runs all over the vertexes $\boldsymbol{v} = \{v_1, \ldots, v_k\}$ of A and $sign(\boldsymbol{v}) = 1$, if $v_k = l_k$ for an even number of coordinates, while $sign(\boldsymbol{v}) = -1$, otherwise;
2. *if $\boldsymbol{s} = \{s_1, \ldots, s_d\}$ is such that $s_j = 0$ for some j then $\mathcal{C}(\boldsymbol{s}) = 0$;*
3. *$\mathcal{C}(+\infty, \ldots, +\infty, s, +\infty, \ldots, +\infty) = s, \quad \forall s \in [0, +\infty]$.*

The result for Lévy copulae corresponding to Sklar's theorem is the following

Theorem 1.11 (Kallsen & Tankov, 2006). *Let U be the tail integral corresponding to a d-dimensional CRV with Lévy density ρ, i.e.*

$$U(s_1, \ldots, s_d) = \int_{s_1}^{+\infty} \cdots \int_{s_d}^{+\infty} \rho(u_1, \ldots, u_d)\mathrm{d}u_1 \cdots \mathrm{d}u_d$$

and $U_1, \ldots, U_d$ the marginal tail integrals, i.e.

$$U_j(s) = \int_0^{+\infty} \cdots \int_0^{+\infty} \int_s^{+\infty} \int_0^{+\infty} \cdots \int_0^{+\infty} \rho(u_1, \ldots, u_d)\mathrm{d}u_1 \cdots \mathrm{d}u_d = \int_s^{+\infty} \rho_j(u_j)\mathrm{d}u_j \qquad \text{for } j = 1, \ldots, d.$$

Then there exists a d-dimensional copula $\mathcal{C}$ such that for all $\boldsymbol{s} \in [0, +\infty]^d$,

$$U(s_1, \ldots, s_d) = \mathcal{C}(U_1(s_1), \ldots, U_d(s_d))$$

After applying Theorem 1.11, we can recover the multivariate Lévy density in the following way

$$\rho(\mathrm{d}s_1, \ldots, \mathrm{d}s_d) = \frac{\partial}{\partial u_1, \ldots \partial u_d}\mathcal{C}(u_1, \ldots, u_2)\Big|_{u_1 = U_1(s_1), \cdots, u_d = U_d(s_d)} \rho_1(\mathrm{d}s_1) \cdots \rho_d(\mathrm{d}s_d)$$

For example, a bivariate CRV with independent components is obtained with

$$\mathcal{C}(s_1, s_2) = s_1\delta_{+\infty}(s_2) + s_2\delta_{+\infty}(s_1)$$

while maximal dependence between the two components of a bivariate CRV corresponds to the copula

$$\mathcal{C}(s_1,\, s_2) = \min\{s_1, s_2\}$$

A interesting example of Lévy copula is the Clayton Lévy copula, defined as

$$\mathcal{C}_\theta(s_1,\, s_2) = (s_1^{-\theta} + s_2^{-\theta})^{-1/\theta} \qquad \text{for } \theta \in (0, +\infty)$$

The attractive feature of Clayton's copula is that it depends only on one parameter, θ, that fully characterizes the degree of dependence between the resulting CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$ and

$$\lim_{\theta \to 0} \mathcal{C}_\theta(s_1,\, s_2) = s_1\delta_{+\infty}(s_2) + s_2\delta_{+\infty}(s_1),$$

$$\lim_{\theta \to +\infty} \mathcal{C}_\theta(s_1,\, s_2) = \min\{s_1, s_2\}.$$

Compound Random Measures

Another interesting proposal that makes use of CRVs are compound random measures (Griffin & Leisen, 2017).

Definition 1.16 (Compound random measures)**.** *Compound random measures are the coordinates of a CRV defined by a score distribution h and a directing Lévy process with intensity v^* such that the Lévy density of the vector is*

$$\rho(s_1, \ldots, s_d) = \int h(s_1, \ldots, s_d|z)v^*(\mathrm{d}z)$$

where $h(\cdot|z)$ is a probability mass function or a p.d.f.

Compound random measures admit the following series representation

$$\tilde{\mu}_j = \sum_{k=1}^{+\infty} m_{j,k} J_k \delta_{\theta_k}$$

where $m_{j,k} \overset{iid}{\sim} h$ and $\tilde{\eta} = \sum_{k=1}^{+\infty} J_k\delta_{\theta_k}$ is a CRM with Lévy intensity $v^*(\mathrm{d}z)\alpha(\mathrm{d}x)$. The vector of random probability measures defined by normalizing each dimension of the CRV of compound random measures is called normalized compound random measure (NCoRM).

1.4.2 Hierarchical processes

An undoubtedly notable strategy to create dependent processes are hierarchical constructions, according to which dependent processes may be seen as conditionally independent and identically distributed from a certain random probability measure over the space of

probability measures, eg. a DP with random base measure. The most famous example within this class is the hierarchical Dirichlet process (HDP) introduced by Teh et al. (2006)

$$\tilde{p}_j \mid \tilde{p}_0 \overset{iid}{\sim} DP(\theta_j, \tilde{p}_0) \qquad j = 1, \dots, d \qquad \tilde{p}_0 \sim DP(\theta, P_0)$$

Hierarchical CRM

The hierarchical structure has been recently generalized to CRMs by Camerlenghi et al. (2021).

Definition 1.17 (Hierarchical CRMs). *$(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ is said to be a vector of hierarchical CRMs if*

$$\begin{aligned} \tilde{\mu}_j \mid \tilde{\mu}_0 &\overset{ind}{\sim} CRM(v_j) \qquad j = 1, \dots, d \\ \tilde{\mu}_0 &\sim CRM(v_0) \end{aligned}$$

with v_j and v_0 Lévy intensities of the form

$$\begin{aligned} v_j(\mathrm{d}s, \mathrm{d}x) &= \rho_j(s)\, \mathrm{d}s\, \tilde{\mu}_0(\mathrm{d}x) \qquad j = 1, \dots, d \\ v_0(\mathrm{d}s, \mathrm{d}x) &= \rho_0(s)\, \mathrm{d}s\, \theta_0 P_0(\mathrm{d}x) \end{aligned}$$

The joint Laplace transform of a bivariate vector of hierarchical CRMs $(\tilde{\mu}_1, \tilde{\mu}_2)$ is given by

$$\mathbb{E}[e^{-\tilde{\mu}_1(f_1) - \tilde{\mu}_2(f_2)}] = e^{-\theta_0 \int_{\mathbb{X}} \psi^{(0)}[\psi^{(1)}(f_1) + \psi^{(2)}(f_2)]\, P_0(\mathrm{d}x)}$$

where $\psi^{(l)}(f) = \int_{\mathbb{R}^+} [1 - e^{-sf(x)}] \rho_l(s) ds$, for $l \in \{0, 1, 2\}$.
Notice that the vector $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ of hierarchical CRMs is a CRV only conditionally to $\tilde{\mu}_0$, but not marginally.

Hierarchical NRMIs

The hierarchical structure for NRMIs provided in Camerlenghi et al. (2019b) is as in the following definition.

Definition 1.18 (Hierarchical NRMIs). *$(\tilde{p}_1, \dots, \tilde{p}_d)$ is said to be a vector of hierarchical NRMIs if*

$$\begin{aligned} \tilde{p}_j \mid \tilde{p}_0 &\overset{ind}{\sim} NRMI(\rho, \theta, \tilde{p}_0) \qquad j = 1, \dots, d \\ \tilde{p}_0 &\sim NRMI(\rho_0, \theta_0, P_0) \end{aligned}$$

where $\tilde{p} \sim NRMI(\rho, \theta, P_0)$ denotes that $\tilde{p}$ is obtained normalizing a CRM with Lévy intensity $v(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\theta P_0(\mathrm{d}x)$.

Notice that, if one uses gamma CRMs to define hierarchical NRMIs, the resulting process is a HDP. Thus, even if HDPs have been firstly introduced by Teh et al. (2006) exploiting the series representation in (1.5), they admit a representation in terms of NRMIs.

1.4.3 Nested processes

Another well-studied and popular approach to model dependent random probability measures are nested processes, where the random probabilities are again conditionally sampled from a random measure over the space of random probability measures. The Dirichlet version of such processes is the nested Dirichlet process (NDP), has been proposed by Rodriguez et al. (2008) and reads

$$\begin{aligned} \tilde{p}_j \mid \tilde{p}_0 &\overset{iid}{\sim} \tilde{p}_0 \\ \tilde{p}_0 &\sim DP(\theta_0, DP(\theta, P_0)). \end{aligned}$$

The NDP has been extended to the broader class of NRMIs in Camerlenghi et al. (2019a). More details on nested and hierarchical processes can be found in Chapter 2.

Chapter 2

Dependent Species Sampling Processes

In this chapter, we define and study in detail a general class of models, where the observables are obtained by firstly sampling a random partition and then associating independent and identically distributed values to the sets of observations in the partition. We name the priors in this class *dependent species sampling processes*, or *multivariate species sampling processes* (mSSPs), because they are the natural generalization of the species sampling processes (SSPs) of Pitman (1996) to a multivariate setting.
In classical species sampling problems, a random sample $(X_1, \dots, X_n)$ is extracted from a population of subjects and each observed value corresponds to the species of a drawn individual. Denoting with $\tilde{p}$ the unknown distribution of species in the population, clearly we have

$$X_i \mid \tilde{p} \overset{iid}{\sim} \tilde{p} \qquad \text{for } i = 1, \dots, n.$$

Therefore, to develop Bayesian models for species sampling problems, one need to define a prior over the unknown distribution $\tilde{p}$. In the univariate setting, the problem can be tackled relying on the large class of priors introduced by Pitman (1996) as generalization of the Dirichlet process of Ferguson (1973).

Definition 2.1 (Species sampling process - SSP)**.** *A random probability measure $\tilde{p}$ is a species sampling process (SSP) if*

$$\tilde{p} \overset{a.s.}{=} \sum_{h \geq 1} \pi_h \delta_{\theta_h} + \left(1 - \sum_{h \geq 1} \pi_h\right) P_0$$

where the atoms $(\theta_h)_{h \geq 1}$ are i.i.d from the non-atomic distribution P_0, the weights $\boldsymbol{\pi} = (\pi_h)_h$ are such that $\mathbb{P}[0 \leq \pi_h \leq 1] = 1$ for any h, and atoms and weights are independent. Moreover, if $\sum_{h \geq 1} \pi_h \overset{a.s.}{=} 1$, $\tilde{p}$ is said proper.

An interesting extension of the species sampling problem arises when many samples are drawn from multiple populations that may share some species. In this chapter, we study

this extension and the correspondent class of processes. Analogously to species sampling processes, which encompass as special cases the vast majority of priors within the exchangeable framework, also their multivariate version generalizes a great number of priors for partial exchangeability, e.g. single-atoms DDP, hierarchical NRMIs, nested NRMIs, GM-dependent NRMIs, NCoRM, etc. See Section 1.4. Thus, this class provides a unifying point of view to study existing partially exchangeable models, understanding their common features and delineate how to construct new priors within and outside this class. It is important to clarify that, while mSSPs are a natural generalization of SSPs, they are not a trivial one. Indeed, while defining and studying mSSP, we had to carefully consider the dependence induced across populations, which is an aspect completely absent in SSPs.
The structure of the chapter is the following. In Section 2.1, we provide a definition of partial exchangeability, suited for an arbitrary number of populations and that is alternative but equivalent to Definition 1.12. Then, given the central role played by random partitions in mSSP, Section 2.3 is entirely devoted to partially exchangeable partitions. In particular, Section 2.2.1 formally introduces partially exchangeable partitions keeping the sample size n fixed, while Section 2.2.2 extends the framework to collections of random partitions that arise when n increases. Section 2.3 focuses on the law of the observables under a mSSP, while Section 2.4 introduces mSSPs, their moments and a full characterization of their law. Section 2.5 presents a subclass of mSSPs, which we named *regular* and for which an outstanding characterization of dependence in terms of correlation between the processes holds true. Finally, Section 2.6 provides a core algorithm to make inference for any mSSP, based on predictive distributions. It has to be intended as starting point to derive MCMC algorithms for specific dependent processes within the class.

2.1 Partial exchangeability for an arbitrary number of populations

As already mentioned in Section 1.4, when statistical data are sampled from a number J of distinct populations, the homogeneity assumption of *exchangeability* is too restrictive since it does not take into account heterogeneity across populations. On the other hand, the assumption of independence across populations does not allow to borrow information across experiments in the Bayesian learning. A natural compromise between the aforementioned extreme cases is partial exchangeability (de Finetti, 1938), introduced in Section 1.4, that entails exchangeability within but not across different populations, while still allowing for dependence between them.
An alternative way to formalize this framework is to encode population labels in the following way. We assume the number J of distinct populations/groups to be fixed a-priori, as it usually happens in statistical applications, for each sample size n. Moreover, for every n, we define a partition $\mathscr{D}_n = \{D_1, \dots, D_J\}$ of $[n] = \{1, \dots, n\}$, where the D_j are in order of appearance, i.e. that $1 \in D_1$ and the smallest element of $[n] - \bigcup_{i=1}^{j-1} D_i$ is in D_j, for all $j = 2, \dots, J$. $\mathscr{D}_n$ is used to encode the population labels corresponding to the first n obser-

vations, in the sense that $i \in D_j$ means that the i-th observed value comes from population with label j. Notice that D_j will be equal to the empty set if none of the n extracted observations belong to population j. Let us now consider those permutations of the observable that do not change the labels assigned by $\mathcal{D}_n$.

Definition 2.2 ($\mathcal{D}_n$-invariant permutation)**.** *A $\mathcal{D}_n$-invariant permutation is a permutation σ of $[n]$ that, when written in cycle notation, is such that elements in the same cycle do not belong to different sets in the partition $\mathcal{D}_n$, which is equivalent to*

$$\{\sigma(D_1), \dots, \sigma(D_J)\} = \{D_1, \dots, D_J\}$$

Letting the sample size n increase, a sequence $\mathcal{D} = (\mathcal{D}_n)_{n\geq 1}$ of partitions will arise and for every n, $\mathcal{D}_n$ is the partition obtained from $\mathcal{D}_{n+1}$ leaving the element $n+1$ out.
We can provide the definition of partial exchangeability as follows.

Definition 2.3. *A sequence of random variables $\boldsymbol{X} = \{X_i,\, i \geq 1\}$ is partially exchangeable with respect to $\mathcal{D}$ if and only if, for every natural number n and every $\mathcal{D}_n$-invariant permutation σ,*

$$(X_1, \dots, X_n) \overset{d}{=} (X_{\sigma(1)}, \dots, X_{\sigma(n)}).$$

Notice that when $J = 2$, Definition 2.3 is equivalent to Definition 1.12, while when $J = 1$, Definition 2.3 coincides with Definition 1.1. For sake of notation, in the following we add the population labels as superscripts of the Xs'. Thus, we denote with $X_i^{(j_i)}$ the i-th observation in the sample, where j_i is the label of the population from which the i-th observation has been extracted. In the following we are assuming that the sequence $\{X_i^{(j_i)} : i \geq 1 \text{ and } j_i = j\}$ is infinite dimensional for every $j \in [J]$ and that, $\forall j \in [J]$ and $\forall m \in \mathbb{N}$, $\exists n \in \mathbb{N}$ such that $\text{card}(D_j) \geq m$. When this construction is used to model statistical data from J distinct populations, this coincides with assuming that, at least in principle, it is possible to sample an infinite number of times from any population of interest and that $\forall j \in [J]$ and $\forall m \in \mathbb{N}$, it always exists a finite sample size n such that, after observing a total of n observations, at least m of those come from population j. Thanks to de Finetti's theorem, we can characterize the partial exchangeable sequence $\boldsymbol{X} = \{X_i^{(j_i)},\, i \geq 1\}$ as arising from a vector of J dependent random probabilities. More precisely, for every sample size n, it holds

$$\begin{aligned} X_i^{(j_i)} \mid (\tilde{p}_1, \dots, \tilde{p}_J) &\sim \tilde{p}_{j_i} \quad \text{for } i = 1, \dots, n \\ (\tilde{p}_1, \dots, \tilde{p}_J) &\sim \mathcal{L}, \end{aligned} \tag{2.1}$$

where $\mathcal{L}$ takes the role of the prior in the Bayes-Laplace paradigm and controls dependence, thus borrowing of information, across different populations.
There has been an increasing literature devoted to nonparametric models for partially exchangeable data (see Section 1.4), a great number of which correspond to almost-surely discrete priors. See Müller et al. (2018) and Quintana et al. (2020) for recent reviews. The

almost-sure discreteness of the elements in $(\tilde{p}_1, \ldots, \tilde{p}_J)$ induces ties across and within populations, generating a random partition of the observable random variables in $\boldsymbol{X}$. Moreover, the same structure can be recovered in mixture models, when almost-sure discrete probabilities measures are used as mixing measures; in this case a random partition is defined on an underlying non-observable sequence of random variables (cf. Section 1.2.2). For all the instances proposed in the literature, the law of the random partition is a crucial aspect, because it controls dependence across samples and borrowing of strength.

2.2 Partially exchangeable random partitions

2.2.1 Finite partially exchangeable random partitions

Denote with $\mathcal{D}_n = \{D_1, \ldots, D_J\}$ a (non–random) partition of the set of the first n natural numbers $[n] = \{1, \ldots, n\}$, where J is fixed and the sets D_j are in order of appearance and possibly empty. Let I_j be the number of elements in D_j, so that $n = I_1 + \cdots + I_J$. When $\mathcal{D}_n$ is used to encode the population labels of observable data, I_j is the size of the subsample extracted from the population with label j. Consider now $\Pi_n = \{A_1, \ldots, A_K\}$, a random partition of $[n]$, where A_l is not empty for all $l = 1, \ldots, K$ and the A_l are in order of appearance.

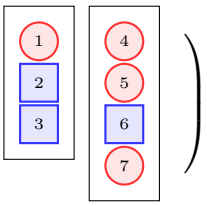
Definition 2.4 (Finite Partially Exchangeable Random Partition)**.** *Π_n is a finite partially exchangeable random partition with respect to $\mathcal{D}_n$ if and only if its distribution is invariant with respect to any $\mathcal{D}_n$-invariant permutation σ, i.e.*

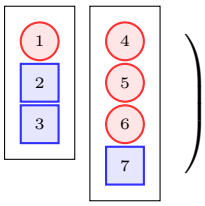
$$\mathbb{P}\left(\Pi_n = \{\sigma(A_1), \ldots, \sigma(A_K)\}\right) = \mathbb{P}\left(\Pi_n = \{A_1, \ldots, A_K\}\right)$$

Denoting with $\sigma(\Pi_n)$ the partition obtained swapping the element of $[n]$ according to σ, clearly by definition we have

$$\Pi_n \stackrel{d}{=} \sigma(\Pi_n)$$

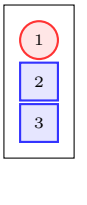
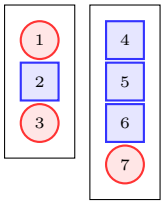
Example 2.1. *Let $n = 7$ and $\mathcal{D}_n = \{\{1, 2, 3\}, \{4, 5, 6, 7\}\}$. Consider a realization of Π_n given by $\Pi_n = \{\{1, 4, 5, 7\}\{2, 3, 6\}\}$ and the $\mathcal{D}_n$-invariant permutation $\sigma = (6, 7)$, if Π_n is a finite partially exchangeable partition we have*

$$\mathbb{P}\left(\Pi_n = \right.$$


$$\left.\right) = \mathbb{P}\left(\Pi_n = \right.$$


$$\left.\right)$$

where the red circles denote that the element belongs to A_1, while the blue squares denotes that the element belongs to A_2.
Consider instead, the $\mathcal{D}_n$-invariant permutation $\sigma = (1, 2)(4, 5)(6, 7)$, if Π_n is a finite partially exchangeable partition, ordering the sets according to the order of appearance, we have

$$\mathbb{P}\left(\Pi_n = \right.$$  $$\left.\right) = \mathbb{P}\left(\Pi_n = \right.$$  $$\left.\right)$$

Denote with $n_{l,j}$ the cardinality of $A_l \cap D_j$ and collect all the cardinalities in a $K \times J$ matrix of counts $\boldsymbol{n}$, whose element at position (l, j) is given by $n_{l,j}$. In the case which $\mathcal{D}_n$ encodes the population labels of observable data, Π_n can usually be interpreted as a latent or observable clustering structure. Thus, $n_{l,j}$ is the number of units from sample j that belong to cluster l. We use $\boldsymbol{n}_j$ to denote the j-th column of $\boldsymbol{n}$.

Lemma 2.1. *If Π_n is a finite partially exchangeable random partition with respect to $\mathcal{D}_n$, then*

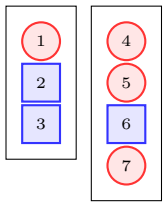
$$\mathbb{P}(\Pi_n = \{A_1, \ldots, A_K\}) = f_n(\boldsymbol{n}) \tag{2.2}$$

Proof. We want to prove that the probability law of a partially exchangeable random partition Π_n is a function only of the matrix of counts $\boldsymbol{n}$. Note that, given $\mathcal{D}_n$, any realization of the partition $\Pi_n = \{A_1, \ldots, A_k\}$ can be deterministically described as function of $\boldsymbol{n}$ and a number J of sequences $S_j = (x_{j,1}, x_{j,2}, \ldots, x_{j,I_j})$, for $j = 1, \ldots, J$. S_j is a sequence containing the elements in D_j, in a specific order, so that $A_1 = \{x_{j,i} : i = 1, \ldots, n_{1,j},\ j = 1, \ldots, J\}$, $A_2 = \{x_{j,i} : i = n_{1,j} + 1, \ldots, n_{1,j} + n_{2,j},\ j = 1, \ldots, J\}$, and so on and so forth. Therefore $\Pi_n = h(\boldsymbol{n}, S_1, \ldots, S_J)$, where h is used to denote a deterministic function. Finally, we have that if $h(\boldsymbol{n}, S_1, \ldots, S_J) = \{A_1, \ldots, A_k\}$, then for every collection of permutations $(\pi_j)_{j=1}^J$, there exists a $\mathcal{D}_n$-invariant permutation σ such that $h(\boldsymbol{n}, \pi_1(S_1), \ldots, \pi_J(S_J)) = \{\sigma(A_1), \ldots, \sigma(A_k)\}$ and since $\mathbb{P}(\{\sigma(A_1), \ldots, \sigma(A_K)\}) = \mathbb{P}(\{A_1, \ldots, A_K\})$ by hypothesis, we have that $\mathbb{P}(\{A_1, \ldots, A_K\}) = f(\boldsymbol{n})$. □

Notice that if a $\mathcal{D}_n$-invariant permutation is applied to the elements of $\{A_1, \ldots, A_K\}$, such permutation may at most induce permutations of the rows of $\boldsymbol{n}$. To clarify this point we provide an example.

Example 2.2. *Let $n = 7$ and $\mathcal{D}_n = \{\{1, 2, 3\}, \{4, 5, 6, 7\}\}$. Consider a realization of Π_n given by $\Pi_n = \{\{1, 4, 5, 7\}\{2, 3, 6\}\}$, then the resulting matrix $\boldsymbol{n}$ is*

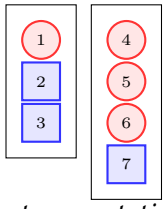
$$\boldsymbol{n} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$$

and therefore $\mathbb{P}\left(\Pi_n = \right.$  $\left.\right) = f_n\left(\mathbf{n}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{n}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}\right)$.

Consider the $\mathcal{D}_n$-invariant permutation $\sigma_1 = (2,\ 3)(4,\ 5,\ 6,\ 7)$. The resulting partition is $\sigma(\Pi_n) =$

$\{\{1,4,5,6\},\{2,3,7\}\}$ *and the resulting matrix of counts is*

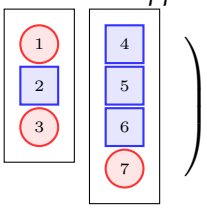
$$\boldsymbol{n}^{(\sigma_1)} = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$$

and therefore $\mathbb{P}\left(\Pi_n = \right.$  $\left.\right) = f_n\left(\mathbf{n}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{n}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}\right)$.

Consider the $\mathscr{D}_n$*-invariant permutation* $\sigma_2 = (1,\,2,\,3)(4,\,5,\,6,\,7)$*. The resulting partition (ordering the sets according to the order of appearance) is* $\sigma(\Pi_n) = \{\{1,3,7\},\{2,4,5,6\}\}$ *and the resulting matrix of counts is*

$$\boldsymbol{n}^{(\sigma_2)} = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

where the two rows have been swapped.

Therefore $\mathbb{P}\left(\Pi_n = \right.$  $\left.\right) = f_n\left(\mathbf{n}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{n}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$.

It turns out that all the functions f_n satisfying certain conditions are the law of a finite partially exchangeable random partition. However, before introducing such conditions and prove the characterization, it is important to clarify a technical aspect about the support of the matrix of counts $\boldsymbol{n}$. It is straightforward to see that, given the partition $\mathscr{D}_n$ and fixing K, $\boldsymbol{n}$ satisfies the following

(n-i) $\boldsymbol{n} \in \mathbb{N}_0^{K\times J}$, where $\mathbb{N}_0 = \{0,1,2\ldots\}$;

(n-ii) $\sum_{l=1}^{K} n_{l,j} = I_j$;

(n-iii) $\sum_{j=1}^{J} n_{l,j} > 0$.

Therefore we may define

$$\rho_K(I_1,\ldots,I_J) = \{\boldsymbol{n} : \boldsymbol{n} \in \mathbb{N}_0^{K\times J},\ \sum_{l=1}^{K} n_{l,j} = I_j,\ \ \sum_{j=1}^{J} n_{l,j} > 0 \text{ for } l = 1,\ldots,K; j = 1,\ldots,J\}$$

However, not all matrices in $\rho_K(I_1,\ldots,I_J)$ correspond to a partition (in order of appearance), when such partition exists we say that $\boldsymbol{n}$ is a *compatible matrix of counts* according to $\mathscr{D}_n$ (or, shortly, $\boldsymbol{n}$ is $\mathscr{D}_n$-compatible). To clarify why not all the matrices in $\rho_K(I_1,\ldots,I_J)$ correspond to a partition, we provide the following two examples.

Example 2.3. *Let $n = 7$, $D_n = \{\{1,2,3\},\{4,5,6,7\}\}$ and K=2. The matrix*

$$\boldsymbol{n} = \begin{pmatrix} 0 & 3 \\ 3 & 1 \end{pmatrix}$$

is not a compatible matrix of counts according to $\mathscr{D}_n$. The order of appearance requires $n_{1,1} > 0$.

Example 2.4. *Let $n = 7$, $D_n = \{\{1,3,4\},\{2,5,6,7\}\}$ and K=3. The matrix*

$$\boldsymbol{n} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 4 \end{pmatrix}$$

is not a compatible matrix of counts according to $\mathscr{D}_n$. The order of appearance requires $n_{1,2}+n_{2,2} > 0$.

Finally, we denote with

$$\rho_K^*(I_1,\dots,I_J) = \{\boldsymbol{n} : \boldsymbol{n} \in \rho_K(I_1,\dots,I_J),\ \boldsymbol{n} \text{ is } \mathscr{D}_n\text{-compatible}\}$$

and

$$\bar{\rho}_n^*(I_1,\dots,I_J) = \bigcup_{K=1}^{n} \rho_K^*(I_1,\dots,I_J).$$

Since the support of $\boldsymbol{n}$ depends on $\mathscr{D}_n$, the detailed and correct notation for the matrix of counts should be $\boldsymbol{n}_{\mathscr{D}_n}$, but for sake of notation we omit the subscript.

Proposition 2.1. *A random partition Π_n is a finite partially exchangeable random partition with respect to $\mathscr{D}_n$ if and only if there exists a function f_n such that*

$$\mathbb{P}(\Pi_n = \{A_1,\dots,A_K\}) = f_n(\boldsymbol{n}) \tag{2.3}$$

where f_n satisfies the three following conditions:

- *(ff-i)* $f_n : \bar{\rho}_n^*(I_1,\dots,I_J) \to [0,1]$
- *(ff-ii)* $\sum f_n(\boldsymbol{n}) = 1$, *where the sum runs all over the space $\mathscr{P}([n])$ of partitions (in order of appearance) of $[n]$.*
- *(ff-iii)* $f_n((\boldsymbol{n}_1,\dots,\boldsymbol{n}_J)) = f_n((\alpha(\boldsymbol{n}_1),\dots,\alpha(\boldsymbol{n}_J)))$, *for every α permutation of K elements that generates a compatible matrix of counts according to $\mathscr{D}_n$.*

Proof. The *if* part can be proved noticing that conditions (ff-i) and (ff-ii) ensure that f_n encodes a probability measure over the space of partitions of $[n]$ and thus imply the existence of a random partion Π_n. Moreover by condition (ff-iii), we have that for every $\mathscr{D}_n$-invariant permutation σ

$$\mathbb{P}(\{\sigma(A_1),\dots,\sigma(A_K)\}) = \mathbb{P}\left(\{A_1,\dots,A_K\}\right)$$

The *only if* part is proved noticing that (2.3) holds true by Lemma 2.1 and being Π_n a random partition of $[n]$, conditions (ff-i) and (ff-ii) are trivially satisfied. Moreover, since Π_n is partially exchangeable with respect to $\mathcal{D}_n$, for each α permutation of K elements that generates a compatible matrix of counts according to $\mathcal{D}_n$ and each compatible matrix of counts $\boldsymbol{n}$, there exists a $\mathcal{D}_n$-invariant permutation of σ and a realization $\{A_1, \ldots, A_K\}$ of Π_n such that

$$\mathbb{P}(\{\sigma(A_1), \ldots, \sigma(A_K)\}) = f_n((\alpha(\boldsymbol{n}_1), \ldots, \alpha(\boldsymbol{n}_J))).$$

□

2.2.2 Infinite partially exchangeable random partitions

When random partitions are used to model statistical data, usually one admits that new observations may be collected, thus what is more interesting is the study of the sequence of random partitions which is obtained letting n vary. When considering a whole sequence of partitions, we need to require a condition of coherence.

Definition 2.5 (Coherent sequence of partitions)**.** *A sequence $\mathcal{D} = (\mathcal{D}_n)_{n\geq 1}$ of partitions of [n] is said **coherent** if, for every n, $\mathcal{D}_n$ is the partition obtained from $\mathcal{D}_{n+1}$ leaving the element $n+1$ out.*

Denote with $\mathcal{D} = (\mathcal{D}_n)_{n\geq 1}$ a coherent sequence of partitions of $[n]$. Consider an almost surely coherent sequence $\Pi = (\Pi_n)_{n\geq 1}$ of random partitions, such that Π_n is a random partition of $[n]$.

Definition 2.6 (Infinite Partially Exchangeable Random Partition)**.** Π *is an (infinite) partially exchangeable random partition with respect to $\mathcal{D}$ if and only if Π_n is a finite partially exchangeable random partition with respect to $\mathcal{D}_n$, for every $n > 0$, which is*

$$\mathbb{P}\left(\Pi_n = \{\sigma(A_1), \ldots, \sigma(A_K)\}\right) = \mathbb{P}\left(\Pi_n = \{A_1, \ldots, A_K\}\right)$$

for every $\mathcal{D}_n$-invariant permutation σ, for every n.

Proposition 2.2. Π *is an infinite partially exchangeable random partition with respect to $\mathcal{D}$ if and only if there exists a function f such that*

$$\mathbb{P}(\Pi_n = \{A_1, \ldots, A_K\}) = f(\boldsymbol{n}) \qquad \forall\, n \geq 1 \tag{2.4}$$

where f is a function satisfying the three following conditions:

(f-i) $f : \bigcup_{n=1}^{+\infty} \bar{\rho}_n^*(I_1, \ldots, I_J) \to [0, 1]$,

(f-ii) $f(1) = 1$ *and* $f(\boldsymbol{n}) = \sum_{l=1}^{K+1} f(\boldsymbol{n}^{lj+})$

where j is the index of the set in $\mathcal{D}_{n+1}$ that contains the element $n+1$, $\boldsymbol{n}^{lj+}$ denotes the matrix

whose entries are equal to those of $\boldsymbol{n}$ *except the* (l, j)*-th element that has been increased by 1. Clearly,* $\boldsymbol{n}^{(K+1)j+}$ *has one row more than* $\boldsymbol{n}$*.*

(f-iii) $f((\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J)) = f((\alpha(\boldsymbol{n}_1), \ldots, \alpha(\boldsymbol{n}_J)))$*,*
for every α *permutation of* K *elements that generates a* $\mathcal{D}_n$*-compatible matrix of counts.*

We call the function f ***partially exchangeable partition probability function*** *(pEPPF).*

Proof. First of all, we prove that, given a coherent sequence of partitions $\mathcal{D}$ of $[n]$, for any function f satisfying conditions (f-i)-(f-iii) there exists Π infinite partially exchangeable random partition with respect to $\mathcal{D}$. Notice that condition (f-ii) implies condition (ff-ii) in Proposition 2.1, so that for every fixed n, there exists Π_n a finite partially exchangeable random partition with respect to $\mathcal{D}_n$. As for the almost surely coherence of $\Pi = (\Pi_n)_{n\geq 1}$, notice that condition (f-ii) ensures that the marginal law of Π_{n+1} gives positive probability to those partitions coherent to Π_n, for every Π_n such that $\mathbb{P}[\Pi_n] > 0$ and by Ionescu-Tulcea extension theorem there exists an almost surely coherent sequence Π whose marginals at every n are provided by the function f. The *only if* part follows directly from the proof of Proposition 2.1. □

Note that if $\mathcal{D}_n$ contains only a set equal to $[n]$ for every $n \geq 1$, than the definitions of partially exchangeable partition and pEPPF coincide with the definitions of exchangeable partition and exchangeable partition probability function (EPPF) of Pitman (1996). We conclude this section proving that for every partially exchangeable partition probability function it can always be constructed a corresponding partially exchangeable sequence of random variables.

Theorem 2.1. *Given a coherent sequence of partitions* $\mathcal{D}$ *of* $[n]$*, for any function* f *satisfying conditions (f-i)-(f-iii), there exists a partially exchangeable sequence* $\boldsymbol{X}$ *whose partition* Π*, defined by the random equivalence relation* $i \sim i'$ *iff* $X_i = X'_i$*, is a partially exchangeable random partition with respect to* $\mathcal{D}$ *and whose law is controlled by* f*. Vice versa, for any partially exchangeable sequence* $\boldsymbol{X}$ *with respect to* $\mathcal{D}$*, the random partition* Π*, define by the random equivalence relation* $i \sim i'$ *iff* $X_i = X'_i$*, is a partially exchangeable random partition with respect to* $\mathcal{D}$ *.*

Proof. We want to show that for every partially exchangeable random partition Π, there exists a partially exchangeable sequence $\boldsymbol{X}$, that induces Π through the equivalence relation $i \sim i'$ iff $X_i = X'_i$. Before doing so, we recall that an infinite sequence $\boldsymbol{X}$ is partially exchangeable with respect to a coherent sequence of partitions $\mathcal{D}$ of $[n]$ if and only if, for every $n \geq 1$ and any $\mathcal{D}_n$-invariant permutation σ

$$(X_1, \ldots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)}).$$

Now let $\theta_1, \theta_2, \ldots$ be an i.i.d. sample from a diffuse distribution H, such that $\theta_i \neq \theta_j$ with probability 1. Define then $X_i = \theta_{g(i)}$, where the function $g : \mathbb{N} \to \mathbb{N}$ is such that for

each value i returns the label of set that contains i in the partition Π_i. So that Π is the infinite exchangeable partition induced by $\boldsymbol{X}$. Lastly, we have to prove that $\boldsymbol{X}$ is partially exchangeable. Notice that the law of $(X_1, \dots, X_n)$ is completely characterized by the laws of $(\theta_1, \dots, \theta_n)$ and (Π_n), which are both invariant under any $\mathcal{D}_n$-invariant permutation σ. Indeed, $(\theta_1, \dots, \theta_n) \stackrel{d}{=} (\theta_{\sigma(1)}, \dots, \theta_{\sigma(n)})$ and, by definition of partially exchangeable random partition, $\Pi_n \stackrel{d}{=} \sigma(\Pi_n)$. To prove the last part of the theorem, it is enough to notice that the partition is a deterministic function of the sequence, thus its law has to preserve the invariance of the sequence to $\mathcal{D}_n$-invariant permutations. □

2.3 Multivariate species sampling model

In (univariate) species sampling problems, a random sample $(X_1, \dots, X_n)$ is drawn from a population of individuals and each observed value corresponds to the species of a drawn individual. An extension of this problem arises when the sample is drawn from multiple populations that may share some species. In this case, a population label is associated to each observation: $X_i^{(j_i)}$ is i-th observation whose population label is j_i. The partition $\mathcal{D}_n$ of $[n]$, introduced in the previous sections, is used to keep track of the population from which each observation has been sampled, so $i \in D_j$ if and only if $j_i = j$. Moreover, we denote with X_l^* the l-th species that appears during sampling and with $n_{l,j}$ the number of observed individuals extracted from population j that belong to the species X_l^*.
The definitions of species sampling model and sequence of Pitman (1996) can be generalized to multiple populations as follows.

Definition 2.7 (Multivariate species sampling model)**.** *We say that $\boldsymbol{X}$ follows a multivariate species sampling model (mSSM), if there exists an infinite partially exchangeable random partition $\Pi = (\Pi_n)_{n\geq 1}$ with respect to $\mathcal{D}$ such that, for every n, n-dimensional vectors of elements in $\boldsymbol{X}$ can be generated by*

1. *sampling Π_n (via pEPPF)*
2. *sampling independent and identically distributed unique values $\{X_1^*, \dots, X_K^*\}$, with X_l^* sampled from a non-atomic measure H (independently from Π_n), for $l = 1, \dots, K$.*

Note that it follows trivially by definition that a sequence following a mSSM is partially exchangeable with respect to $\mathcal{D}$.

Definition 2.8 (Multivariate species sampling sequence)**.** *We say that $\boldsymbol{X}$ is a multivariate species sampling sequence (mSSS) with respect to $\mathcal{D}$ if the predictive distribution of $\boldsymbol{X}$ is given by*

$$\begin{aligned} X_1^{(j_1)} &\sim H \\ X_{n+1}^{(j_{n+1})} \mid \boldsymbol{X}_{1:n} &\sim \sum_{l=1}^{K} p_{j_{n+1},l}(\boldsymbol{n})\, \delta_{X_l^*} + p_{j_{n+1},K+1}(\boldsymbol{n})\, H \end{aligned} \tag{2.5}$$

where K is the number of different species/unique values observed in the sample $\boldsymbol{X}_{1:n} = \{X_i^{(j_i)} : i = 1, \dots, n\}$, $\boldsymbol{n} = (n_{l,j} : l = 1, \dots, K, j = 1, \dots, J)$ is the matrix of counts before observing $X_{n+1}^{(j_{n+1})}$ and H is a non-atomic distribution. While the collection of functions $(p_{j,l}, l = 1, 2, \dots)_{j=1}^{J}$ are such that:

(p-i) $p_{j,l}(\boldsymbol{n}) \geq 0$,

(p-ii) $\sum_{l=1}^{K+1} p_{j,l}(\boldsymbol{n}) = 1, \forall\, \boldsymbol{n}$ *and* $\forall j = 1, \dots J$,

(p-iii) $p_{j,l}(\boldsymbol{n}) p_{j',r}(\boldsymbol{n}^{lj+}) = p_{j',r}(\boldsymbol{n}) p_{j,l}(\boldsymbol{n}^{rj'+})$, $\forall j, j' \in \{1, \dots, J\}$ *and* $\forall l, r$,

(p-iv) $p_{j,l}((\boldsymbol{n}_1, \dots, \boldsymbol{n}_J)) = p_{j,\alpha^{-1}(l)}((\alpha(\boldsymbol{n}_1), \dots, \alpha(\boldsymbol{n}_J)))$, *for every α permutation of K elements that generates a $\mathscr{D}_n$-compatible matrix of counts.*

The collection of functions $p_{j,l}$ is called ***multivariate prediction probability function*** *(mPPF).*

Notice that conditions (p-iii) and (p-iv) are generalization to the partial exchangeable setting of the conditions provided in (Lee et al., 2013). Conditions (p-iii) and (p-iv) guarantee that the joint distribution defined by (2.5) is invariant with respect to $\mathscr{D}_n$-invariant permutations, for every n, i.e. they ensure partial exchangeability. Condition (p-iii) may be thought as invariance for *future* observations, thus, when $j_{n+1} = j_{n+2}$, it implies

$$\mathbb{P}(X_{n+1}^{(j_{n+1})} = X_l^*, X_{n+2}^{(j_{n+2})} = X_r^* \mid \boldsymbol{X}_{1:n}) = \mathbb{P}(X_{n+1}^{(j_{n+1})} = X_r^*, X_{n+2}^{(j_{n+2})} = X_l^* \mid \boldsymbol{X}_{1:n})$$

Condition (p-iii) guarantees that the joint distribution is invariant with respect to all $\mathscr{D}_n$-invariant permutations that do not induce any change on the matrix of counts $\boldsymbol{n}$ (cf. Example 2.2). This is a consequence of the fact that any permutation can be expressed as product of transpositions. Condition (p-iv) instead may be seen as the invariance condition for *past* observations because it implies that

$$\mathbb{P}[X_{n+1}^{(j_{n+1})} = X_l^* \mid \boldsymbol{X}_{1:n}] = \mathbb{P}[X_{n+1}^{(j_{n+1})} = X_l^* \mid \sigma(\boldsymbol{X}_{1:n})]$$

for any $\mathscr{D}_n$-invariant permutation σ. Ultimately it guarantees the invariance also for those permutations that cause row swapping (cf. Example 2.2). Next theorem shows the equivalence between Definition 2.7 and Definition 2.8 above and has as trivial corollary partial exchangeability of any mSSS.

Theorem 2.2. *$\boldsymbol{X}$ follows a multivariate species sampling model if and only if $\boldsymbol{X}$ is a multivariate species sampling sequence, i.e. Definitions 2.7 and 2.8 are equivalent.*

Proof. To prove the *only if* part, let $\boldsymbol{X}$ be a mSSM with pEPPF f, setting n equal 1, one trivially obtain $X_{1,1} \sim H$. For $n > 1$, define

$$p_{j,l}(\boldsymbol{n}) = \frac{f(\boldsymbol{n}^{lj+})}{f(\boldsymbol{n})} \qquad \forall\, \boldsymbol{n},\ \ l = 1, \dots, k+1 \text{ and } j = 1, \dots, J. \tag{2.6}$$

Therefore, since f returns the probability of the partition, we have that

$$p_{j,l}(\boldsymbol{n}) = \mathbb{P}\left[X_{n+1}^{(j_{n+1})} = X_l^* \mid \boldsymbol{X}_{1:n}\right]$$

Thus, the predictive distribution of $\boldsymbol{X}$ coincides with the expression in (2.5) and the collection of functions $p_{j,l}$ trivially satisfy conditions (p-i) and (p-ii). Moreover, condition (p-iii) can be obtained computing

$$p_{j,l}(\boldsymbol{n})p_{j',r}(\boldsymbol{n}^{lj+}) = \frac{f(\boldsymbol{n}^{lj+})}{f(\boldsymbol{n})}\frac{f((\boldsymbol{n}^{lj+})^{rj'+})}{f(\boldsymbol{n}^{lj+})} = \frac{f((\boldsymbol{n}^{lj+})^{rj'+})}{f(\boldsymbol{n})} = p_{j',r}(\boldsymbol{n})p_{j,l}(\boldsymbol{n}^{rj'+})$$

while (p-iv) follows combining (2.6) and (f-iii) in Proposition 2.2.
Let us assume now that $\boldsymbol{X}$ is a multivariate species sampling sequence and consider the partitions Π_n defined by the equivalence relation $i \sim i'$ iff $X_i^{(j_i)} = X_{i'}^{(j_{i'})}$, for any $X_i^{(j_i)}$ and $X_{i'}^{(j_{i'})} \in \boldsymbol{X}$. The *if* part is proved if one is able to show that the law of such partition is invariant with respect to any $\mathscr{D}_n$-invariant permutation, for every n. Consider a realization $\Pi_n = \{A_1, \ldots, A_K\}$ and denote with Π_i the partition obtained by Π_n considering only the first i observations and with $\boldsymbol{n}^{(i)}$ the matrix of counts corresponding to Π_i, by (2.5), we have that

$$\mathbb{P}\left(\Pi_n = \{A_1, \ldots, A_K\}\right) = \prod_{i=2}^{n} p_{j_i,l_i}(\boldsymbol{n}^{(i-1)})$$

where j_i and l_i are the population and species label of the i-th observation. For any $\mathscr{D}_n$ invariant permutation σ, we get

$$\mathbb{P}\left(\Pi_n = \{\sigma(A_1), \ldots, \sigma(A_K)\}\right) = \prod_{i=2}^{n} p_{j_i,\alpha^{-1}(l_{\sigma(i)})}\left(\alpha(\boldsymbol{n}_1^{(\sigma(i)-1)}), \ldots, \alpha(\boldsymbol{n}_J^{(\sigma(i)-1)})\right)$$

By applying (p-iii) and, then, (p-iv) we have

$$\begin{aligned}\mathbb{P}\left(\Pi_n = \{\sigma(A_1), \ldots, \sigma(A_K)\}\right) &= \prod_{i=2}^{n} p_{j_i,\alpha^{-1}(l_i)}\left(\alpha(\boldsymbol{n}_1^{(i-1)}), \ldots, \alpha(\boldsymbol{n}_J^{(i-1)})\right) \\ &= \prod_{i=2}^{n} p_{j_i,l_i}\left(\boldsymbol{n}_1^{(i-1)}, \ldots, \boldsymbol{n}_J^{(i-1)}\right)\end{aligned}$$

which completes the proof. □

As one would expect, if we consider just a subset of populations in $\boldsymbol{X}$, the resulting sequence is still a multivariate species sampling sequence as clarified by the following two propositions.

Proposition 2.3. *Defining $\boldsymbol{X}_j = \{X_i^{(j_i)} \in \boldsymbol{X} : j_i = j\}$, if $\boldsymbol{X}$ is a multivariate species sampling sequence, then marginally $\boldsymbol{X}^{(-j)} = \{\boldsymbol{X}_1, \ldots \boldsymbol{X}_{j-1}, \boldsymbol{X}_{j+1}, \ldots, \boldsymbol{X}_J\}$ is a multivariate species*

sampling sequence.

Proof. The proof follows from Definition 2.7 (or equivalently from Definition 2.8) after a marginalization over the observations from population j. □

Proposition 2.4. *If $\boldsymbol{X}$ is a multivariate species sampling sequence, then marginally $\boldsymbol{X}_j$ is a species sampling sequence.*

Proof. The proof follows from Definition 2.7 (or equivalently from Definition 2.8) after a marginalization over the observations in populations with labels $j' \neq j$ and exploiting the results in Pitman (1996). □

Notice that, on the one hand, the exchangeable case can always be recovered within our framework, setting $J = 1$, and therefore all the following results hold also in the exchangeable case. On the other hand, the basic idea underlying Definition 2.7 is even more general and can be used to define further generalizations, even beyond partial exchangeability, as long as they correspond to a generative process such that observations can be sampled in two steps

1. sample a partition (even not partially exchangeable);
2. given the partition, sample the unique values independently from a non-atomic measure.

where two conditions have been relaxed with respect to Definition 2.7: partial exchangeability of the partition and independence between the partition and the unique values. More details on this can be found in Chapter 6.

2.3.1 Correlation between observables

Typically, when dealing with partial exchangeable sequences, it is of interest to compute the correlation between two observations (both within and across samples), because such correlation provides insights on the Bayesian learning mechanism induced by the model. It turns out that in all mSSMs the correlation is equal to the pEPPF for $n = 2$ observations, as shown in the proposition here below.

Proposition 2.5. *If $\boldsymbol{X} \sim$ mSSM,*

$$Corr(X_i^{(j_i)}, X_l^{(j_l)}) = \mathbb{P}(X_i^{(j_i)} = X_l^{(j_l)}) = \mathbb{P}(X_i^{(j_i)} = X_l^{(j_l)} \mid X_l^{(j_l)})$$

Proof. Define the random variable Z, so that $Z = 1$, if $X_i^{(j_i)} = X_{i'}^{(j_{i'})}$, and $Z = 0$, otherwise.

The first equality follows from

$$\begin{aligned}\text{Cov}(X_i^{(j_i)}, X_{i'}^{(j_{i'})}) &= \mathbb{E}\left[\text{Cov}(X_i^{(j_i)}, X_{i'}^{(j_{i'})} \mid Z)\right] + \text{Cov}\left(\mathbb{E}\left[X_i^{(j_i)} \mid Z\right], \mathbb{E}\left[X_{i'}^{(j_{i'})} \mid Z\right]\right) \\ &= \mathbb{E}\left[\text{Cov}(X_i^{(j_i)}, X_{i'}^{(j_{i'})} \mid Z)\right] + 0 \\ &= \text{Cov}(X_i^{(j_i)}, X_{i'}^{(j_{i'})} \mid Z = 1)\mathbb{P}(X_i^{(j_i)} = X_{i'}^{(j_{i'})}) + 0 \\ &= \mathbb{P}(X_i^{(j_i)} = X_{i'}^{(j_{i'})})\text{Var}(X^*).\end{aligned}$$

where $X^* \sim H$ and the last equality follows by the independence between the partition and the unique values. □

Therefore in multivariate species sampling models the correlation between any couple of observations (even those drawn from different populations) is always non-negative. However, positive correlation between observations across samples is not implied by partial exchangeability (see Chapter 3). This result tells us also that all dependence and borrowing of strength reside in the sharing of the underlying common atoms. Notice that to prove Proposition 2.5, the partial exchangeability property of the partition is actually not needed, the only requirement to obtain the first equality is the independence of the unique values sampled from H, while to get the second equality one also need independence between the unique values and the partition.

2.4 Multivariate species sampling process

As already mentioned, multivariate species sampling models generate sequences of observations that are partially exchangeable. Thus, by de Finetti's theorem, there exists a vector of underlying random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_J)$, from which the observed data can be seen as independent random samples.

Definition 2.9 (Multivariate species sampling process 1)**.** *When $\boldsymbol{X}$ is distributed according to a multivariate species sampling model, the associated vector of random probabilities $(\tilde{p}_1, \ldots, \tilde{p}_J)$ in (2.1) is a multivariate species sampling process (mSSP).*

Multivariate species sampling processes are a generalization of the species sampling processes of Pitman (1996) as underlined by the following corollary.

Corollary 2.1. *If $(\tilde{p}_1, \ldots, \tilde{p}_J) \sim$ mSSP then marginally $\tilde{p}_j \sim$ SSP.*

Proof. The proof follows from Definition 2.7 (or equivalently from Definition 2.8) after a marginalization over the observations in populations with labels $j' \neq j$ and exploiting the results in Pitman (1996). □

To move from univariate species sampling processes to their multivariate version the crucial aspect to study is the dependence between the elements in the vector $(\tilde{p}_1, \ldots, \tilde{p}_J)$. In

order to do so, we provide in the following both marginal and mixed moments as well as a full characterization of the joint law of the processes. These results can also be seen as generalizations to the general classes of SSP and mSSP of the results of joint moments of normalized completely random measures in the seminal work by James et al. (2006) and, more recently, of hierarchical normalized completely random measures in Argiento et al. (2020).

Proposition 2.6. *If $\boldsymbol{X}$ follows a mSSM with associated $(\tilde{p}_1, \dots, \tilde{p}_J) \sim$ mSSP (for every measurable A such that $0 < H(A) < 1$),*

$$Var\{\tilde{p}_j(A)\} = \mathbb{P}(X_i^{(j)} = X_l^{(j)})H(A)\{1 - H(A)\}$$

for any i and l such that $i \neq l$ and $X_i^{(j_i)}$ and $X_l^{(j_l)}$ come from population with label j, i.e. $j_i = j_l = j$.

Proof.

$$\mathbb{E}\{\tilde{p}_j(A)^2\} = \mathbb{P}(X_i^{(j)} \in A, X_l^{(j)} \in A).$$

Then we disintegrate with respect to $\{X_i^{(j)} = X_l^{(j)}\}$ to recover independence.

$$\begin{aligned}\mathbb{P}(X_i^{(j)} \in A, X_l^{(j)} \in A) = \mathbb{P}(X_i^{(j)} = X_l^{(j)})\mathbb{P}(X_i^{(j)} \in A, X_l^{(j)} \in A \mid X_i^{(j)} = X_l^{(j)}) + \\ + \mathbb{P}(X_i^{(j)} \neq X_l^{(j)})\mathbb{P}(X_i^{(j)} \in A, X_l^{(j)} \in A \mid X_i^{(j)} \neq X_l^{(j)}) = \\ = \mathbb{P}(X_i^{(j)} = X_l^{(j)})H(A) + \mathbb{P}(X_i^{(j)} \neq X_l^{(j)})H(A)^2.\end{aligned}$$

Finally, $\text{Var}\{\tilde{p}_j(A)\} = \mathbb{E}\{\tilde{p}_j(A)^2\} - \mathbb{E}\{\tilde{p}_j(A)\}^2 = \mathbb{P}(X_i^{(j)} = X_l^{(j)})H(A)\{1 - H(A)\}$. □

Therefore the variance can be expressed through two multiplicative terms, one being the probability of a tie within the population (which depends on the pEPPF only), the other corresponding to the variance of the random variable $\mathbb{1}_A(X_i^{(j)})$, which depends on $H(A)$ only. Note that

$$\mathbb{P}(\mathbb{1}_A(X_i^{(j)}) = 1) = \mathbb{P}(X_i^{(j)} \in A) = \mathbb{P}(X_i^{(j)} \in A, X_l^{(j)} \in A \mid X_i^{(j)} = X_l^{(j)}) = H(A)$$

2.4.1 Correlation between multivariate species sampling processes

Proposition 2.7. *If $\boldsymbol{X}$ follows a mSSM with associated $(\tilde{p}_1, \dots, \tilde{p}_J) \sim$ mSSP (for every measurable A such that $0 < H(A) < 1$),*

$$Corr\{\tilde{p}_j(A), \tilde{p}_k(A)\} = \frac{Corr(X_i^{(j)}, X_m^{(k)})}{\sqrt{Corr(X_i^{(j)}, X_l^{(j)})}\sqrt{Corr(X_m^{(k)}, X_n^{(k)})}}$$

and

$$Corr\{\tilde{p}_j(A), \tilde{p}_k(A)\} = \frac{\mathbb{P}(X_i^{(j)} = X_m^{(k)})}{\sqrt{\mathbb{P}(X_i^{(j)} = X_l^{(j)})}\sqrt{\mathbb{P}(X_m^{(k)} = X_n^{(k)})}}$$

for any i, l, m and n such that $i \neq l$, $m \neq n$, $X_i^{(j_i)}$ and $X_l^{(j_l)}$ come from population with label j, i.e. $j_i = j$ and $j_l = j$ and $X_m^{(j_m)}$ and $X_n^{(j_n)}$ come from population with label k, i.e. $j_m = k$ and $j_n = k$.

Proof.

$$\mathbb{E}\{\tilde{p}_j(A)\tilde{p}_k(A)\} = \mathbb{P}(X_i^{(j)} \in A, X_{k,1} \in A).$$

Then we disintegrate with respect to $\{X_i^{(j)} = X_m^{(k)}\}$ to recover independence.

$$\begin{aligned}\mathbb{P}(X_i^{(j)} \in A, X_m^{(k)} \in A) =& \mathbb{P}(X_i^{(j)} = X_m^{(k)})\mathbb{P}(X_i^{(j)} \in A, X_m^{(k)} \in A \mid X_i^{(j)} = X_m^{(k)}) + \\ &+ \mathbb{P}(X_i^{(j)} \neq X_m^{(k)})\mathbb{P}(X_i^{(j)} \in A, X_m^{(k)} \in A \mid X_i^{(j)} \neq X_m^{(k)}) = \\ =& \mathbb{P}(X_i^{(j)} = X_m^{(k)})H(A) + \mathbb{P}(X_i^{(j)} \neq X_m^{(k)})H(A)^2.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{Cov}\{\tilde{p}_j(A), \tilde{p}_k(A)\} &= \mathbb{E}\{\tilde{p}_j(A)\tilde{p}_k(A)\} - \mathbb{E}\{\tilde{p}_j(A)\}\mathbb{E}\{\tilde{p}_k(A)\} = \\ &= \mathbb{P}(X_i^{(j)} = X_m^{(k)})H(A)\{1 - H(A)\}.\end{aligned}$$

Finally, use Proposition 2.6 to compute the correlation and apply Proposition 2.5. □

Proposition 2.7 expresses the dependence between the underlying random probabilities in terms of the law of the sequence $\boldsymbol{X}$, in particular as a function of the pEPPF for $n = 2$. This result provides an interesting explanation on why the correlation does not depend on the specific measurable set A: correlation between the random probabilities is a consequence uniquely of ties between the variables in $\boldsymbol{X}$.
In the following for sake of simplicity we omit to specify that A is measurable and that $0 < H(A) < 1$ (i.e. $\tilde{p}_j(A)$ has not a degenerate distribution).

Corollary 2.2. *If $\boldsymbol{X}$ follows a mSSM with associated $(\tilde{p}_1, \ldots, \tilde{p}_J) \sim$ MSSP then*

(c-i) $Corr\{\tilde{p}_j(A), \tilde{p}_k(A)\} \geq 0$.

(c-ii) $Corr\{\tilde{p}_j(A), \tilde{p}_k(A)\} = 0$ *iff* $\mathbb{P}(X_i^{(j)} = X_m^{(k)}) = 0$.

(c-iii) If $\tilde{p}_j$ and $\tilde{p}_k$ are equal in distribution then

$$Corr\{\tilde{p}_j(A), \tilde{p}_k(A)\} = \frac{Corr(X_i^{(j)}, X_m^{(k)})}{Corr(X_i^{(j)}, X_l^{(j)})} = \frac{\mathbb{P}(X_i^{(j)} = X_m^{(k)})}{\mathbb{P}(X_i^{(j)} = X_l^{(j)})}.$$

Proof. The corollary trivially follows from Proposition 2.7. □

Note that result (c-iii) provides a very straightforward interpretation of what happens when the probability of a tie across samples approaches the probability of a tie within, leading to increases in the correlation towards one.

Example 2.5 (Hierarchical Dirichlet processes)**.** *If $(\tilde{p}_1, \tilde{p}_2)$ are distributed accordingly to a HDP, i.e.*

$$\tilde{p}_i \mid \tilde{p}_0 \overset{i.i.d.}{\sim} DP(\theta, \tilde{p}_0), \quad \tilde{p}_0 \sim DP(\theta_0, P_0),$$

then

$$\mathbb{P}(X_i^{(1)} = X_l^{(1)}) = 1 - \frac{\theta\,\theta_0}{(1+\theta)(1+\theta_j)},$$

$$\mathbb{P}(X_i^{(1)} = X_m^{(2)}) = \frac{1}{1+\theta_0},$$

and

$$Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = \frac{1+\theta}{1+\theta_0+\theta}$$

Therefore,

$$\lim_{\theta_0 \to +\infty} Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = 0 \qquad \lim_{\theta \to +\infty} Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = 1$$

Example 2.6 (Nested Dirichlet processes)**.** *If $(\tilde{p}_1, \tilde{p}_2)$ are distributed accordingly to a NDP, as in Section 1.4.3 then*

$$\mathbb{P}(X_i^{(1)} = X_l^{(1)}) = \frac{1}{1+\theta},$$

$$\mathbb{P}(X_i^{(1)} = X_m^{(2)}) = \frac{1}{(1+\theta_0)(1+\theta)},$$

and

$$Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = \frac{1+\theta}{(1+\theta_0)(1+\theta)}$$

Therefore,

$$\lim_{\theta_0 \to +\infty} Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = 0 \qquad \lim_{\theta_0 \to 0} Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = 1$$

Example 2.7 (GM-dependent NRMIs)**.** *If $(\tilde{p}_1, \tilde{p}_2)$ are GM-dependent NRMIs, i.e. the normalized version of the CRMs described in Section 1.4.1, with DP process marginals, then*

$$\mathbb{P}(X_i^{(1)} = X_l^{(1)}) = \frac{1}{1+\theta},$$

$$\mathbb{P}(X_i^{(1)} = X_m^{(2)}) = \theta\,(1-z)\,{}_3F_2(\theta\,(1-z)+2,\,1,\,1;\,\theta+2,\,\theta+2;\,1),$$

and

$$Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = (1-z)\,\frac{\theta}{1+\theta}\,{}_3F_2(\theta\,(1-z)+2,\,1,\,1;\,\theta+2,\,\theta+2;\,1).$$

Therefore,

$$\lim_{z\to 1} Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = 0 \qquad \lim_{z\to 0} Corr\{\tilde{p}_1(A), \tilde{p}_2(A)\} = 1$$

2.4.2 Higher moments in multivariate species sampling processes

Proposition 2.8. *If $\boldsymbol{X}_j \sim SSM$ (with associated SSP denoted by $\tilde{p}_j$) then, for every natural number q,*

$$\mathbb{E}\{\tilde{p}_j(A)^q\} = \mathbb{E}\big\{H(A)^{K^{(j)}_{1:q}}\big\},$$

where $K^{(j)}_{1:q}$ is the random number of species in a sample of size q from $\boldsymbol{X}_j$, i.e. $\boldsymbol{X}_{j,1:q}$.

Proof.

$$\mathbb{E}\{\tilde{p}_j(A)^q\} = \mathbb{P}\{\boldsymbol{X}_{j,1:q} \in A^q\}.$$

Then we disintegrate with respect to the partition $\Pi^{(j)}_q$ of $\boldsymbol{X}_{j,1:q}$ to recover independence and aggregate by symmetry induced by exchangeability.

$$\begin{aligned}
&\sum_{\Pi^{(j)}_q \in \mathscr{P}(\boldsymbol{X}_{j,1:q})} \mathbb{P}\big\{\boldsymbol{X}_{j,1:q} \in A^q \mid \Pi^{(j)}_q\big\}\mathbb{P}\big\{\Pi^{(j)}_q\big\} = \\
&= \sum_{s=1}^{q} H(A)^s \sum_{\Pi^{(j)}_q \in \mathscr{P}(\boldsymbol{X}_{j,1:q}):K_q=s} \mathbb{P}\big\{\Pi^{(j)}_q\big\} = \\
&= \sum_{s=1}^{q} H(A)^s \sum_{(n_1,\dots,n_s)\in\rho_s(q)} \frac{1}{s!}\binom{q}{n_1,\dots,n_s} f(n_1,\dots,n_s) = \\
&= \sum_{s=1}^{q} H(A)^s \mathbb{P}(K^{(j)}_q = s) = \mathbb{E}\big\{H(A)^{K^{(j)}_q}\big\}.
\end{aligned}$$

where f is the EPPF associated to $\Pi^{(j)}_q$ and $\binom{q}{n_1,\dots,n_s} = \frac{q!}{n_1!\cdots n_s!}$. □

Proposition 2.9. *Let $\{A_1,\dots,A_h\}$ be a family of pairwise disjoint measurable sets. If $\boldsymbol{X}_j \sim SSM$ (with associated SSP denoted by $\tilde{p}_j$) then, for every sequence of natural numbers $q_1, q_2, \dots, q_h$,*

$$\begin{aligned}
&\mathbb{E}\{\tilde{p}_j(A_1)^{q_1}\cdots\tilde{p}_j(A_h)^{q_h}\} = \\
&= \mathbb{E}\left[H(A_1)^{K^{(j)}_{1:q_1}} H(A_2)^{K^{(j)}_{q_1+1:q_2}} \cdots H(A_h)^{K^{(j)}_{q_{h-1}+1:q_h}} \mid E_{\neq}\right]\mathbb{P}(E_{\neq})
\end{aligned}$$

where $K^{(j)}_{a:b}$ is the number of species in the "block of observations" from the a-th to the b-th observation, in a sample of size $q_1+\cdots+q_h$ from $\boldsymbol{X}_j$. We denote each block of observations as $\boldsymbol{X}_{j,a:b}$. $E_{\neq}$ is the event of no shared species across the blocks of observations.

Proof. For notational convenience we prove the proposition for $h = 2$. The general case can be proven in the same exact way.

$$\mathbb{E}\{\tilde{p}_j(A_1)^{q_1}\tilde{p}_j(A_2)^{q_2}\} = \mathbb{P}\{\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2}\}$$

where $q = q_1 + q_2$. Denote now with $\mathcal{A}_{q_1,q_2} \subset \mathcal{P}(\boldsymbol{X}_{j,1:q})$ the set of all possible partitions $\Pi_q^{(j)}$ of the elements in $\boldsymbol{X}_{j,1:q}$ such that the elements in $\boldsymbol{X}_{j,1:q_1}$ and in $\boldsymbol{X}_{j,q_1+1:q_2}$ do not belong to the same set according to $\Pi_q^{(j)}$. It follows that

$$\begin{aligned}
\mathbb{P}\{\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2}\} =& \mathbb{P}\{(\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2}) \cap (\Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2})\} = \\
=& \mathbb{P}(\Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2})\mathbb{P}\{\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2} \mid \Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2}\} = \\
=& \sum_{s_1=1}^{q_1}\sum_{s_2=1}^{q_2} \mathbb{P}(\Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2}, K_{1:q_1}^{(j)} = s_1, K_{q_1+1:q_2}^{(j)} = s_2)\times \\
& \times \mathbb{P}\{\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2} \mid \Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2}, K_{1:q_1}^{(j)} = s_1, K_{q_1+1:q_2}^{(j)} = s_2\} = \\
=& \sum_{s_1=1}^{q_1}\sum_{s_2=1}^{q_2} H(A_1)^{s_1}H(A_2)^{s_2}\mathbb{P}(\Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2}, K_{1:q_1}^{(j)} = s_1, K_{q_1+1:q_2}^{(j)} = s_2) \\
=& E\left[H(A_1)^{K_{1:q_1}^{(j)}}H(A_2)^{K_{q_1+1:q_2}^{(j)}} \mid \Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2}\right]\mathbb{P}(\Pi_q^{(j)} \in \mathcal{A}_{q_1,q_2})
\end{aligned}$$

□

Theorem 2.3. *If $\boldsymbol{X} \sim mSSM$ (with associated mSSP denoted by $(\tilde{p}_1, \ldots, \tilde{p}_J)$) then, for every sequence of natural numbers $q_1, q_2, \ldots, q_J$,*

$$\mathbb{E}\{\tilde{p}_1(A)^{q_1} \cdots \tilde{p}_J(A)^{q_J}\} = \mathbb{E}\big\{H(A)^{K_{q_1,\ldots,q_J}}\big\},$$

where $K_{q_1,\ldots,q_J}$ is the overall number of species observed in a sample from $\boldsymbol{X}$, which contains q_j observations from population j, for each $j \in \{1, \ldots, J\}$, i.e. $\boldsymbol{X}_{1:q_1,\ldots,1:q_J}$.

Proof.

$$\mathbb{E}\{\tilde{p}_1(A)^{q_1} \cdots \tilde{p}_J(A)^{q_J}\} = \mathbb{P}\{\boldsymbol{X}_{j,1:q_j} \in A^{q_j} : j = 1, \ldots, J\},$$

Then we disintegrate with respect to the possible partitions Π_q of $\boldsymbol{X}_{1:q_1,\ldots,1:q_J}$ to recover independence and aggregate by symmetry.

$$\sum_{\Pi_q \in \mathcal{P}(\boldsymbol{X}_{1:q_1,\ldots,1:q_J})} \mathbb{P}\big\{\boldsymbol{X}_{j,1:q_j} \in A^{q_j} : j = 1, \ldots, J \mid \Pi_q\big\}\mathbb{P}\big\{\Pi_q\big\} =$$

$$
\begin{aligned}
&= \sum_{s=1}^{q} H(A)^s \sum_{\Pi_q \in \mathscr{P}(\boldsymbol{X}_{1:q_1,\ldots,1:q_J}):K_{q_1,\ldots,q_J}=s} f(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_J) \\
&= \sum_{s=1}^{q} H(A)^s \mathbb{P}(K_{q_1,\ldots,q_J}=s) \\
&= \mathbb{E}\big\{H(A)^{K_{q_1,\ldots,q_J}}\big\}.
\end{aligned}
$$

where f is the pEPPF associated to Π_q. □

Theorem 2.4. *Let $\{A_1,\ldots,A_J\}$ be a family of pairwise disjoint measurable sets. If $\boldsymbol{X} \sim mSSM$ (with associated mSSP denoted by $(\tilde{p}_1,\ldots,\tilde{p}_J)$) then, for every sequence of natural numbers q_1,q_2, …,q_J,*

$$
\mathbb{E}\{\tilde{p}_1(A_1)^{q_1}\cdots\tilde{p}_J(A_J)^{q_J}\} = E\left[H(A_1)^{K^{(1)}_{1:q_1}}\cdots H(A_J)^{K^{(J)}_{1:q_J}} \mid E_{\neq}\right]\mathbb{P}(E_{\neq}).
$$

where $K^{(j)}_{1:q_j}$ is the number of species from population j, observed in a sample from $\boldsymbol{X}$, which contains q_j observations from population j, for each $j \in \{1,\ldots,J\}$, and $E_{\neq}$ is the event of no shared species across populations in the same sample.

Proof. First we go on the level of observations

$$
\mathbb{E}\Big\{\prod_{j=1}^{J}\tilde{p}_j(A_j)^{q_j}\Big\} = \mathbb{P}\big\{\boldsymbol{X}_{1:q_1,\ldots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j}\big\},
$$

Denote now with $\mathscr{A}_{q_1,\ldots,q_J} \subset \mathscr{P}(\boldsymbol{X}_{1:q_1,\ldots,1:q_J})$ the set of all possible partitions Π_q of the elements in $\boldsymbol{X}_{1:q_1,\ldots,1:q_J}$ such that the elements in $\boldsymbol{X}_{j,1:q_1}$ and in $\boldsymbol{X}_{j',1:q_{j'}}$ do not belong to the same set, for any $j \neq j'$ according to Π_q.

$$
\begin{aligned}
&\mathbb{P}\left\{\boldsymbol{X}_{1:q_1,\ldots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j}\right\} = \mathbb{P}\left\{(\boldsymbol{X}_{1:q_1,\ldots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j}) \cap (\Pi_q \in \mathscr{A}_{q_1,\ldots,q_J})\right\} = \\
&= \mathbb{P}\left(\Pi_q \in \mathscr{A}_{q_1,\ldots,q_J}\right)\mathbb{P}\left\{\boldsymbol{X}_{1:q_1,\ldots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j} \mid \Pi_q \in \mathscr{A}_{q_1,\ldots,q_J}\right\} = \\
&= \sum_{s_1=1}^{q_1}\cdots\sum_{s_J=1}^{q_J}\mathbb{P}(\Pi_q \in \mathscr{A}_{q_1,\ldots,q_J}, K^{(1)}_{q_1}=s_1,\ldots,K^{(J)}_{q_J}=s_J)\times \\
&\quad\times\mathbb{P}\left\{\boldsymbol{X}_{1:q_1,\ldots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j} \mid \Pi_q \in \mathscr{A}_{q_1,\ldots,q_J}, K^{(1)}_{q_1}=s_1,\ldots,K^{(J)}_{q_J}=s_J\right\} =
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{s_1=1}^{q_1} \cdots \sum_{s_J=1}^{q_J} H(A_1)^{s_1} \cdots H(A_J)^{s_J} \times \\
&\quad \times \mathbb{P}(\Pi_q \in \mathcal{A}_{q_1,\ldots,q_J}, K_{q_1}^{(1)} = s_1, \ldots, K_{q_J}^{(J)} = s_J) \\
&= E\left[H(A_1)^{K_{q_1}^{(1)}} \cdots H(A_J)^{K_{q_J}^{(J)}} \mid \Pi_q \in \mathcal{A}_{q_1,\ldots,q_J} \right] \mathbb{P}(\Pi_q \in \mathcal{A}_{q_1,\ldots,q_J})
\end{aligned}
$$

□

2.4.3 Characterization of multivariate species sampling processes

Finally, in the next theorem we provide a characterization of observable in terms of random probabilities and vice versa. We show that in the series representation of mSSP there is a latent infinite set of i.i.d. atoms shared across all the marginal SSP, while the partition law is controlled by the weights. Thus, at least conceptually, mSSP are strongly connected with common atoms dependent nonparametric processes (MacEachern, 1999, 2000; Quintana et al., 2020).

Theorem 2.5. $\boldsymbol{X} \sim$ *mSSM if and only if*

$$
\tilde{p}_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h} \delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) H, \qquad \text{for } j = 1, \ldots, J,
$$

where θ_h *are i.i.d from* H *and independent from* $\boldsymbol{\pi} = (\pi_{j,h})_{j,h}$.

Proof. To prove the *if part* of the theorem, note that conditionally on $\tilde{p}_1$, $X_1^{(j_1)}$ is either equal to some θ_h or sampled from H. However, since θ_h are i.i.d from H, marginally

$$
X_1^{(j_1)} \sim H.
$$

To compute the distribution of $X_{n+1}^{(j_{n+1})}$ given $\boldsymbol{X}_{1:n}$, let us choose any arbitrary order for the variables in $(\theta_h)_{h\geq 1}$ and introduce two sequences of auxiliary random variables: $\{c_i : i = 1, \ldots, n+1\}$, such that

$$
c_i = \begin{cases} h & \text{iff } X_i^{(j_i)} = \theta_h \\ 0 & \text{iff } X_i^{(j_i)} \neq \theta_h \text{ for any } h \end{cases}
$$

and $\{\phi_i : i = 1, \ldots, n+1\}$, such that $\phi_i \overset{iid}{\sim} H$ with ϕ_i independent from $(\theta_h)_{h\geq 1}$ and $(\pi_{j,h})_{j,h}$. So, one have that

$$
X_i^{(j_i)} \mid (\theta_h)_{h\geq 1},\, c_i,\, \phi_i \overset{ind}{\sim} \mathbb{1}_{\{c_i \geq 1\}}\, \delta_{\theta_{c_i}} + \mathbb{1}_{\{c_i = 0\}}\, \delta_{\phi_i}.
$$

Moreover denote with c_l^*, for $l = 1, \ldots, K$, the unique values of c_i in order of appearance of the observable.

Thus,

$$\begin{aligned}\mathbb{P}\left(X_{n+1}^{(j_{n+1})} \in A \mid \boldsymbol{X}_{1:n}\right) &= \sum_{h\geq 0} \mathbb{P}\left(X_{n+1}^{(j_{n+1})} \in A \mid c_{n+1}=h, \boldsymbol{X}_{1:n}\right) \mathbb{P}(c_{n+1}=h \mid \boldsymbol{X}_{1:n}) \\ &= \sum_{h\geq 1} \mathbb{P}\left(\theta_h \in A \mid c_{n+1}=h, \boldsymbol{X}_{1:n}\right) \mathbb{P}(c_{n+1}=h \mid \boldsymbol{X}_{1:n}) + \\ &\quad + \mathbb{P}\left(\phi_{n+1} \in A \mid c_{n+1}=0, \boldsymbol{X}_{1:n}\right) \mathbb{P}(c_{n+1}=0 \mid \boldsymbol{X}_{1:n})\end{aligned}$$

using the notation $\boldsymbol{c}_{1:n} = \{c_i : i = 1 \dots n\}$, we have

$$\mathbb{P}\left(\theta_h \in A \mid c_{n+1}=h, \boldsymbol{X}_{1:n}\right) = \mathbb{E}\left[\mathbb{P}\left(\theta_h \in A \mid \boldsymbol{c}_{1:n}, c_{n+1}=h, \boldsymbol{X}_{1:n}\right) \mid c_{n+1}=h, \boldsymbol{X}_{1:n}\right]$$

where

$$\mathbb{P}\left(\theta_h \in A \mid \boldsymbol{c}_{1:n}, c_{n+1}=h, \boldsymbol{X}_{1:n}\right) = \begin{cases} H(A) & \text{if } h \neq c_l^* \quad \forall l \\ \mathbb{1}_{\{X_l^* \in A\}} & \text{if } h = c_l^* \end{cases}$$

computing the expected value

$$\begin{aligned}\mathbb{P}\left(\theta_h \in A \mid c_{n+1}=h, \boldsymbol{X}_{1:n}\right) &= H(A)\, \mathbb{P}(h \neq c_l^* \mid c_{n+1}=h, \boldsymbol{X}_{1:n}) \\ &\quad + \sum_{l=1}^{K} \mathbb{1}_{\{X_l^* \in A\}} \mathbb{P}(h = c_l^* \mid c_{n+1}=h, \boldsymbol{X}_{1:n}) \\ &= H(A)\, \mathbb{P}(c_{n+1} \neq c_l^* \mid c_{n+1}=h, \boldsymbol{X}_{1:n}) \\ &\quad + \sum_{l=1}^{K} \mathbb{1}_{\{X_l^* \in A\}} \mathbb{P}(c_{n+1} = c_l^* \mid c_{n+1}=h, \boldsymbol{X}_{1:n})\end{aligned}$$

Putting everything together we have

$$\begin{aligned}\mathbb{P}\left(X_{n+1}^{(j_{n+1})} \in A \mid \boldsymbol{X}_{1:n}\right) &= H(A)\, \mathbb{P}(c_{n+1} \neq c_l^* \mid \boldsymbol{X}_{1:n}) \\ &\quad + \sum_{l=1}^{K} \mathbb{1}_{\{X_l^* \in A\}} \mathbb{P}(c_{n+1} = c_l^* \mid \boldsymbol{X}_{1:n})\end{aligned}$$

Finally, note that (by hypothesis and de Finetti's theorem) we have that $\boldsymbol{X}$ is partially exchangeable. Thus, by Theorem 2.1, the random partition Π, define by the equivalence relation $i \sim i'$ iff $X_i^{(j_i)} = X_{i'}^{(j_i')}$, is partially exchangeable with respect to $\mathcal{D}$. Moreover, by Thereom 2.2, we know that the partial exchangeability of Π is a necessary and sufficient condition to prove that the functions $\mathbb{P}(c_{n+1} = c_l^* \mid \boldsymbol{X}_{1:n})$ are mPPF according to Definition 2.8.

To prove the *only if* part, recall that by Proposition 2.4 we have that $\mathbf{X}_j$ is marginally a species sampling model (SSM) and thus by Pitman (1996) there exists a species sampling

process

$$\tilde{p}_j \overset{a.s.}{=} \sum_{h\geq 1} \tilde{\pi}_{j,h}\delta_{\tilde{\theta}_{j,h}} + \left(1 - \sum_{h\geq 1} \tilde{\pi}_{j,h}\right) H_j\,,$$

where $\tilde{\theta}_{j,h} \overset{iid}{\sim} H_j$, such that $X_i^{(j_i)} \mid \tilde{p}_{j_i} \overset{ind}{\sim} \tilde{p}_{j_i}$, for $i \geq 1$. Note that, by definition of SSP, marginally $X_i^{(j_i)} \sim H_{j_i}$ and, by Definition 2.7 of mSSM, $X_i^{(j_i)} \sim H$, for every $i \geq 1$. Therefore $H_j = H$ for $j = 1, \ldots, J$. Thus for any fixed $j \in [J]$, $\tilde{\theta}_{j,h} \overset{iid}{\sim} H$, for $h \geq 1$. Provided these marginal laws for $\tilde{p}_j$, for $j \in [J]$, we have to derive the joint law of $(\tilde{p}_1, \ldots, \tilde{p}_J)$. More precisely, we are interested just in the joint law of the random atoms. Note that given $X_i^{(j_i)} \neq X_{i'}^{(j_i')}$, $X_i^{(j_i)}$ and $X_{i'}^{(j_i')}$ are independent. This implies that given $\tilde{\theta}_{j,h} \neq \tilde{\theta}_{j',h'}$, $\tilde{\theta}_{j,h}$ and $\tilde{\theta}_{j',h'}$ are independent. However, there can be ties between the unique random atoms $\tilde{\theta}_{j,h}$ across different j's.
Let $\theta_1, \theta_2, \ldots$ be the unique random atoms across $\tilde{\theta}_{1,1}, \ldots, \tilde{\theta}_{J,1}, \tilde{\theta}_{1,2}, \ldots, \tilde{\theta}_{J,2}, \ldots$ (in a given order). Therefore, we can rewrite

$$\tilde{p}_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}\theta_h + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) H,$$

where

$$\begin{cases} \pi_{j,h} = \tilde{\pi}_{j,h'} & \text{if } \theta_h = \tilde{\theta}_{j,h'} \\ \pi_{j,h} = 0 & \text{if } \nexists\ \tilde{\theta}_{j,h'} \text{ s.t. } \theta_h = \tilde{\theta}_{j,h'} \end{cases}$$

□

2.5 Regular mSSP

Theorem 2.5 provides a new definition of mSSP which is

Definition 2.10 (Multivariate species sampling process 2)**.** *A vector of random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is a mSSP if*

$$\tilde{p}_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}\delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) P_0, \qquad \text{for } j = 1, \ldots, J,$$

where the atoms $(\theta_h)_{h\geq 1}$ are i.i.d from the non-atomic distribution P_0, the weights $\boldsymbol{\pi} = (\pi_{j,h})_{j,h}$ are such that $\mathbb{P}[0 \leq \pi_{j,h} \leq 1] = 1$ for any j and h, and atoms and weights are independent. Moreover, if $\sum_{h\geq 1} \pi_{j,h} \overset{a.s.}{=} 1$, for any j, $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is said proper.

Definition 2.10 clearly clarifies the link between mSSPs and SSPs (cf. Definition 2.1). Indeed, starting from it, it is straightforward to prove that each coordinate of a mSSP is marginally a SSP and that, basically, a mSSP arises when many SSP share the same atoms

$(\theta_h)_{h\geq 1}$. However, it is important to notice that the conditions imposed on the weights $\boldsymbol{\pi}$ by Definition 2.10 are very mild. Since the $\pi_{j,h}$'s can be almost surely null, the random probabilities $\tilde{p}_1, \ldots, \tilde{p}_J$ may actually share just few or even none of the atoms with positive probability. Even though Definition 2.10 clearly highlights the connection between SSP and mSSP, it may be convenient to adopt a different notation that allows to distinguish between those atoms that can actually be shared across different probability measures and those that are specific to a certain process. This can be done representing each process only in terms of those weights that are not almost surely null, as done in the following.
Let us start considering a bivariate mSSP $(\tilde{p}_1, \tilde{p}_2)$. Definition 2.10 can be equivalently restated writing each measure as

$$\tilde{p}_j \overset{a.s.}{=} \sum_{h=1}^{H} \pi_{j,h}^{(1,2)} \delta_{\theta_h} + \sum_{k=1}^{K_j} \pi_{j,k}^{(j)} \delta_{\eta_{j,k}} + \left(1 - \sum_{h=1}^{H} \pi_{j,h}^{(1,2)} - \sum_{k=1}^{K_j} \pi_{j,k}^{(j)}\right) P_0, \qquad \text{for } j = 1, 2.$$

where $\mathbb{P}[\pi_{j,h}^{(\cdot)} > 0] > 0$, H and K_j are non-random and have value in $\{0, 1, \ldots, +\infty\}$, and all atoms are i.i.d. from P_0. We adopt the convention $\sum_{h=1}^{0} x = 0$. Notice that, accordingly to this notation, θ_h is now an atom shared with positive probability by $\tilde{p}_1$ and $\tilde{p}_2$, while $\eta_{j,k}$ is specific to the measure $\tilde{p}_j$ and cannot be shared. Moreover, each of the two measures can be written without lost of generality as a mixture of three components, defining

$$\omega_j^{(1,2)} = \sum_{h=1}^{H} \pi_{j,h}^{(1,2)} \qquad \bar{\pi}_{j,h}^{(1,2)} = \begin{cases} \pi_{j,h}^{(1,2)} / \omega_j^{(1,2)} & \text{if } \omega_j^{(1,2)} > 0 \\ 0 & \text{if } \omega_j^{(1,2)} = 0 \end{cases}$$

and

$$\omega_j = \sum_{k=1}^{K_j} \pi_{j,k}^{(j)} \qquad \bar{\pi}_{j,k}^{(j)} = \begin{cases} \pi_{j,k}^{(j)} / \omega_j & \text{if } \omega_j > 0 \\ 0 & \text{if } \omega_j = 0 \end{cases}.$$

and as shown in the following definition.

Definition 2.11 (Multivariate species sampling model 3)**.** *A bivariate vector of random probability measures $(\tilde{p}_1, \tilde{p}_2)$ is said a mSSP if*

$$\tilde{p}_j \overset{a.s.}{=} \omega_j^{(1,2)} \sum_{h=1}^{H} \bar{\pi}_{j,h}^{(1,2)} \delta_{\theta_h} + \omega_j \sum_{k=1}^{K_j} \bar{\pi}_{j,k}^{(j)} \delta_{\eta_{j,k}} + \left(1 - \omega_j^{(1,2)} - \omega_j\right) P_0, \qquad \textit{for } j = 1, 2. \tag{2.7}$$

where all atoms are i.i.d. from P_0 and independent from the weights and the weights are such that

- $\mathbb{P}[\bar{\pi}_{j,h}^{(\cdot)} > 0] > 0$, *for any j, h,*
- $\sum_{h=1}^{H} \bar{\pi}_{j,h}^{(1,2)} \overset{a.s}{=} \sum_{k=1}^{K_j} \bar{\pi}_{j,k}^{(j)} \overset{a.s}{=} 1$, *for any j,*
- $\mathbb{P}[0 \leq \omega_j^{(1,2)} \leq 1] = \mathbb{P}[0 \leq \omega_j \leq 1] = 1$.

Thus, the components of the mixtures are such that the first corresponds to possibly shared species, the second is a idiosyncratic measure corresponding to almost surely non-shared species and the last is the improper part of the process. Moreover, we can easily interpret all the parameters involved: θ_h is a species shared with positive probability between the two populations, $\omega_j^{(1,2)}$ is the overall frequency of individuals in population j whose species can be possibly found also in the other population, and H is the number of possibly shared species.

The equivalence between Definition 2.10 for $J = 2$ and Definition 2.11 is straightforward. However, the advantage of Definition 2.11 is twofold. Firstly, we can immediately distinguish between shared and non-shared species. Secondly, this representation allows also to identify a notable subclass of mSSPs, which we name *regular* and arises imposing a simple independence condition between the weights associated to non-shared species, as done in the following definition.

Definition 2.12 (Regular mSSP)**.** *A bivariate mSSP $(\tilde{p}_1, \tilde{p}_2)$ is said regular if $K_1 = K_2 = 0$ or if the weights $\bar{\pi}_{j,k}^{(j)}$ in (2.7) are such that*

$$\left(\bar{\pi}_{1,k}^{(1)}\right)_{k=1}^{K_1} \perp \left(\bar{\pi}_{2,k}^{(2)}\right)_{k=2}^{K_2}.$$

A J-variate mSSP $(\tilde{p}_1, \ldots, \tilde{p}_J)$, with $J > 2$, is said regular if $(\tilde{p}_j, \tilde{p}_k)$ is a regular mSSP for any $j, k \in \{1, \ldots, J\}$.

Intuitively, regularity requires that relative frequencies within non-shared species are independent across populations, and thus, regularity is considered trivially satisfied also when there are not non-shared species. Regular mSSP differ from non-regular mSSPs in the fact that the dependence structure in regular mSSPs admits an outstanding characterization in terms of correlation between the measures. The same result does not holds true in the general class of mSSP, resulting in fundamental differences between regular and non-regular processes. Secondly, regular mSSP are the subclass which appears to be more of interest in statistics, since it includes all mSSPs studied and used in Bayesian nonparametrics till today (e.g. hierarchical processes, nested processes, dependent normalized random measures, etc.) the peculiarity of regular mSSPs is the fact that within this class, the correlation completely characterizes the dependence structure between any two processes. In fact, as shown in next theorem, it is impossible to construct a zero-correlated regular mSSP whose components are not pairwise independent.

Theorem 2.6. *If $(\tilde{p}_1, \ldots, \tilde{p}_J)$ is a regular mSSP, then*

$$Corr(\tilde{p}_j(A), \tilde{p}_k(A)) = 0 \quad \textit{iff} \quad \tilde{p}_j \perp \tilde{p}_k,$$

Proof. Let us consider the representation of $(\tilde{p}_j, \tilde{p}_k)$ as mixtures of three components

$$\tilde{p}_j \overset{a.s.}{=} \omega_j^{(j,k)} \sum_{h=1}^{H} \bar{\pi}_{j,h}^{(j,k)} \delta_{\theta_h} + \omega_j \sum_{l=1}^{K_j} \bar{\pi}_{j,l}^{(j)} \delta_{\eta_{j,l}} + \left(1 - \omega_j^{(j,k)} - \omega_j\right) P_0$$

and

$$\tilde{p}_k \overset{a.s.}{=} \omega_k^{(j,k)} \sum_{h=1}^{H} \bar{\pi}_{k,h}^{(j,k)} \delta_{\theta_h} + \omega_k \sum_{l=1}^{K_k} \bar{\pi}_{k,l}^{(k)} \delta_{\eta_{k,l}} + \left(1 - \omega_k^{(j,k)} - \omega_k\right) P_0.$$

then we have that $\text{Corr}(\tilde{p}_j(A), \tilde{p}_k(A)) = 0$ iff $\text{pr}(X_{1,j} = X_{1,k}) = 0$ iff $\omega_j^{(j,k)} \overset{a.s}{=} \omega_k^{(j,k)} \overset{a.s}{=} 0$. Thus

$$\tilde{p}_j \overset{a.s.}{=} \omega_j \sum_{l=1}^{K_j} \bar{\pi}_{j,l}^{(j)} \delta_{\eta_{j,l}} + (1 - \omega_j) P_0$$

and

$$\tilde{p}_k \overset{a.s.}{=} \omega_k \sum_{l=1}^{K_k} \bar{\pi}_{k,l}^{(k)} \delta_{\eta_{k,l}} + (1 - \omega_k) P_0.$$

and therefore $\tilde{p}_j \perp \tilde{p}_k$. □

Notice however that not all types of regular mSSP can achieve exactly zero correlation. In particular those which do not have idiosincratic and improper components (i.e. $\omega_1^{(1,2)} \overset{a.s}{=} \omega_2^{(1,2)} = 1$), such as hierarchical constructions, cannot produce zero correlation.

2.6 Inference and marginal algorithm

Proposition 2.10. *If $\boldsymbol{X} \sim$ mSSM with pEPPF f, then a marginal algorithm can be derived as*

$$X_{n+1}^{(j_{n+1})} \mid \boldsymbol{X}_{1:n} = \begin{cases} X_l^* & w.p. \ \frac{f([n_{1,1},\ldots,n_{1,l}+1,\ldots,n_{1,c}],\boldsymbol{n}_2\ldots,\boldsymbol{n}_J)}{f([n_{1,1},\ldots,n_{1,l},\ldots,n_{1,c}],\boldsymbol{n}_2,\ldots,\boldsymbol{n}_J)} \\ X_{new}^* & w.p. \ \frac{f([n_{1,1},\ldots,n_{1,l},\ldots,n_{1,c},1],[\boldsymbol{n}_2,0]\ldots,[\boldsymbol{n}_J,0])}{f([n_{1,1},\ldots,n_{1,l},\ldots,n_{1,c}],\boldsymbol{n}_2,\ldots,\boldsymbol{n}_J)} \end{cases}$$

Notice that, by Theorem 2.2, we know that

$$p_{j,l}(\boldsymbol{n}) = \frac{f(\boldsymbol{n}^{lj+})}{f(\boldsymbol{n})} \qquad \forall\, \boldsymbol{n}, \ \ l = 1, \ldots, K+1 \text{ and } j = 1, \ldots, J.$$

Therefore the same algorithm can be expressed using mPPF instead of ratios of pEPPF. From this result, it is straightforward to generalized also the well known algorithms for Dirichlet process mixture model as those in Neal (2000). However, in order to simplify the ratio and have analytical simple results and computationally efficient samplers from Proposition 2.10 is often necessary to introduce some data augmentation to recover product form (e.g. composition of Gibbs type priors).

2.6.1 Probability of a new species

One of the crucial aspects typically considered to choose across different (univariate) SSP is the probability of observing a new species for X_{n+1} not included in the sample $X_1, \dots, X_n$ already observed (see, for instance, De Blasi et al., 2015). Extending the same approach to the multivariate case, we compare here the probability of observing a new species in a certain population which is not included in a sample already observed from a different population. For sake of simplicity, consider the case of a bivariate mSSP $(\tilde{p}_1, \tilde{p}_2)$ and the usual sampling procedure $X_{i,j} \mid (\tilde{p}_1, \tilde{p}_2) \overset{iid}{\sim} p_j$, the quantity of interest is

$$\mathbb{P}[X_{2,1} \text{ is "new"} \mid X_{1,1}, \dots, X_{1,n}].$$

Notice that if, as it usually happens, $\tilde{p}_1 \overset{d}{=} \tilde{p}_2$, then the probability coincides with $\mathbb{P}[X_{1,1}$ is "new" $\mid$ $X_{2,1}, \dots, X_{2,n}]$. We conclude this chapters with the following examples, where we provide the probability of a new species for some of the most famous mSSP.

Example 2.5 (Continue). *If $(\tilde{p}_1, \dots, \tilde{p}_J)$ is distributed accordingly to a HDP, then*

$$\mathbb{P}[X_{2,1} \textit{ is "new"} \mid X_{1,1}, \dots, X_{1,n}] = \sum_{\boldsymbol{l}} \frac{\alpha_0}{\alpha_0 + |\boldsymbol{l}|}\, p(\boldsymbol{l} \mid X_{1,1}, \dots, X_{1,n}), = f_{HDP}\left(n,\, K^{(1)}_{1:n}, \boldsymbol{n}\right)$$

where the sum runs over all $\boldsymbol{l} = (l_1, \dots, l_{K^{(1)}_{1:n}})$, such that $l_h \in \{1, \dots, n_h\}$, where $K^{(1)}_{1:n}$ is the number of distinct species observed in $X_{1,1}, \dots, X_{1,n}$ and n_h is the number of subjects corresponding to the h-th specie in order of appearance. Using the Chinese Franchise metaphor, l_h is the number of tables in the first restaurant serving the h-th dish. Therefore

$$p(\boldsymbol{l} \mid X_{1,1}, \dots, X_{1,n}) \propto \frac{\alpha_0^{K^{(1)}_{1:n}} \alpha^{|\boldsymbol{l}|}}{(\alpha_0)_{|\boldsymbol{l}|}} \prod_{h=1}^{K^{(1)}_{1:n}} \frac{(l_h - 1)!}{(\alpha)_{n_h}} |s(n_h, l_h)| \mathbb{1}_{\{1,\dots,n_h\}}(l_h)$$

where $|s(n,k)|$ are the signless Stirling numbers of the first kind and $(x)_n$ is a Pochhammer symbol representing the rising factorial (cf. Camerlenghi et al., 2018, 2019b).
If $(\tilde{p}_1, \dots, \tilde{p}_J)$ is distributed accordingly to a HPY, then

$$\mathbb{P}[X_{2,1} \textit{ is "new"} \mid X_{1,1}, \dots, X_{1,n}] = \sum_{\boldsymbol{l}} \frac{\alpha_0 + K^{(1)}_{1:n}\sigma_0}{\alpha_0 + |\boldsymbol{l}|} p(\boldsymbol{l} \mid X_{1,1}, \dots, X_{1,n}) = f_{HPY}\left(n,\, K^{(1)}_{1:n}, \boldsymbol{n}\right)$$

where

$$p(\boldsymbol{l} \mid X_{1,1}, \dots, X_{1,n}) \propto \frac{\prod_{t=1}^{|\boldsymbol{l}|-1}(\alpha + t\,\sigma)}{(\alpha_0 + 1)_{|\boldsymbol{l}|-1}} \prod_{h=1}^{K^{(1)}_{1:n}} \frac{\mathcal{C}(n_h, l_h; \sigma)}{\sigma^{l_h}} (1 - \sigma_0)_{l_h - 1} \mathbb{1}_{\{1,\dots,n_h\}}(l_h)$$

where $\mathcal{C}(n,k;\sigma) = \frac{1}{k!}\sum_{i=0}^{k}(-1)^i \binom{k}{i}(-i\sigma)_n$ is the generalized factorial coefficient (cf. Camerlenghi

et al., *2018*, *2019b*).

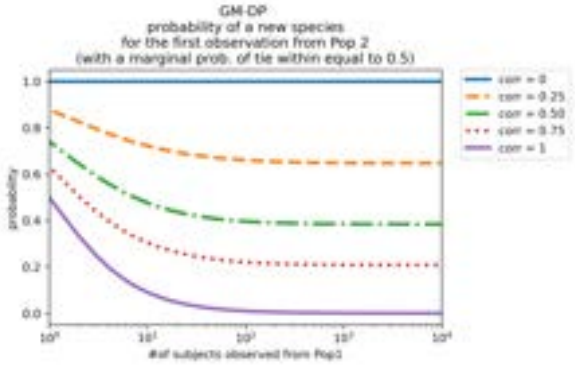
Example 2.6 (Continue). *If* $(\tilde{p}_1, \ldots, \tilde{p}_J)$ *is distributed accordingly to a NDP, then*

$$\mathbb{P}[X_{2,1} \textit{ is "new"} \mid X_{1,1}, \ldots, X_{1,n}] = \frac{\theta_0}{\theta_0 + 1} + \frac{\theta}{(\theta_0 + 1)(\theta + n)} = f_{NDP}\,(n,\, \boldsymbol{n})$$

Example 2.7 (Continue). *If* $(\tilde{p}_1, \tilde{p}_2)$ *is distributed accordingly to a GM-DP, then*

$$\mathbb{P}[X_{2,1} \textit{ is "new"} \mid X_{1,1}, \ldots, X_{1,n}] = f_{GM}\left(n,\, K^{(1)}_{1:n}, \boldsymbol{n}\right)$$

The following figure represents the expected value of $f_{GM}\left(n,\, K^{(1)}_{1:n}, \boldsymbol{n}\right)$*, as function of* n*.*



Chapter 3

Dependent Processes with Full-Range Borrowing of Information

One of the results in the previous chapter is that correlation across observations extracted from different populations (which, for brevity, we also call *correlation across samples*) is non-negative within the class of mSSM. However, as we show in this third chapter, such condition is not implied by the assumption of partial exchangeability of the observables. Moreover, it is important to stress that the correlation across groups is a key ingredient in the definition of the Bayesian learning mechanism induced by any model for partial exchangeable data and, in particular, it controls how borrowing of information across samples is performed. Thus, controlling and quantifying the correlation across samples and its sign should not be regarded as secondary aspects while developing a model.

In this chapter, we extend the study of the correlation also to those existing dependent priors that do not belong to the class of mSSMs, finding that no existing model permits to effectively control the borrowing of information across samples. Therefore, we define a new dependent process that we called *NRMI with full-range borrowing of information* (n-FuRBI) and through which it is possible to freely control the correlation across samples.

The structure of the chapter is the following. The next two sections clarify the framework, the main goals, and some aspects connected to the idea of borrowing of information. Section 3.3 provides general results for dependent processes and an extensive discussion over the correlation induced by these models. Section 3.4 and 3.5 define n-FuRBI, together with their main a-priori properties and the induced correlation structure. Section 3.6 illustrates in depth a posterior characterization of n-FuRBI, with a focus on the predictive distributions. In Section 3.7 we provide MCMC algorithms to derive posterior inference. The chapter concludes with Section 3.8 where detailed applications on simulated and real data can be found.

3.1 Overview and main goals

Consider two sequences of observations $\boldsymbol{X} = (X_i)_{i\geq 1}$ and $\boldsymbol{Y} = (Y_j)_{j\geq 1}$, we recall that they are said to be partially exchangeable if and only if, for all sample sizes (n_1, n_2) and all permutations (π_1, π_2), it holds

$$\big((X_i)_{i=1}^{n_1}, (Y_j)_{j=1}^{n_2}\big) \stackrel{d}{=} \big((X_{\pi_1(i)})_{i=1}^{n_1}, (Y_{\pi_2(j)})_{j=1}^{n_2}\big).$$

In the following, we are referring to $\boldsymbol{X}_{1:n_1} = (X_i)_{i=1}^{n_1}$ and $\boldsymbol{Y}_{1:n_2} = (Y_j)_{j=1}^{n_2}$ for finite n_1 and n_2 with the term *samples*. Moreover, de Finetti representation theorem for partial exchangeability (see Section 1.4) states that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are partially exchangeable if and only if there exist two random probability measures $\tilde{p}_1$ and $\tilde{p}_2$ such that

$$\begin{aligned} X_i \mid \tilde{p}_1 &\overset{iid}{\sim} \tilde{p}_1 \qquad \text{for } i = 1, \dots, n_1 \\ Y_j \mid \tilde{p}_2 &\overset{iid}{\sim} \tilde{p}_2 \qquad \text{for } j = 1, \dots, n_2 \\ (\tilde{p}_1, \tilde{p}_2) &\sim Q \end{aligned} \tag{3.1}$$

In Section 1.4 we reviewed many existing proposal for the dependent prior Q in (3.1), which we have divided mainly in two classes: prior based on the series representation of the underlying probability measures and prior based on CRMs. However, regardless of the specific definition, when choosing among these priors we should be interested in the dependence induced at the level of the observables. In this regard, it is useful to notice that within partial exchangeability observations in different groups cannot be more correlated (in absolute sense) that the ones in the same group.

Lemma 3.1. *Consider two partially exchangeable sequences $\boldsymbol{X}$ and $\boldsymbol{Y}$, such that $\tilde{p}_1$ and $\tilde{p}_2$ in equation (3.1) have the same marginal distribution. Then*

$$-Corr(X_i, X_{i'}) \leq Corr(X_i, Y_j) \leq Corr(X_i, X_{i'}),$$

for any i, i' and j.

Proof. Recall that $(X_i, Y_j) \mid \tilde{p}_1, \tilde{p}_2 \overset{\text{iid}}{\sim} \tilde{p}_1 \times \tilde{p}_2$ and notice that

$$\text{Cov}(X_i, Y_j) = \mathbb{E}[\text{Cov}(X_i, Y_j \mid \tilde{p}_1, \tilde{p}_2)] + \text{Cov}(\mathbb{E}[X_i \mid \tilde{p}_1], \mathbb{E}[Y_j \mid \tilde{p}_2]),$$

where the first term equal 0, so that

$$\text{Cov}(X_i, Y_j) = \text{Cov}\left(\int x\, \tilde{p}_1(dx), \int x\, \tilde{p}_2(dx)\right),$$

and analogously

$$\text{Cov}(X_i, X_{i'}) = \text{Cov}\left(\int x\,\tilde{p}_1(dx), \int x\,\tilde{p}_1(dx)\right) = \text{Var}\left(\int x\,\tilde{p}_1(dx)\right).$$

By Cauchy-Schwartz inequality we have that

$$\left[\text{Cov}\left(\int x\,\tilde{p}_1(dx), \int x\,\tilde{p}_2(dx)\right)\right]^2 \leq \text{Var}\left(\int x\,\tilde{p}_1(dx)\right)\text{Var}\left(\int x\,\tilde{p}_2(dx)\right).$$

Lastly assume that $\tilde{p}_1 \overset{d}{=} \tilde{p}_2$, so that Cauchy-Schwartz inequality becomes

$$-\text{Var}\left(\int x\,\tilde{p}_1(dx)\right) \leq \text{Cov}\left(\int x\,\tilde{p}_1(dx), \int x\,\tilde{p}_2(dx)\right) \leq \text{Var}\left(\int x\,\tilde{p}_1(dx)\right).$$

Substituting the expression in terms of the observables we get

$$-\text{Cov}(X_i, X_{i'}) \leq \text{Cov}(X_i, Y_j) \leq \text{Cov}(X_i, X_{i'}),$$

as desired. □

Notice that if observables are actually exchangeable (i.e. the labels of the groups are irrelevant), the upper bound is attained and it can be shown to be non negative. Thus, the closer the correlation is to the lower bound in Lemma 3.1 the farther the model is from exchangeability.

In exchangeable models, the relationship between the observations is typically driven by ties between them. The first goal of this chapter is to show that a similar scenario can be depicted for partially exchangeable models, even outside the class of multivariate species sampling process (see Chapter 2), as long as $\tilde{p}_1$ and $\tilde{p}_2$ are marginally SSP. For partial exchangeable models, the notion of tie is replaced by the one of *hyper-tie*. The latter will be the key object driving the dependence across samples. In particular, in Section 3.3 we show how to compute the correlation between observations in terms of the probability of an hyper-tie. Such representation highlights that: for model based on the series representation the correlation can rarely be computed, while for models based on CRMs the correlation may actually be computed but it turns out to be always non-negative. Thus, it appears that the available literature focused on a subset of possible values for the correlation. Therefore, the second goal of this chapter is to provide a class of priors under which the correlation can be computed explicitly and can be also negative.

We note that all constructions cited in Section 1.4 require that the random probability measures in (3.1) can be written as

$$\begin{cases} \tilde{p}_1 \overset{a.s.}{=} \sum_{k\geq 1} \bar{J}_k \delta_{\theta_k} \\ \tilde{p}_2 \overset{a.s.}{=} \sum_{k\geq 1} \bar{W}_k \delta_{\phi_k} \end{cases} \quad \text{with } \theta_k \overset{\text{i.i.d.}}{\sim} P_0, \;\; \phi_k \overset{\text{i.i.d.}}{\sim} P_0, \tag{3.2}$$

where P_0 is a fixed probability distribution on $\mathbb{X}$ where the random weights $\left(\{\bar{J}_k\}, \{\bar{W}_k\}\right)$ and the atoms $(\{\theta_k\}, \{\phi_k\})$ are independent, i.e., $\tilde{p}_1$ and $\tilde{p}_2$ are marginally SSP. Next Lemma shows that the sign of the correlation depends *only* on the dependence between the atoms.

Lemma 3.2. *Consider two partially exchangeable sequences X and Y, such that the underlying $\tilde{p}_1$ and $\tilde{p}_2$ are as in (3.2). Suppose moreover Corr$(\theta_k, \phi_{k'}) \geq 0$ for any choice of k and k'. Then Corr $(X_i, Y_j) \geq 0$.*

Proof. By definition of covariance we have

$$\text{Cov}(X_i, Y_j) = \text{Cov}\left(\sum_{j\geq 1} J_j\theta_j, \sum_{k\geq 1} W_k\phi_k\right) = \sum_{j\geq 1}\sum_{k\geq 1} \text{Cov}\left(J_j\theta_j, W_k\phi_k\right).$$

For arbitrary j and k we have

$$E\left(J_jW_k\theta_j\phi_k\right) = E[J_jW_k]E[\theta_j\phi_k] \geq E[J_jW_k]E[\theta_j]E[\phi_k],$$

since $\text{Cov}(\theta_j, \phi_k) \geq 0$. Denoting $c = E[\theta_j] = E[\phi_k]$, we get

$$\text{Cov}\left(J_j\theta_j, W_k\phi_k\right) \geq c^2\text{Cov}(J_j, W_k).$$

Finally, since $\tilde{p}_1$ and $\tilde{p}_2$ are random probability measures it holds

$$\text{Cov}(X_i, Y_j) \geq c^2\text{Cov}\left(\sum_{j\geq 1} J_j, \sum_{k\geq 1} W_k\right) = 0,$$

that concludes the proof. □

Therefore, in order to obtain a negative correlation we cannot simply allow sharing of the atoms between the two groups, that was a popular way of introducing dependence (e.g. in hierarchical structures); instead, we need a flexible joint distribution for the sequence of atoms. This task is accomplished by the prior proposed in this chapter, n-FuRBI, that allows to attain any possible value for the correlation specified in Lemma 3.1 and, moreover, it encompasses many previous cited constructions as special cases. We will show that it combines the flexibility of the series construction with the analytical tractability derived by CRMs.

3.2 Borrowing of information

As already mentioned the correlation between observations is useful not only to quantify the dependence induced by the prior Q, but it connects also to the notion of *borrowing of information*. This term was first coined by John Tukey (Brillinger, 2002) and popularized

in reference to Stein's paradox and empirical Bayes techniques in Efron & Morris (1977). More generally, statisticians refer to borrowing of information when many samples contributes to inference related to just one sample. Imagine to collect the data $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$, while being actually interested only on the parameter $\tilde{p}_1$ in (3.1) associated to $\boldsymbol{X}$. The simplest approach could be to disregard the second sample, with the disadvantage of losing possibly useful information. The typical borrowing instead consists in using all the observations and shrinking the estimates for different samples towards each other, that happens when the prior induces positive correlation between observations in different samples. A toy example to clarify this concepts is the following. Let us consider the situation in which observations coming from two different populations have been collected and we assume that

$$\begin{aligned} X_i \mid \mu_x &\overset{iid}{\sim} \mathrm{N}(\mu_x, 1) \qquad \text{for } i = 1, \dots, n \\ Y_j \mid \mu_y &\overset{iid}{\sim} \mathrm{N}(\mu_y, 1) \qquad \text{for } j = 1, \dots, m \end{aligned}$$

To obtain a working model, one has to specify a certain prior over μ_x and μ_y. The main well-known strategies we may employ are the following

- Modeling μ_x and μ_y as independent, which ultimately means that we do not consider the information coming from one sample to be relevant for the inference of the other.
- Assuming $\mu_x \overset{a.s.}{=} \mu_y$, which is reasonable only if we have strong prior information regarding the fact that the distribution in the two populations is the same.
- Modeling μ_x and μ_y as dependent, assuming positive correlation between them. This ultimately corresponds to the idea that, if our posterior estimate of μ_y is higher that our prior guess than we should increase also our guess about μ_x.

To clarify the last point, let us compare a typical strategy used to perform borrowing of information, which is provided by the following prior

$$\begin{aligned} \mu_x \mid \mu_0 &\sim \mathrm{N}(\mu_0, 1) \\ \mu_y \mid \mu_0 &\sim \mathrm{N}(\mu_0, 1) \\ \mu_0 &\sim \mathrm{N}(\nu, 1) \end{aligned} \tag{3.3}$$

with the correspondent independent prior, which preserves the same marginals

$$\begin{gathered} \mu_x \sim \mathrm{N}(\nu, 2) \qquad \mu_y \sim \mathrm{N}(\nu, 2) \\ \mu_x \perp \mu_y \end{gathered}$$

Let us assume that only the second sample has been observed and we note what happens to the distribution of μ_x under the two specifications. Under independence, there is no

change and thus

$$p(\mu_x \mid \boldsymbol{Y}_{1:m}) \equiv \mathrm{N}(\nu,\, 2)$$

while under model (3.3) the new distribution of μ_x is

$$p(\mu_x \mid \boldsymbol{Y}_{1:m}) \propto \int_{-\infty}^{+\infty} p(\mu_x \mid \mu_0) p(\mu_0 \mid \boldsymbol{Y}_{1:m}) \mathrm{d}\mu_0 \equiv \mathrm{N}\left(\frac{1}{2m+1}\nu + \frac{2m}{2m+1}\frac{\nu+\bar{y}}{2},\, 1 + \frac{m+1}{2m+1}\right)$$

where the expected value can be rewritten as

$$\mathbb{E}[\mu_x \mid \boldsymbol{Y}_{1:m}] = \nu + \frac{m}{2m+1}(\bar{y} - \nu)$$

Thus, when $\bar{y} > \nu$ the borrowing results in an increase in the expected value of μ_x, while if $\bar{y} < \nu$ we will observe a decrease in the expected value of μ_x.

Finally, imagine that also the first sample has been observed. The point estimates under a square loss function for μ_x under independence and model (3.3) are respectively

$$\hat{\mu}_x = \frac{n}{n+1/2}\bar{x} + \frac{1/2}{n+1/2}\nu$$

$$\hat{\mu}_x = \frac{n}{n+1/2}\bar{x} + \frac{1/2}{n+1/2}\left[\nu + \frac{m}{2m+1}(\bar{y} - \nu)\right].$$

The same reasoning holds true if we consider the prediction of X_1 given Y_1: in the independent case we have

$$p(X_1 \mid Y_1) = \mathrm{N}(\nu, 3)$$

while, using classical borrowing we have

$$p(X_1 \mid Y_1) = \mathrm{N}\left(\nu + \frac{1}{3}(Y_1 - \nu), 2 + \frac{2}{3}\right)$$

However, some applications may still require to use the information in $\boldsymbol{Y}_{1:m}$ to improve the inference on $\tilde{p}_1$ and the prediction of X_1, but without such assumption of positive correlation between $X_1(\mu_x)$ and $Y_1(\mu_y)$.

Think for example about different investments in financial markets, whose returns are related, but may exhibit opposite behaviour. Our proposal allows to consider any interesting choice: independence, classical shrinkage, but also repulsion of estimates for different samples, generating what we call *full–range borrowing of information*. Finally, notice that the repulsive behaviour proposed in this chapter is different from the one of the repulsive priors, introduced in Petralia et al. (2012); Quinlan et al. (2017), that instead consider repulsive distributions for atoms of the same random probability measure.

3.3 General results on dependent processes

As stated in Section 3.1, the vast majority of dependent processes introduced in the literature are almost surely discrete and marginal SSPs, therefore they admit a representation as in (3.2). Moreover, each atom in $\tilde{p}_1$ is independent from all atoms of $\tilde{p}_2$ except one, and the vice versa is true for the atoms of $\tilde{p}_2$, i.e. $\theta_k \perp \phi_h$ for $k \neq h$. For all those priors, $\tilde{p}_1$ and $\tilde{p}_2$ can be seen, without loss of generality, as the projection over different coordinates of other two processes, namely p_1 and p_2, such that

$$p_1 \overset{a.s.}{=} \sum_{k\geq 1} \bar{J}_k \delta_{(\theta_k,\phi_k)}, \quad p_2 \overset{a.s.}{=} \sum_{k\geq 1} \bar{W}_k \delta_{(\theta_k,\phi_k)}, \quad (\theta_k, \phi_k) \overset{i.i.d.}{\sim} G_0, \tag{3.4}$$

where G_0 is a probability distribution on $\mathbb{X} \times \mathbb{X}$. So that

$$\tilde{p}_1(\cdot) = p_1(\cdot \times \mathbb{X}) \qquad \tilde{p}_2(\cdot) = p_2(\mathbb{X} \times \cdot). \tag{3.5}$$

In other words, $\tilde{p}_1$ and $\tilde{p}_2$ are deterministic transformation of p_1 and p_2, that are random probability measures on $\mathbb{X} \times \mathbb{X}$ that share the same atoms and have any dependence structure on the weights. We will also require that the weights are independent of the atoms, as in all the constructions mentioned in Section 1.4. Notice that in this case p_1 and p_2 actually constitute a mSSP (see Chapter 2). Finally, for ease of exposition, we are going to take $\tilde{p}_1$ and $\tilde{p}_2$ with the same marginal distribution, even if this is not strictly necessary.

Almost sure discreteness implies that a sample from the random probability measure $\tilde{p}_1$ (or $\tilde{p}_2$) will display ties with positive probability. The probability of a tie, i.e. a coincidence of any two observations i and j in the same sample, is

$$\beta := \mathbb{P}[X_i = X_j] = \sum_{k\geq 1} \mathbb{E}[\bar{J}_k^2] = \sum_{k\geq 1} \mathbb{E}[\bar{W}_k^2] = \mathbb{P}[Y_i = Y_j] \tag{3.6}$$

with $(\bar{J}_k)_{k\geq 1}$ and $(\bar{W}_k)_{k\geq 1}$ equal in distribution since we are assuming, for simplicity, that $\tilde{p}_1$ and $\tilde{p}_2$ are equal in distribution. When considering jointly the two samples, the concept of tie can be replaced by the one of *hyper-tie*, that is two observations in different samples coinciding with components having the same label. According to (3.1), its probability is given by

$$\gamma := \sum_{k\geq 1} \mathbb{P}[X_i = \theta_k,\, Y_j = \phi_k] = \sum_{k\geq 1} \mathbb{E}[\bar{J}_k \bar{W}_k]. \tag{3.7}$$

Sampling from components with the same label is equivalent to sampling the same atom at the level of the underlying (p_1, p_2) in (3.4). Clearly, when the atoms are shared between $\tilde{p}_1$ and $\tilde{p}_2$, i.e. $G_0(\mathrm{d}\theta, \mathrm{d}\phi) = P_0(\mathrm{d}\theta)\delta_{\{\theta\}}(\mathrm{d}\phi)$, as it happens for instance with hierarchical processes (see Camerlenghi et al., 2019b), a hyper-tie corresponds to an actual tie between observations in different samples.

Lemma 3.3. *Consider $(\tilde{p}_1, \tilde{p}_2)$ as in (3.5). Then $0 \leq \gamma \leq \beta$ and $\beta = \gamma$ if and only if $\bar{W}_k \overset{a.s.}{=} \bar{J}_k$ for any k.*

Proof. Recall that

$$\beta := \mathbb{E}[\bar{J}_k^2] = \sum_{k\geq 1} \mathbb{E}[\bar{W}_k^2] \qquad \gamma := \sum_{k\geq 1} \mathbb{E}[\bar{J}_k \bar{W}_k].$$

Since

$$\mathbb{E}[\bar{J}_k \bar{W}_k] \leq \sqrt{\mathbb{E}[\bar{J}_k^2]\mathbb{E}[\bar{W}_k^2]} = \mathbb{E}[\bar{J}_k^2]$$

it follows that $\gamma \leq \beta$. Moreover, the equality holds if and only if $\bar{J}_k \overset{a.s}{=} a_k + \bar{W}_k$, for any k, with $a_k \in \mathbb{R}$, however the equality of marginal distributions implies $a_k = 0$. □

Hyper-ties play a crucial role in measuring the dependence between observables across groups, as the ties do for the dependence between observables within groups. Indeed, consider the following specification

$$X_i \mid \tilde{p}_1 \overset{\text{i.i.d.}}{\sim} \tilde{p}_1, \qquad Y_j \mid \tilde{p}_2 \overset{\text{i.i.d.}}{\sim} \tilde{p}_2, \qquad (\tilde{p}_1, \tilde{p}_2) \sim Q, \tag{3.8}$$

where Q is the law of any process described in (3.5). The next Proposition provides the correlation between observations, within and across groups.

Proposition 3.1. *Let $(\tilde{p}_1, \tilde{p}_2)$ be as in (3.5) and consider model (3.8).Then*

$$Corr(X_i, X_{i'}) = Corr(Y_j, Y_{j'}) = \beta \qquad i \neq i' \text{ and } j \neq j'$$

and

$$Corr(X_i, Y_j) = \gamma\, \rho_0 \qquad \text{for all } i, j$$

where ρ_0 is the correlation between two random variables jointly sampled from G_0.

Proof. As far as the correlation across sequences is concerned, we start by computing the moments, denoting $c = \mathbb{E}[\theta_j] = \mathbb{E}[\phi_k]$,

$$\mathbb{E}[X_i Y_j] = \sum_{j\geq 1}\sum_{k\geq 1} \mathbb{E}[\bar{J}_j \bar{W}_k]\mathbb{E}[\theta_j \phi_k] = c^2 \sum_{j\neq k} \mathbb{E}[\bar{J}_j \bar{W}_k] + \sum_{k\geq 1} \mathbb{E}[\bar{J}_k \bar{W}_k]\mathbb{E}[\theta_k \phi_k],$$

and

$$\mathbb{E}[X_i]\mathbb{E}[Y_j] = c^2 \sum_{j\geq 1}\sum_{k\geq 1} \mathbb{E}[J_j]\mathbb{E}[W_k].$$

Adding and subtracting $c^2 \sum_{k\geq 1} \mathbb{E}[\bar{J}_k \bar{W}_k]$ we get

$$\begin{aligned}\text{Cov}(X_i, Y_j) &= \text{Cov}(\theta, \phi) \sum_{k\geq 1} \mathbb{E}[\bar{J}_k \bar{W}_k] + c^2 \sum_{j\geq 1}\sum_{k\geq 1} \mathbb{E}[\bar{J}_j \bar{W}_k] - c^2 \sum_{j\geq 1}\sum_{k\geq 1} \mathbb{E}[J_j]\mathbb{E}[W_k] \\ &= \gamma\, \text{Cov}(\theta, \phi) + c^2 - c^2,\end{aligned}$$

since $\sum_{j\geq 1}\bar{J}_j = \sum_{k\geq 1}\bar{W}_k = 1$ almost surely. The result follows noticing that $\text{Var}(X_i) = \text{Var}(\theta)$ and $\text{Var}(Y_j) = \text{Var}(\phi)$.
Correlation within each sequence can be derived as a particular case of the above computations, with $\gamma = \beta$ and $\rho_0 = 1$. □

Thus, the correlation between observations in the *same* sample is the probability of a tie; instead, correlation between observations from *different* samples is given by the probability of a hyper-tie, multiplied by the correlation between atoms. It is clear that the latter can be negative, suitably choosing the joint distribution of the atoms; negative correlation is induced when G_0 exhibits a repulsive behaviour. Thus, choosing G_0 appropriately, for instance as a bivariate normal, it is easy to tune the correlation with the available prior knowledge. The following corollary shows the values that can be attained, once the marginal law is specified.

Corollary 3.1. *Let $(\tilde{p}_1, \tilde{p}_2)$ be as in (3.5) and consider model (3.8). If the marginal distribution of $\tilde{p}_1$ and $\tilde{p}_2$ is fixed, then*

$$\textit{Corr}(X_i, Y_j) \in [-\beta, \beta],$$

and the extreme values are attained if and only if the jumps are equal and $\rho_0 = \pm 1$.

Proof. It is clear that $\gamma \geq 0$. Then, we need to maximize $\mathbb{E}[\bar{J}_k, \bar{W}_k]$ for any $k \geq 1$. Since $\tilde{p}_1$ and $\tilde{p}_2$ have fixed marginals, it is equivalent to maximize $\text{Corr}(\bar{J}_k, \bar{W}_k)$. It is well-known that the correlation is the greatest when $\bar{J}_k \overset{\text{a.s.}}{=} \bar{W}_k + c$, with fixed c; since they share the same marginals, $\bar{J}_k \overset{\text{a.s.}}{=} \bar{W}_k$ and the result follows. □

Interestingly, notice that the extreme case of $\text{Corr}(X_i, Y_j) = \beta$ is attained with equal weights and atoms and corresponds to full *exchangeability*. Null correlation, instead, is attained when atoms are uncorrelated or when there is probability zero of hyper-ties. Lastly, maximum negative correlation $\text{Corr}(X_i, Y_j) = -\beta$, attained with equal weights and perfectly negatively correlated atoms, can be thought as the *opposite* case with respect to exchangeability, at least in terms of correlation. Ties and hyper-ties play a similar role also in the predictive structure, as the next Lemma shows.

Lemma 3.4. *Let $(\tilde{p}_1, \tilde{p}_2)$ be as in (3.2) and consider model (3.8) and let $A, B \in \mathcal{X}$. Then*

$$\mathbb{P}\left(X_1 \in A, X_2 \in B\right) = \beta P_0(A \cap B) + (1-\beta)\, P_0(A) P_0(B).$$

and

$$\mathbb{P}\left(X_1 \in A, Y_1 \in B\right) = \gamma G_0(A \times B) + (1-\gamma)\, P_0(A) P_0(B).$$

where P_0 is the marginal distribution on $\mathbb{X}$ obtained from G_0.

Proof. Proof follows trivially, noticing that p_1 and p_2 form a mSSP (cf. Definition 2.7). □

The result is indeed quite intuitive. If X_1 and Y_1 form a hyper-tie (with probability γ) they come from the same pair of atoms and need to be sampled jointly; otherwise they refer to different atoms and are sampled independently. The same happens inside each group, where X_1 and X_2 are equal with probability β.

Example 3.1 (Hierarchical Dirichlet process)**.** The hierarchical Dirichlet process (Teh et al., 2006) is characterized by the hierarchical representation

$$\tilde{p}_i \mid \tilde{p}_0 \overset{\text{i.i.d.}}{\sim} \text{DP}(\theta, \tilde{p}_0), \quad \tilde{p}_0 \sim \text{DP}(\theta_0, P_0),$$

where P_0 is a diffuse measure and $\text{DP}(\alpha, H)$ denotes the law of a Dirichlet process with concentration parameter $\alpha > 0$ and baseline distribution H. Since the $\tilde{p}_i$'s share the atoms, an hyper-tie corresponds to an actual tie between observations in different samples, so that with simple computations we get

$$\beta = \text{Corr}(X_i, X_j) = 1 - \frac{\theta\theta_0}{(1+\theta)(1+\theta_0)},$$

$$\gamma = \text{Corr}(X_i, Y_j) = \frac{1}{1+\theta_0}.$$

Thus, the correlation is forced to be positive, with θ_0 tuning the dependence; see Example 1 in Camerlenghi et al. (2019b) for more details.

Example 3.2 (Hierarchical Pitman-Yor)**.** *If*

$$\tilde{p}_i \mid \tilde{p}_0 \overset{\text{i.i.d.}}{\sim} PY(\sigma, \theta, \tilde{p}_0), \quad \tilde{p}_0 \sim PY(\sigma_0, \theta_0, P_0),$$

where P_0 is a diffuse measure, the $\tilde{p}_i$'s share the entire sequence of atoms and an hyper-tie corresponds to an actual tie between observations in different samples, so that with simple computations we get

$$\beta = Corr(X_i, X_j) = 1 - \frac{(\theta+\sigma)(\theta_0+\sigma+_0)}{(1+\theta)(1+\theta_0)},$$

$$\gamma = Corr(X_i, Y_j) = \frac{1-\sigma_0}{1+\theta_0}.$$

Thus, also in this case the correlation is forced to be positive, with θ_0 and σ_0 tuning the dependence across samples.

It should now be clear that it is crucial to suitably set γ in order to tune the level of dependence. However, its knowledge in closed form is restricted to few cases and for general dependent processes cannot be easily derived. Therefore, it seems we are facing a trade-off. On the one hand we have dependent processes based on the stick-breaking representation, that guarantee high flexibility while sacrificing the availability of analytical results; on the other hand we have constructions based on completely random measures, for which an

extensive theory has been developed, but that are not as manageable for introducing dependence, since all the existing instances produce non-negative correlation across samples. In the following we will show how to combine the best of both worlds through the n-FuRBI, defined in next section: they are flexible processes that can attain any value for the correlation between the observables and for which marginal urn schemes can be derived. Their construction is based on completely random measures and completely random vectors, reviewed in Section 1.3 and Section 1.4.1.

3.4 Full range borrowing of information NRMIs

In this section we introduce n-FuRBI, starting from a completely random vector. Again, for simplicity, we consider only the case of two samples with the same a priori marginal distribution, however the extension is straightforward.

Definition 3.1. *Consider a CRV (μ_1, μ_2) on $(\mathbb{X} \times \mathbb{X}, \mathcal{X} \otimes \mathcal{X})$ with Lévy intensity*

$$v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x_2) = \rho(\mathrm{d}s_1, \mathrm{d}s_2)\alpha(\mathrm{d}x_1, \mathrm{d}x_2),$$

where $\alpha(\mathrm{d}x_1, \mathrm{d}x_2) = \theta G_0(\mathrm{d}x_1, \mathrm{d}x_2)$ and G_0 is a non-atomic probability measure over $(\mathbb{X}\times\mathbb{X}, \mathcal{X} \otimes \mathcal{X})$ such that $G_0(\cdot \times \mathbb{X}) = G_0(\mathbb{X} \times \cdot) = P_0(\cdot)$. Then $\tilde{\mu}_1$ and $\tilde{\mu}_2$ defined as

$$\tilde{\mu}_1(\cdot) = \mu_1(\mathbb{X} \times \cdot) \qquad \tilde{\mu}_2(\cdot) = \mu_2(\cdot \times \mathbb{X}) \tag{3.9}$$

are called FuRBI completely random measures (FuRBI CRMs) with underlying Lévy intensity v. The normalized versions $\tilde{p}_j(\cdot) = \frac{\tilde{\mu}_j(\cdot)}{\tilde{\mu}_j(\mathbb{X})}$ for $j = 1, 2$ are said n-FuRBIs.

Essentially, firstly a pair of random measures endowed with the same locations is constructed on the product space $\mathbb{X} \times \mathbb{X}$; as a second step, the coordinates of each pair of atoms are split. Notice that in general FuRBI CRMs are not CRVs, because the joint sampling of the atoms forbids the independence of increments of the vector. In this regard it may be useful to underline the difference between FuRBI-CRMs and classical CRVs, exploiting the difference in the underlying PPs. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ form a CRV, then there exists a PP, N, on $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{X}$ with a certain intensity $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x)$ such that

$$\tilde{\mu}_1(\mathrm{d}x) = \int_{\mathbb{R}^+\times\mathbb{R}^+} s_1 \, N(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x)$$

and

$$\tilde{\mu}_2(\mathrm{d}x) = \int_{\mathbb{R}^+\times\mathbb{R}^+} s_2 \, N(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x).$$

Instead, if $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are FuRBI CRMs, then there exists a PP, N, on $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{X} \times \mathbb{X}$ with

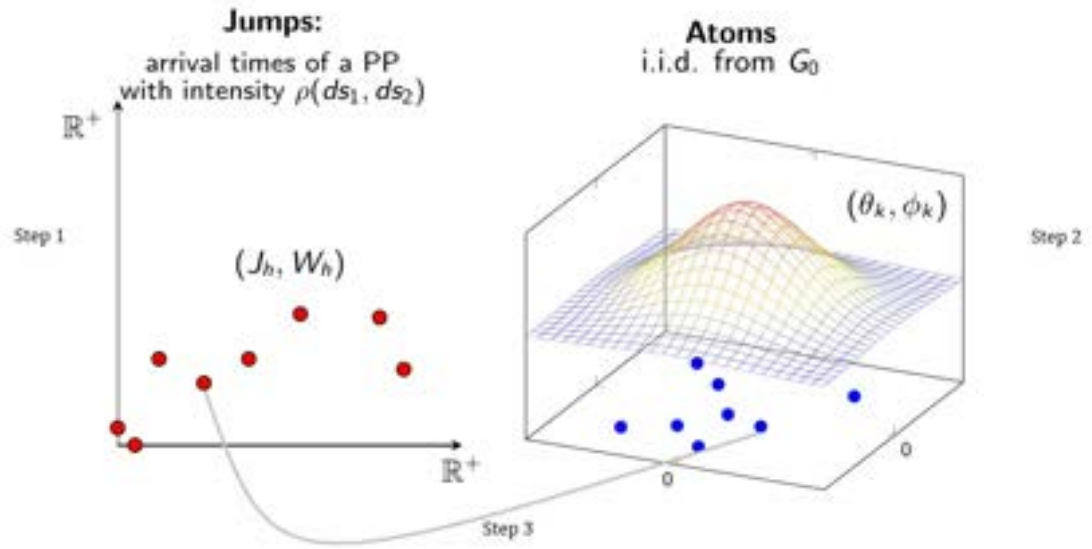


Figure 3.1: Simulation steps for a bivariate vector of FuRBI CRMs with underlying intensity $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x) = \rho(\mathrm{d}s_1, \mathrm{d}s_2)\, G_0(\mathrm{d}x_1, \mathrm{d}x_2)$. At step 1 arrival times of a PP on $\mathbb{R}^+ \times \mathbb{R}^+$ with intensity $\rho(\mathrm{d}s_1, \mathrm{d}s_2)$ are sampled, at step 2 a bivariate atom from $G_0(\mathrm{d}x_1, \mathrm{d}x_2)$ is sampled for each couple of jumps, at step 3 the i.i.d. atoms are associated to the couples of jumps. To get the correspondent NRMIs, it is enough to normalize the two sequences of jumps.

a certain intensity $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x_2)$ such that

$$\tilde{\mu}_1(\mathrm{d}x) = \int_{\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{X}} s_1 \, N(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x, \mathrm{d}x_2)$$

and

$$\tilde{\mu}_2(\mathrm{d}x) = \int_{\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{X}} s_2 \, N(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x).$$

Figure 3.1 displays the sampling steps to generate FuRBI CRMs starting from the Poisson process, if compared to Figure 1.2 it highlights that the main difference between the two constructions lies in the dimension of the atoms to be sampled. Finally, looking at the series representations of CRMs in Theorem 1.8, we have that the dependence between FuRBI-CRMs is driven both by the dependence between the weights encoded by $\rho(\mathrm{d}s_1, \mathrm{d}s_2)$ and by the dependence of the atoms encoded by G_0, as shown in the next proposition.

Proposition 3.2. *Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be FuRBI CRMs as defined in Definition 3.1, then $\forall\, A, B \in \mathcal{X}$,*

$$\begin{aligned} Cov(\tilde{\mu}_1(A), \tilde{\mu}_2(B)) = & \, Cov\left(\tilde{\mu}_1(\mathbb{X}), \tilde{\mu}_2(\mathbb{X})\right) \, P_0(A)P_0(B) + \\ & + \sum_{k \geq 1} \mathbb{E}[J_k \, W_k] \left(G_0(A \times B) - P_0(A)P_0(B)\right) \end{aligned}$$

Proof.

$$\mathbb{E}[\tilde{\mu}_1(A)\,\tilde{\mu}_2(B)] = \mathbb{E}\left[\sum_{k\geq 1} J_k \delta_{\theta_k}(A) \sum_{k'\geq 1} W_{k'}\delta_{\phi_{k'}}(B)\right] = \sum_{k\geq 1}\sum_{k'\geq 1}\mathbb{E}[J_k W_{k'}\delta_{\theta_k}(A)\delta_{\phi_{k'}}(B)]$$

Moreover, by the homogeneity of the two CRMs, we have

$$\begin{aligned}\mathbb{E}[\tilde{\mu}_1(A)\,\tilde{\mu}_2(B)] &= \sum_{k\geq 1}\sum_{k'\geq 1}\mathbb{E}[J_k W_{k'}]\mathbb{E}[\delta_{\theta_k}(A)\delta_{\phi_{k'}}(B)] \\ &= \sum_{k\geq 1}\sum_{k'\geq 1}\mathbb{E}[J_k W_{k'}]\left(G_0(A\times B)\mathbb{1}_{\{k=k'\}} + P_0(A)Q_0(B)\mathbb{1}_{\{k\neq k'\}}\right)\end{aligned}$$

While the product between the marginal first moments is

$$\begin{aligned}\mathbb{E}[\tilde{\mu}_1(A)]\,\mathbb{E}[\tilde{\mu}_2(B)] &= \sum_{k\geq 1}\mathbb{E}[J_k]\mathbb{E}[\delta_{\theta_k}]\sum_{k'\geq 1}\mathbb{E}[W_{k'}]\,E[\delta_{\phi_{k'}}] \\ &= \sum_{k\geq 1}\sum_{k'\geq 1}\mathbb{E}[J_k]\,\mathbb{E}[W_{k'}]\,P_0(A)Q_0(B)\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{Cov}[\tilde{\mu}_1(A),\tilde{\mu}_2(B)] &= \sum_{k\geq 1}\mathbb{E}[J_k W_k]\,G_0(A\times B) - \sum_{k\geq 1}\mathbb{E}[J_k]\,\mathbb{E}[W_k]\,P_0(A)Q_0(B) \\ &+ \sum_{k\geq 1}\sum_{k'\neq k}\mathbb{E}[J_k W_{k'}]P_0(A)Q_0(B) - \sum_{k\geq 1}\sum_{k'\neq k}\mathbb{E}[J_k]\,\mathbb{E}[W_{k'}]\,P_0(A)Q_0(B)\end{aligned}$$

Notice now that

$$\begin{aligned}\mathrm{Cov}\left(\sum_{k\geq 1}J_k,\,\sum_{k'\geq 1}W_{k'}\right) &= \sum_{k\geq 1}\sum_{k'\geq 1}\mathrm{Cov}(J_k,W_{k'}) \\ &= \sum_{k\geq 1}\mathrm{Cov}(J_k,W_k) + \sum_{k\geq 1}\sum_{k'\neq k}\mathrm{Cov}(J_k,W_{k'})\end{aligned}$$

Therefore we have

$$\begin{aligned}&\mathrm{Cov}[\tilde{\mu}_1(A),\,\tilde{\mu}_2(B)] = \sum_{k\geq 1}\mathbb{E}[J_k W_k]\,G_0(A\times B) - \sum_{k\geq 1}\mathbb{E}[J_k]\,\mathbb{E}[W_k]\,P_0(A)Q_0(B) \\ &+ P_0(A)Q_0(B)\left(\mathrm{Cov}\left(\sum_{k\geq 1}J_k,\,\sum_{k'\geq 1}W_{k'}\right) - \sum_{k\geq 1}\mathrm{Cov}(J_k,W_k)\right) \\ &= \mathrm{Cov}\left(\sum_{k\geq 1}J_k,\,\sum_{k'\geq 1}W_{k'}\right)P_0(A)Q_0(B) + \sum_{k\geq 1}\mathbb{E}[J_k W_k]\;(G_0(A\times B) - P_0(A)Q_0(B))\end{aligned}$$

□

The theorem shows that the covariance between two FuRBI CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$ when evaluated on some Borel set $A \in \mathcal{X}$ is given by the sum of a first term that depends on the covariance between $\tilde{\mu}_1(\mathbb{X})$ and $\tilde{\mu}_2(\mathbb{X})$ and a second terms that depends on the difference $G_0(A \times A) - P_0(A)P_0(A)$. If the atoms are independent so that $G_0(A \times A) = P_0(A)^2$, the covariance simplifies to

$$\text{Cov}(\tilde{\mu}_1(A), \tilde{\mu}_2(A)) = \text{Cov}\left(\tilde{\mu}_1(\mathbb{X}), \tilde{\mu}_2(\mathbb{X})\right) P_0(A)^2.$$

While FuRBI CRMs do not form a CRV, they admit a representation in terms of a CRV in the product space, namely (μ_1, μ_2) in Definition 3.1, and this is useful to characterize the joint law of the FuRBI CRMs, as shown in the following Proposition.

Proposition 3.3. *Consider a vector of FuRBI CRMs $(\tilde{\mu}_1, \tilde{\mu}_2)$. Then*

- *Both $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are CRMs with $\rho(\mathrm{d}s) = \int_{\mathbb{R}_+} \rho(\mathrm{d}s_1, \mathrm{d}s)$ and intensity $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\theta P_0(\mathrm{d}x)$.*
- *For any $A, B \in \mathcal{X}$, the following equality holds*

$$\begin{aligned}&\mathbb{E}\left[e^{-\lambda_1\tilde{\mu}_1(A)-\lambda_2\tilde{\mu}_2(B)}\right] = \\ &\quad = exp\left\{-G_0(A \times B^c)\psi(\lambda_1) - G_0(A^c \times B)\psi(\lambda_2) - G_0(A \times B)\psi_b(\lambda_1, \lambda_2)\right\},\end{aligned} \tag{3.10}$$

  where ψ denotes the common marginal Laplace exponent and ψ_b the joint Laplace exponent of (μ_1, μ_2).

- *The joint law of $(\tilde{\mu}_1, \tilde{\mu}_2)$ is characterized by the joint Lévy intensity of (μ_1, μ_2).*

Proof. The first point follows from the Lévy-Khintchine representation for a CRV. As regards the second point, we have

$$\begin{aligned}\mathbb{E}\left[\exp\{-\lambda_1\tilde{\mu}_1(A) - \lambda_2\tilde{\mu}_2(B)\}\right] &= \mathbb{E}\left[\exp\{-\lambda_1\mu_1(A \times \mathbb{X}) - \lambda_2\mu_2(\mathbb{X} \times B)\}\right] \\ &= \mathbb{E}\big[\exp\{-\lambda_1\mu_1(A \times B^c) - \lambda_1\mu_1(A \times B) \\ &\quad - \lambda_2\mu_2(A^c \times B) - \lambda_2\mu_2(A \times B)\}\big]\end{aligned} \tag{3.11}$$

By independence of the increments, we have $\mu_1(C)$ independent from $\mu_2(D)$, with $C \cap D = \emptyset$, so the right hand side reads

$$\begin{aligned}\mathbb{E}\big[\exp\{-\lambda_1\tilde{\mu}_1(A) - \lambda_2\tilde{\mu}_2(B)\}\big] &= \mathbb{E}\left[\exp\{-\lambda_1\mu_1(A \times B^c)\}\right]\mathbb{E}\left[\exp\{-\lambda_2\mu_2(A^c \times B)\}\right] \\ &\quad \times \mathbb{E}\left[\exp\{-\lambda_1\mu_1(A \times B) - \lambda_2\mu_2(A \times B)\}\right]\end{aligned} \tag{3.12}$$

□

Moving to the normalized measures, notice that the n-FuRBIs $(\tilde{p}_1, \tilde{p}_2)$ admit a representation as in (3.2) and (3.4). Next Proposition shows that the β and γ associated to any couple of n-FuRBI can be computed through their Laplace exponents.

Proposition 3.4. *Consider $(\tilde{p}_1, \tilde{p}_2)$ n-FuRBI. Then the probability of a tie is given by*

$$\beta = -\int_{\mathbb{R}_+} u \left\{ \frac{\mathrm{d}^2}{\mathrm{d}u^2} \psi(u) \right\} e^{-\psi(u)} \, \mathrm{d}u,$$

while the probability of a hyper-tie reads

$$\gamma = -\int_{\mathbb{R}_+^2} \left\{ \frac{\partial^2}{\partial u_1 \partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \, \mathrm{d}u_1 \mathrm{d}u_2.$$

In order to compute the probability of a hyper-tie and prove Proposition 3.4, we will show

$$\mathbb{P}\left(X \in A, Y \in B\right) = P_0(A) P_0(B) \left(1 - \delta\right) + G_0(A \times B)\delta,$$

with $A, B \in \mathcal{X}^2$ and

$$\delta := -\int_{\mathbb{R}_+^2} \left\{ \frac{\partial^2}{\partial u_1 \partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \, \mathrm{d}u_1 \mathrm{d}u_2.$$

If follows then that the probability of a hyper-tie will be given exactly by δ, which coincides with γ. We start with three technical Lemmas.

Lemma 3.5. *It holds*

$$\int_{\mathbb{R}_+^2} \left\{ \frac{\partial}{\partial u_1} \psi_b(u_1, u_2) \right\} \left\{ \frac{\partial}{\partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \, \mathrm{d}u_1 \mathrm{d}u_2 = 1 - \delta. \tag{3.13}$$

Proof. Integrating by parts

$$\begin{aligned}
&\int_0^\infty \left\{ \frac{\partial}{\partial u_1} \psi_b(u_1, u_2) \right\} \left\{ \frac{\partial}{\partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \, \mathrm{d}u_1 = \\
&\quad = -\int_0^\infty \left\{ \frac{\partial}{\partial u_2} \psi_b(u_1, u_2) \right\} \left\{ \frac{\partial}{\partial u_1} e^{-\psi_b(u_1, u_2)} \right\} \mathrm{d}u_1 = \\
&\quad = \left[\left[-\left\{ \frac{\partial}{\partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \right]_0^\infty + \int_0^\infty \left\{ \frac{\partial^2}{\partial u_1 \partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \, \mathrm{d}u_1 \right] = \\
&\quad = \left[\left\{ \frac{\partial}{\partial u_2} \psi_b(0, u_2) \right\} e^{-\psi_b(0, u_2)} + \int_0^\infty \left\{ \frac{\partial^2}{\partial u_1 \partial u_2} \psi_b(u_1, u_2) \right\} e^{-\psi_b(u_1, u_2)} \, \mathrm{d}u_1 \right].
\end{aligned} \tag{3.14}$$

Notice

$$\int_0^\infty \left\{ \frac{\mathrm{d}}{\mathrm{d}u_2} \psi_b(0, u_2) \right\} e^{-\psi_b(0, u_2)} \, \mathrm{d}u_2 = 1, \tag{3.15}$$

by the Fundamental Theorem of Calculus. Thus the result follows immediately. □

Lemma 3.6. *Let $C \in \mathcal{X}^2$. Then*

$$\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\mu_1(C)\mu_2(C)\right] \mathrm{d}u_1\mathrm{d}u_2 = G_0(C)^2\,(1-\delta) + G_0(C)\delta \tag{3.16}$$

Proof. By independence of the increments it follows

$$\begin{aligned}
&\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\mu_1(C)\mu_2(C)\right] \mathrm{d}u_1\mathrm{d}u_2 = \int_{\mathbb{R}_+^2} \mathbb{E}\big[e^{-u_1\mu_1(C)-u_2\mu_2(C)-u_1\mu_1(C^c)-u_2\mu_2(C^c)} \\
&\quad \mu_1(C)\mu_2(C)\big] \mathrm{d}u_1\mathrm{d}u_2 = \int_{\mathbb{R}_+^2} \mathbb{E}\big[e^{-u_1\mu_1(C)-u_2\mu_2(C)}\mu_1(C)\mu_2(C)\big]\mathbb{E}\big[e^{-u_1\mu_1(C^c)-u_2\mu_2(C^c)}\big] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\quad = \int_{\mathbb{R}_+^2} \mathbb{E}\left[\frac{\partial}{\partial u_1}\frac{\partial}{\partial u_2}e^{-u_1\mu_1(C)-u_2\mu_2(C)}\right]\mathbb{E}\big[e^{-u_1\mu_1(C^c)-u_2\mu_2(C^c)}\big] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\quad = \int_{\mathbb{R}_+^2} \frac{\partial}{\partial u_1}\frac{\partial}{\partial u_2}\left\{\mathbb{E}\left[e^{-u_1\mu_1(C)-u_2\mu_2(C)}\right]\right\}\mathbb{E}\big[e^{-u_1\mu_1(C^c)-u_2\mu_2(C^c)}\big] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\quad\quad = \int_{\mathbb{R}_+^2} \frac{\partial}{\partial u_1}\frac{\partial}{\partial u_2}\left\{e^{-G_0(C)\psi_b(u_1,u_2)}\right\}e^{-G_0(C^c)\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2 = \\
&\quad\quad = \int_{\mathbb{R}_+^2} \frac{\partial}{\partial u_1}\left\{-G_0(C)\frac{\partial}{\partial u_2}\psi_b(u_1,u_2)e^{-G_0(C)\psi_b(u_1,u_2)}\right\}e^{-G_0(C^c)\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2 = \\
&= \int_{\mathbb{R}_+^2} \left(G_0(C)^2\frac{\partial}{\partial u_1}\psi_b(u_1,u_2)\frac{\partial}{\partial u_2}\psi_b(u_1,u_2)\right)e^{-G_0(C)\psi_b(u_1,u_2)}e^{-G_0(C^c)\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2 + \\
&+ \int_{\mathbb{R}_+^2} \left(-G_0(C)\frac{\partial}{\partial u_1\partial u_2}\psi_b(u_1,u_2)\right)e^{-G_0(C)\psi_b(u_1,u_2)}e^{-G_0(C^c)\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2 = \\
&\quad\quad = \int_{\mathbb{R}_+^2} \left(G_0(C)^2\frac{\partial}{\partial u_1}\psi_b(u_1,u_2)\frac{\partial}{\partial u_2}\psi_b(u_1,u_2)\right)e^{-\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2 + \\
&\quad\quad + \int_{\mathbb{R}_+^2} \left(-G_0(C)\frac{\partial}{\partial u_1\partial u_2}\psi_b(u_1,u_2)\right)e^{\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2
\end{aligned}$$

By Lemma 3.5 it follows

$$\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\mu_1(C)\mu_2(C)\right] \mathrm{d}u_1\mathrm{d}u_2 = G_0(C)^2\,(1-\delta) + G_0(C)\delta$$

□

Lemma 3.7. *Let $C, D \in \mathcal{X}$ such that $C \cap D = \emptyset$. Then*

$$\begin{aligned}
&\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\mu_1(C)\mu_2(D)\right] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\quad = G_0(C)G_0(D)\,(1-\delta)
\end{aligned} \tag{3.17}$$

Proof. Denote $Y = (C \cup D)^c$. Since C and D are disjoint, by independence of increments it holds

$$\begin{aligned}
&\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\mu_1(C)\mu_2(D)\right] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(C\cup D)-u_2\mu_2(C\cup D)}\mu_1(C)\mu_2(D)\right] \mathbb{E}\left[e^{-u_1\mu_1(Y)-u_2\mu_2(Y)}\right] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(C)-u_2\mu_2(C)}\mu_1(C)\right] \mathbb{E}\left[e^{-u_1\mu_1(D)-u_2\mu_2(D)}\mu_2(D)\right] \times \\
&\qquad \times \mathbb{E}\left[e^{-u_1\mu_1(Y)-u_2\mu_2(Y)}\right] \mathrm{d}u_1\mathrm{d}u_2 = \\
&\int_{\mathbb{R}_+^2} \frac{\partial}{\partial u_1}\left\{e^{-G_0(C)\psi_b(u_1,u_2)}\right\} \frac{\partial}{\partial u_2}\left\{e^{-G_0(D)\psi_b(u_1,u_2)}\right\} \times \\
&\qquad \times e^{-G_0(Y)\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2 = \\
&\qquad G_0(C)G_0(D)\int_{\mathbb{R}_+^2} \left(\frac{\partial}{\partial u_1}\psi_b(u_1,u_2)\frac{\partial}{\partial u_2}\psi_b(u_1,u_2)\right) e^{-\psi_b(u_1,u_2)} \mathrm{d}u_1\mathrm{d}u_2
\end{aligned} \tag{3.18}$$

Then apply Lemma 3.5. □

Finally we can derive the proof of Proposition 3.4.

Proof of Proposition 3.4. Let $A, B \in \mathcal{X}^2$, then we have

$$\begin{aligned}
\mathbb{P}\left(X \in A, Y \in B\right) &= \mathbb{E}\left[\frac{\tilde{\mu}_1(A)}{\tilde{\mu}_1(\mathbb{X})}\frac{\tilde{\mu}_2(B)}{\tilde{\mu}_2(\mathbb{X})}\right] = \mathbb{E}\left[\frac{\mu_1(A\times\mathbb{X})}{\mu_1(\mathbb{X}\times\mathbb{X})}\frac{\mu_2(\mathbb{X}\times B)}{\mu_2(\mathbb{X}\times\mathbb{X})}\right] = \\
&= \int_{\mathbb{R}_+^2} \mathbb{E}\left[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\mu_1(A\times\mathbb{X})\mu_2(\mathbb{X}\times B)\right] \mathrm{d}u_1\mathrm{d}u_2 = \\
&= \int_{\mathbb{R}_+^2} \mathbb{E}\Big[e^{-u_1\mu_1(\mathbb{X}\times\mathbb{X})-u_2\mu_2(\mathbb{X}\times\mathbb{X})}\big[\mu_1(A\times B)\mu_2(A\times B) + \\
&\quad \mu_1(A\times B)\mu_2(A^c\times B) + \mu_1(A\times B^c)\mu_2(A\times B) + \\
&\quad \mu_1(A\times B^c)\mu_2(A^c\times B)\big]\Big] \mathrm{d}u_1\mathrm{d}u_2
\end{aligned} \tag{3.19}$$

We compute each integral separately applying Lemmas 3.6 and 3.7 and we get

$$\begin{aligned}
\mathbb{P}\left(X \in A, Y \in B\right) &= G_0(A\times\mathbb{X})G_0(\mathbb{X}\times B)\left(1-\delta\right) + G_0(A\times B)\delta \\
&= P_0(A)P_0(B)\left(1-\delta\right) + G_0(A\times B)\delta,
\end{aligned} \tag{3.20}$$

as desired. Then the probability of a tie in the product space is given exactly by δ, which equals γ. The probability of a tie is given by the particular case $\psi_b(u_1, u_2) = \psi(u_1 + u_2)$,

since

$$-\int_{\mathbb{R}_+^2}\left\{\frac{\partial^2}{\partial u_1\partial u_2}\psi_b(u_1+u_2)\right\}e^{-\psi_b(u_1+u_2)}\,\mathrm{d}u_1\mathrm{d}u_2=-\int_0^\infty\int_0^u\,\mathrm{d}v\left\{\frac{\partial^2}{\partial u^2}\psi_b(u)\right\}e^{-\psi_b(u)}\,\mathrm{d}u,$$

with the change of variables $u=u_1+u_2$ and $v=u_1$. □

Thus, the crucial value of γ can be obtained by computing, analytically or numerically, a bivariate integral. The two results above show a recurrent trait of our approach: interesting quantities will be usually rewritten in terms of the original CRV, in order to exploit its analytical tractability. We conclude this section with two examples of FuRBI CRMs, that also show how some existing constructions can be obtained as special cases.

Example 3.3 (FuRBI CRMs with equal jumps)**.** *Consider the underlying Lévy intensity*

$$v(\mathrm{d}s_1,\mathrm{d}s_2,\mathrm{d}x_1,\mathrm{d}x_2)=\rho(\mathrm{d}s_1)\delta_{s_1}(\mathrm{d}s_2)\,\theta\,G_0(\mathrm{d}x_1,\mathrm{d}x_2).$$

The series representation of the corresponding FuRBI CRMs is

$$\tilde{\mu}_1\overset{a.s.}{=}\sum_{k\geq1}W_k\delta_{\theta_k}\qquad\tilde{\mu}_2\overset{a.s.}{=}\sum_{k\geq1}W_k\delta_{\phi_k}\qquad\text{with }\ (\theta_k,\phi_k)\overset{i.i.d}{\sim}G_0.$$

Therefore, $\gamma=\beta$, that is the probability of a tie or a hyper-tie is the same.

Example 3.4 (Extended Compound FuRBI CRMs)**.** *Consider the joint underlying Lévy intensity*

$$v(\mathrm{d}s_1,\mathrm{d}s_2,\mathrm{d}x_1,\mathrm{d}x_2)=\int z^{-2}h(s_1/z,\,s_2/z)\,\mathrm{d}s_1\mathrm{d}s_2v^*(\mathrm{d}z)\,\theta\,G_0(\mathrm{d}x_1,\mathrm{d}x_2),$$

where h is a mass probability function or density function and v^* is a Lévy intensity that satisfies

$$\int z^{-2}\int\min\{1,||\boldsymbol{s}||\}h(s_1/z,\,s_2/z)\,\mathrm{d}s_1\mathrm{d}s_2v^*(\mathrm{d}z),\quad||\boldsymbol{s}||=\sqrt{s_1^2+s_2^2}.$$

The series representation of the corresponding FuRBI CRMs is

$$\tilde{\mu}_1\overset{a.s.}{=}\sum_{k\geq1}m_{1,k}W_k\delta_{\theta_k}\qquad\tilde{\mu}_2\overset{a.s.}{=}\sum_{k\geq1}m_{2,k}W_k\delta_{\phi_k}\qquad\text{with }\ (\theta_k,\phi_k)\overset{i.i.d}{\sim}G_0,$$

where $(m_{1,k},m_{2,k})\overset{iid}{\sim}h$. Notice that when G_0 is degenerate on the main diagonal, one retrieves compound random measure of Griffin & Leisen (2017).

3.4.1 Correlation structure between n-FuRBI

In order to analyze the dependence between $\tilde{p}_1$ and $\tilde{p}_2$ n-FuRBIs, it can be useful to compute the correlation of the random probability measures evaluated on the same Borel set A.

This quantity is of significant interest in the literature, since in all the existing CRM-based models it does not depend on the specific set considered (cf. Proposition 2.7); therefore it is sometimes used as a global measure of dependence. The next proposition illustrates the covariance structure between two n-FuRBIs.

Proposition 3.5. *Let $\tilde{p}_1$ and $\tilde{p}_2$ be n-FuRBI. Then $\forall\, A, B \in \mathcal{X}$ it holds*

$$Cov(\tilde{p}_1(A), \tilde{p}_2(B)) = \gamma \left[G_0(A \times B) - P_0(A)P_0(B)\right]$$

and

$$Corr(\tilde{p}_1(A), \tilde{p}_2(B)) = \frac{\gamma}{\beta} \frac{G_0(A \times B) - P_0(A)P_0(B)}{\sqrt{P_0(A)(1 - P_0(A))P_0(B)(1 - P_0(B))}}.$$

In the particular case of $A = B$ we have

$$Cov(\tilde{p}_1(A), \tilde{p}_2(A)) = \gamma \left[G_0(A \times A) - P_0(A)^2\right],$$

$$Corr(\tilde{p}_1(A), \tilde{p}_2(A)) = \frac{\gamma}{\beta} \frac{G_0(A \times A) - P_0(A)^2}{P_0(A)(1 - P_0(A))}.$$

Proof. Let $A, B \in \mathcal{X}^2$. By de Finetti's Theorem we know

$$\mathbb{E}\left[\tilde{p}_1(A)\tilde{p}_2(B)\right] = \mathbb{P}\left(X \in A, Y \in B\right),$$

and by (3.20) we have

$$\mathbb{E}\left[\tilde{p}_1(A)\tilde{p}_2(B)\right] = G_0(A \times \mathbb{X})G_0(\mathbb{X} \times B)\left(1 - \gamma\right) + G_0(A \times B)\gamma \tag{3.21}$$

Finally

$$\begin{aligned} \text{Cov}\left(\tilde{p}_1(A), \tilde{p}_2(B)\right) &= G_0(A \times \mathbb{X})G_0(\mathbb{X} \times B)\left(1 - \gamma\right) + G_0(A \times B)\gamma - G_0(A \times \mathbb{X})G_0(\mathbb{X} \times B) \\ &= \gamma \left[G_0(A \times B) - G_0(A \times \mathbb{X})G_0(\mathbb{X} \times B)\right]. \end{aligned} \tag{3.22}$$

The correlation follows dividing by the product of the standard deviations, that can be obtained by the above formula since

$$\begin{aligned} \text{Var}\left(\tilde{p}_1(A)\right) &= \text{Cov}\left(\tilde{p}_1(A), \tilde{p}_1(A)\right) = \beta\left[P_0(A) - P_0(A)^2\right] \\ &= \beta P_0(A)\left[1 - P_0(A)\right], \end{aligned}$$

as desired. □

Unlike what usually happens with existing models, the correlation between $\tilde{p}_1(A)$ and $\tilde{p}_2(A)$ can be negative and this happens when A is such that $G_0(A \times A) < P_0(A)^2$, that is when G_0 exhibits a repulsive behaviour. Moreover, the correlation depends on the specific

Borel set on which the two measures are evaluated and, therefore, it has to be interpreted as a local measure of dependence.

Example 3.5 (n-FuRBI with equal jumps)**.** *In this case we showed that $\beta = \gamma$. Therefore*

$$Corr\left(\tilde{p}_1(A), \tilde{p}_2(A)\right) = \frac{G_0(A \times A) - P_0(A)^2}{P_0(A)(1 - P_0(A))}. \tag{3.23}$$

Moreover, thanks to Lemma 3.3, once the joint law G_0 is fixed this is the highest possible correlation in absolute value.

As regards the correlation between the observables, that may be seen as more influential from a modelling perspective, the result of Proposition 3.1 holds.

Example 3.6 (Gamma n-FuRBI with equal jumps)**.** *It is the most general framework in terms of correlation: once the marginal law is fixed, any value in $[-\beta, \beta]$ can be attained. If the common marginal is given by a Dirichlet process, it reads*

$$Corr(X_i, Y_j) = \frac{\rho_0}{1 + \theta}.$$

Choosing appropriately ρ_0 and θ the entire spectrum $(-1, 1)$ becomes available.

3.5 σ-stable n-FuRBI

In this section we provide some details on the specific example of n-FURBIs with σ-stable marginals and underlying Lévy intensity obtained using Clayton's Lévy copula (see Ascolani et al., 2021).

Definition 3.2. *Consider a completely random vector (μ_1, μ_2) on $(\mathbb{X} \times \mathbb{X}, \mathcal{X} \otimes \mathcal{X})$ with Lévy intensity $v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x_2) = \rho(s_1,\, s_2)\mathrm{d}s_1\mathrm{d}s_2\, \alpha(\mathrm{d}x_1, \mathrm{d}x_2)$ such that*

$$\int_0^{+\infty} \rho(s_1,\, s)\mathrm{d}s_1 = \int_0^{+\infty} \rho(s,\, s_2)\mathrm{d}s_2 = \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma}\mathrm{d}s\, \alpha(\mathrm{d}x_1, \mathrm{d}x_2), \qquad 0 < \sigma < 1,$$

$\tilde{\mu}_1(\cdot) = \mu_1(\cdot \times \mathbb{X})$ and $\tilde{\mu}_2(\cdot) = \mu_2(\mathbb{X} \times \cdot)$ are called σ-stable FuRBI CRMs with underlying Lévy intensity v. The random probability measures $\tilde{p}_1$ and $\tilde{p}_2$ obtained normalizing two σ-stable FuRBI CRMs are called σ-stable n-FuRBIs.

In order to obtain a working model which makes use of σ-stable n-FuRBIs, the underlying Lévy intensity v has to be specified. A useful strategy to do so is to use Lévy copulas. See Section 1.4.1. A popular Lévy copula is Clayton's one, which is given by

$$C_\theta(x_1, x_2) = \{x_1^{-\theta} + x_2^{-\theta}\}^{-1/\theta}$$

The attractive feature of Clayton's copula is that it depends only on one parameter, θ, that fully characterizes the degree of dependence between the resulting CRMs μ_1 and μ_2. As consequence, when Clayton's copula is used to specify the law of two n-FuRBIs, θ controls the portion of dependence between $\tilde{p}_1$ and $\tilde{p}_2$ induced by the joint distribution of the weights. In particular when $\theta \to 0$ independence between $\tilde{p}_1$ and $\tilde{p}_2$ is approached, while the case of $\theta \to +\infty$ corresponds to maximal dependence induced by the weights, i.e., the two sequences of weights are equal with probability 1. Applying Clayton's Lévy copula to marginal Lévy σ-stables, one gets the following joint Lévy intensity (see Epifani & Lijoi, 2010)

$$v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x_2; \theta) = \frac{(1+\theta)\,\sigma\,(s_1\,s_2)^{\sigma\theta-1}}{\Gamma(1-\sigma)\,(s_1^{\sigma\theta}+s_2^{\sigma\theta})^{\frac{1}{\theta}+2}}\,\alpha(\mathrm{d}x_1, \mathrm{d}x_2) \tag{3.24}$$

Theorem 3.1. *Consider the sampling model $X_{i,j} \mid \tilde{p}_j \overset{ind}{\sim} \tilde{p}_j$ for $j = 1, 2$ and $i = 1, \ldots, n_j$, where $\tilde{p}_1$ and $\tilde{p}_2$ are n-FuRBIs with underlying joint Lévy intensity provided by (3.24) and denote with ρ_0 the correlation between two random variables jointly sampled from G_0, then*

$$Corr(X_{i,1}, X_{i',2}) = g(\theta)\,\rho_0$$

where $g : \mathbb{R}^+ \to (0, (1-\sigma))$.

Proof. The theorem follows by Proposition 3.1. □

Therefore, for appropriate choices of G_0, and in particular of ρ_0, the correlation between observations in different samples can be negative.

3.5.1 Prior algorithm and simulations

We provide here a simulation study, which outlines the flexibility of the nonparamteric prior introduced in the previous section when $\alpha(\mathrm{d}x_1, \mathrm{d}x_2)$ is a multivariate Gaussian probability measure with zero means, unitary variances and correlation ρ_0. To do so we need an algorithm to sample the infinite dimensional parameters $\tilde{p}_1$ and $\tilde{p}_2$ for different values of the hyperparameters θ and ρ_0. Algorithm 1 do so and it has been obtained adapting the Algorithm 6.15 in Cont & Tankov (2004) to the atom-dependent structure.
We first sample a realization for $\tilde{p}_1$ and then simulate the conditional distribution of $\tilde{p}_2$, given $\tilde{p}_1$, under different hyperparameters choices. Figure 3.2 shows the results in terms of cumulative distributions functions. The plots in the first and second row ($\rho_0 = -1$ and $\rho_0 = -0.5$) show a strong and medium negative correlation between the observables, represented by the opposite behaviour of $\tilde{p}_2$ and $\tilde{p}_1$. While $\tilde{p}_1$ associate high probabilities to positive values, $\tilde{p}_2$ tends to associates high probabilities to negative values. While ρ_0 increases, first the conditional distribution of $\tilde{p}_2$ becomes independent from $\tilde{p}_1$ ($\rho_0 = 0$) and then shows a behaviour similar to that of $\tilde{p}_1$ ($\rho_0 = 0.5$ and $\rho_0 = 1$), corresponding to positive correlation of the observables.

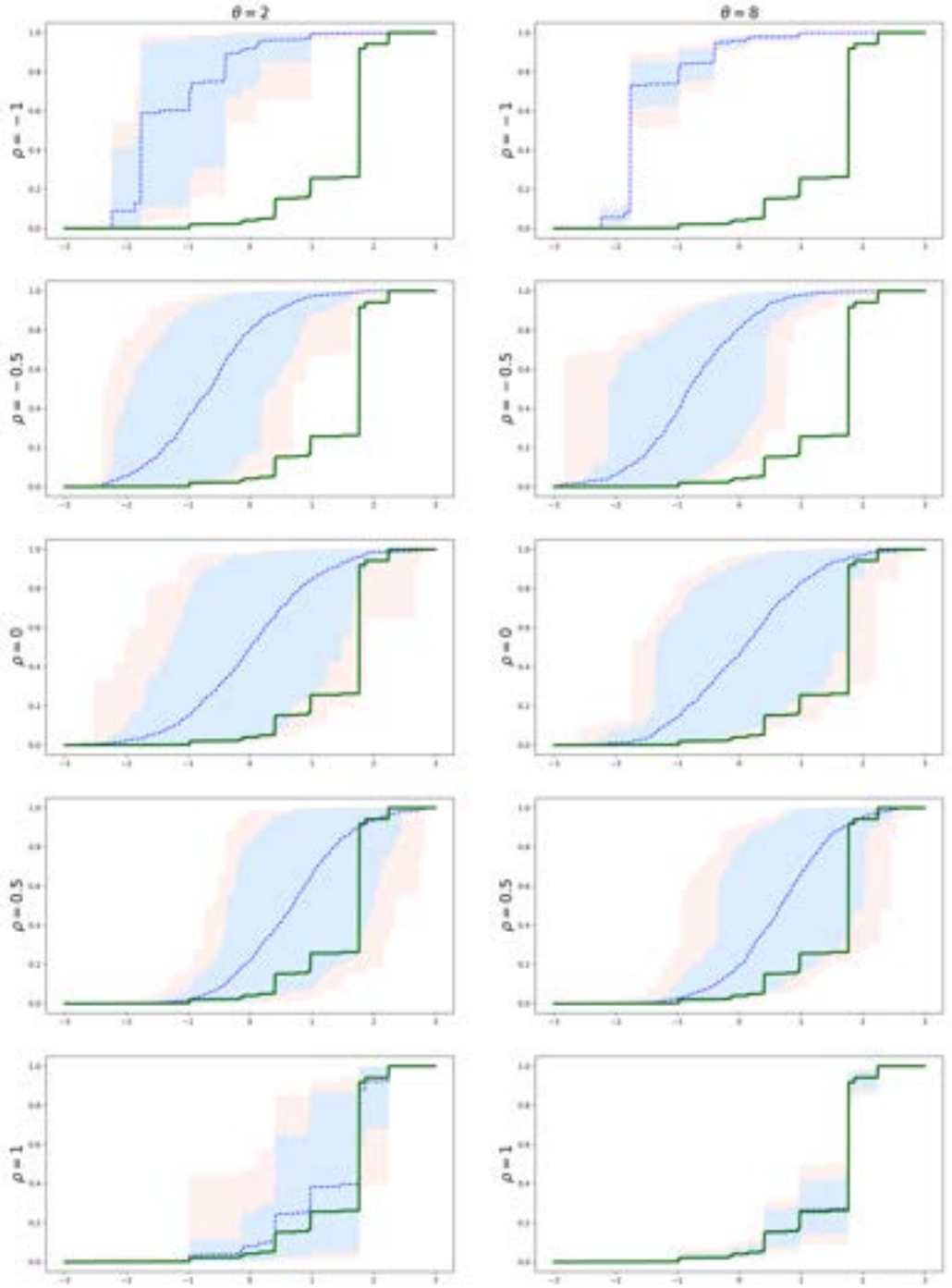


Figure 3.2: Dark green continous line: a realization of the c.d.f corresponding to $\tilde{p}_1$, i.e. $\int_{-\infty}^{x} \tilde{p}_1(\mathrm{d}x)$. Blue dashed lines: conditional expected value of the c.d.f. corresponding to $\tilde{p}_2$, given the realization of $\tilde{p}_1$, i.e. $\mathbb{E}\left[\int_{-\infty}^{x} \tilde{p}_2(\mathrm{d}x) \mid \tilde{p}_1\right]$. Light blue shaded area: 95% probability interval for the c.d.f. corresponding to $\tilde{p}_2$. Pink shaded area: 99% probability interval for the c.d.f. corresponding to $\tilde{p}_2$.

Algorithm 1: Prior Sampler

for $k \leftarrow 0$ **to** K **do**
  Sample T_k from an Exponential(1);
  Compute $S_k^{(1)} = S_{k-1}^{(1)} + T_k$;
  Sample U_k from an U(0, 1);
  Compute $S_k^{(2)} = S_k^{(1)} \left(U_k^{-\theta/(1+\theta)} - 1 \right)^{-\frac{1}{\theta}}$;
  Compute $W_{j,k} = (S_k^{(j)} \sigma \, \Gamma(1-\sigma))^{-\frac{1}{\sigma}}$ for $j = 1, 2$;
  Sample $(\theta_{1,k}, \theta_{2,k})$ from G_0;
end
Compute $\bar{W}_{j,k} = W_{j,k} / \sum_{k=1}^{K} W_{j,k}$ for $j = 1, 2$ and $k = 1, \ldots, K$;
Obtain $\tilde{p}_1 \approx \sum_{k=1}^{K} \bar{W}_{1,k} \delta_{\theta_{1,k}}$ and $\tilde{p}_2 \approx \sum_{k=1}^{K} \bar{W}_{2,k} \delta_{\theta_{2,k}}$

3.6 Posterior characterization

In the previous section we illustrated interesting a priori properties of n-FuRBI. However, the crux of Bayesian models is writing in a manageable way the posterior distribution, i.e., the distribution of $\tilde{p}_1$ and $\tilde{p}_2$ given data from model in (3.8).
Conjugacy is out of question here: even in the exchangeable context it is a property that is shared only by the Dirichlet process (see James et al., 2006). Nevertheless, we will show that, conditional on a set of suitable latent variables, the posterior distribution of the original CRV (μ_1, μ_2) is again a CRV. Through normalization, all the other quantities of interest become available, leading to a simple formulation and viable methods for sampling. It is to stress once again that those results come from the high analytical tractability of CRMs and CRVs.
Thus, consider a sample of n observations $\boldsymbol{X}_{1:n} = (X_1, \ldots, X_n)$ from $\tilde{p}_1$ with unique values $\boldsymbol{X}^* = (X_1^*, \ldots, X_k^*)$ and associated multiplicities $\mathbf{n} = (n_1, \ldots, n_k)$; analogously, consider m observations $\boldsymbol{Y}_{1:m} = (Y_1, \ldots, Y_m)$ from $\tilde{p}_2$ with unique values $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_c^*)$ with multiplicities $\mathbf{m} = (m_1, \ldots, m_c)$. Notice that it is immediate to check for ties: indeed, it suffices to identify equal observations in each sample. Instead, in general hyper-ties cannot be deduced directly from the data: the hidden structure they induce will be the main component of the mentioned latent structure. As already clarified and highlighted in representation (3.4), a hyper-tie is an actual tie in the product space: it means that X_i^* and Y_j^* form a hyper-tie if they correspond to the two coordinates of the same atom. In the following we will denote by (i, j), with $1 \leq i \leq k$ and $1 \leq j \leq c$ a hyper-tie between X_i^* and Y_j^*. Not all unique values will form an hyper-tie, thus for consistency of notation, $(i, c+1)$, with $1 \leq i \leq k$, will denote that X_i^* does not form an hyper-tie with any value in $\boldsymbol{Y}^*$ and $(k+1, j)$, with $1 \leq j \leq c$, will denote that Y_j^* does not form an hyper-tie with any

value in $\boldsymbol{X}^*$. We are now able to define the *compatible latent structures*, i.e., all the collection of hyper-ties across unique values that are consistent with samples $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$.

Definition 3.3. *We say that $\boldsymbol{p} = \{(i_l, j_l)\}_l$ is a compatible latent structure (CLS) for $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$ if*

1. *Each element of $\boldsymbol{X}^*$ forms at most one hyper-tie: for any $1 \leq i \leq k$ there exists exactly one i_l such that $i_l = i$.*

2. *Each element of $\boldsymbol{Y}^*$ forms at most one hyper-tie: for any $1 \leq j \leq c$ there exists exactly one j_l such that $j_l = j$.*

3. *At least one coordinate refers to an element of $\boldsymbol{X}^*$ or $\boldsymbol{Y}^*$: for any l, if $i_l = k+1$ then $j_l \neq c+1$.*

Finally we call $\mathscr{P} = \{\boldsymbol{p} \mid \boldsymbol{p} is a\ CLS\}$ the set of all compatible latent structures.

Once the latent structure $\mathbf{p}$ is fixed, we can collect all the hyper-ties in

$$\Delta_{\mathbf{p}} = \{(i,j) \in \mathbf{p} \mid i \neq k+1 \text{ and } j \neq c+1\},$$

and all the remaining values in

$$\Delta^1_{\mathbf{p}} = \{(i,j) \in \mathbf{p} \mid j = c+1\}, \quad \Delta^2_{\mathbf{p}} = \{(i,j) \in \mathbf{p} \mid i = k+1\}.$$

If X^*_i and Y^*_j form a hyper-tie, it means that (X^*_i, Y^*_j) is an actual atom in representation (3.4). Instead, if X^*_i does not form a hyper-tie, we have a partial knowledge of the original pair: the unknown second coordinate can be sampled from $P_{X^*_i}(\cdot)$, that is the conditional distribution given X^*_i, induced by the joint measure G_0. Analogously happens if Y^*_j does not form a hyper-tie.
We consider the following simplifying notation

$$g_{i,j} = g_0(X^*_i, Y^*_j), \quad g_{i,c+1} = p_0(X^*_i), \quad g_{k+1,j} = p_0(Y^*_j),$$

where g_0 and p_0 are the density functions of G_0 and P_0 respectively, that we assume to exist with respect to suitable dominating measures. Finally, we consider the following integrals

$$\tau_{n,m}(u) = \int_{\mathbb{R}^2_+} e^{-u_1 s_1 - u_2 s_2} s_1^n s_2^m \, \rho(\mathrm{d}s_1, \mathrm{d}s_2), \quad u = (u_1, u_2),$$

where often n and m will be equal to n_i and m_j, with $1 \leq i \leq k+1$, $1 \leq j \leq c+1$ and $n_{k+1} = m_{c+1} = 0$ for consistency.
Thus, the set of latent variables with the associated distributions, conditional on $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$, is given by

- $\mathbf{p} \in \mathscr{P}$, whose probability mass function is proportional to

$$\left(\prod_{(i,j)\in\mathbf{p}} g_{i,j}\right) \int_{\mathbb{R}_+^2} u_1^{n-1} u_2^{m-1} \prod_{(i,j)\in\mathbf{p}} \tau_{n_i,m_j}(u) e^{-\psi_b(u)}\,\mathrm{d}u.$$

- (U_1, U_2), whose density on $\mathbb{R}_+^2$ is proportional to

$$u_1^{n-1} u_2^{m-1} \prod_{(i,j)\in\mathbf{p}} \tau_{n_i,m_j}(u) e^{-\psi_b(u)}$$

- $\{Z_i^x\}_i$, whose distribution is given by $P_{X_i^*}(\cdot)$, for any $i = 1, \dots k$.

- $\{Z_j^y\}_j$, whose distribution is given by $P_{Y_j^*}(\cdot)$, for any $j = 1, \dots, c$.

We are now ready to state the main theorem on posterior characterization.

Theorem 3.2. *Consider samples $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$ from model (3.8), with Q being the law of a n-FuRBI. Then, the distribution of (μ_1, μ_2) conditional on $\boldsymbol{X}_{1:n}$, $\boldsymbol{Y}_{1:m}$ and the set of latent variables $\boldsymbol{p}$, U_1, U_2, $\{Z_i^x\}_i$, $\{Z_j^y\}_j$ is given by*

$$(\hat{\mu}_1, \hat{\mu}_2) + \sum_{(i,j)\in\Delta_p} J_{i,j}\delta_{(X_i^*, Y_j^*)} + \sum_{(i,j)\in\Delta_p^1} J_{i,c+1}\delta_{(X_i^*, Z_i^x)} + \sum_{(i,j)\in\Delta_p^2} J_{k+1,j}\delta_{(Z_j^y, Y_j^*)},$$

where

- *$(\hat{\mu}_1, \hat{\mu}_2)$ is a CRV with intensity $e^{-U_1 s_1 - U_2 s_2}\rho(\mathrm{d}s_1, \mathrm{d}s_2)G_0(\mathrm{d}x)$.*

- *$J_{i,j} = (J_{i,j}^1, J_{i,j}^2)$ are jumps with density proportional to $s_1^{n_i} s_2^{m_j} e^{-U_1 s_1 - U_2 s_2}\rho(\mathrm{d}s_1, \mathrm{d}s_2)$.*

- *$(\hat{\mu}_1, \hat{\mu}_2)$ and $J_{i,j}$ are independent.*

Proof. We need to compute the conditional Laplace functional of (μ_1, μ_2), i.e.

$$\mathbb{E}\left[e^{-\int_{\mathbb{X}^2} h_1(x)\,\mu_1(\mathrm{d}x) - \int_{\mathbb{X}^2} h_2(x)\,\mu_2(\mathrm{d}x)} \mid \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}\right],$$

with $h_i : \mathbb{X}^2 \to \mathbb{R}$ measurable functions. Define $A_j = A_{j,\epsilon} = \{x \in \mathbb{X} \mid d(x, X_i^*) < \epsilon\}$ and $B_j = B_{j,\epsilon} = \left\{x \in \mathbb{X} \mid d(x, Y_j^*) < \epsilon\right\}$, with $1 \le i \le k$ and $1 \le j \le c$, such that $A_i \cap A_j = \emptyset$ and $B_i \cap B_j = \emptyset$ for any $i \neq j$. Moreover, denote

$$A_{k+1} = \left(\cup_{i=1}^k A_i\right)^c, \quad B_{c+1} = \left(\cup_{i=1}^c B_i\right)^c.$$

Thus our goal becomes to compute

$$\begin{aligned}
&\mathbb{E}\big[e^{-\int_{\mathbb{X}^2} h_1(x)\,\mu_1(\mathrm{d}x)-\int_{\mathbb{X}^2} h_2(x)\,\mu_2(\mathrm{d}x)} \mid \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}\big] \\
&= \lim_{\epsilon\to 0} \mathbb{E}\left[e^{-\int_{\mathbb{X}^2} h_1(x)\,\mu_1(\mathrm{d}x)-\int_{\mathbb{X}^2} h_2(x)\,\mu_2(\mathrm{d}x)} \mid \boldsymbol{X}^* \in \times_{j=1}^k A_j, \boldsymbol{Y}^* \in \times_{j=1}^c B_j\right] \\
&= \lim_{\epsilon\to 0} \frac{\mathbb{E}\left[e^{-\int_{\mathbb{X}^2} h_1(x)\,\mu_1(\mathrm{d}x)-\int_{\mathbb{X}^2} h_2(x)\,\mu_2(\mathrm{d}x)} \prod_{j=1}^k \tilde{p}_1(A_j)^{n_j} \prod_{j=1}^c \tilde{p}_2(B_j)^{m_j}\right]}{\mathbb{E}\left[\prod_{j=1}^k \tilde{p}_1(A_j)^{n_j} \prod_{j=1}^c \tilde{p}_2(B_j)^{m_j}\right]}.
\end{aligned} \tag{3.25}$$

We start to evaluate

$$\begin{aligned}
&\mathbb{E}\big[\tilde{p}_1(A_1)^{n_1}\ldots\tilde{p}_1(A_k)^{n_k}\tilde{p}_2(B_1)^{m_1}\tilde{p}_2(B_c)^{m_c}\big] = \\
&\qquad \mathbb{E}\left[\frac{\tilde{\mu}_1(A_1)^{n_1}\ldots\tilde{\mu}_1(A_k)^{n_k}\tilde{\mu}_2(B_1)^{m_1}\tilde{\mu}_2(B_c)^{m_c}}{\tilde{\mu}_1(\mathbb{X})^n\tilde{\mu}_2(\mathbb{X})^m}\right] = \\
&\qquad = \mathbb{E}\left[\frac{\mu_1(A_1\times\mathbb{X})^{n_1}\ldots\mu_1(A_k\times\mathbb{X})^{n_k}\mu_2(\mathbb{X}\times B_1)^{m_1}\mu_2(\mathbb{X}\times B_c)^{m_c}}{\mu_1(\mathbb{X}\times\mathbb{X})^n\mu_2(\mathbb{X}\times\mathbb{X})^m}\right] = \mathcal{G}.
\end{aligned}$$

By Netwon's binomial

$$\mu_1(A_h\times\mathbb{X}) = \sum_{i_1^h+\ldots i_{c+1}^h=n_h} \binom{n_h}{i_1^h,\ldots,i_{c+1}^h} \prod_{r=1}^{c+1} \mu_1^{i_r^h}(A_h\times B_r), \quad h=1,\ldots,k,$$

$$\mu_2(\mathbb{X}\times B_r) = \sum_{j_1^r+\ldots j_{k+1}^r=m_r} \binom{m_r}{j_1^r,\ldots,j_{k+1}^r} \prod_{h=1}^{k+1} \mu_2^{j_h^r}(A_h\times B_r), \quad r=1,\ldots,c.$$

For ease of notation denote

$$\begin{aligned}
\sum_{\mathbf{i},\mathbf{j}} \binom{\mathbf{n}}{\mathbf{i}}\binom{\mathbf{m}}{\mathbf{j}} = &\sum_{i_1^1+\ldots i_{c+1}^1=n_1} \binom{n_1}{i_1^1,\ldots,i_{c+1}^1} \cdots \sum_{i_1^{c+1}+\ldots i_{c+1}^{k+1}=n_{k+1}} \binom{n_{k+1}}{i_1^{k+1},\ldots,i_{c+1}^{k+1}} \\
&\sum_{j_1^1+\ldots j_{k+1}^1=m_1} \binom{m_1}{j_1^1,\ldots,j_{k+1}^1} \cdots \sum_{j_1^{k+1}+\ldots j_{k+1}^{k+1}=m_{k+1}} \binom{m_{k+1}}{j_1^{k+1},\ldots,j_{k+1}^{k+1}}.
\end{aligned}$$

Thus

$$\mathcal{G} = \sum_{\mathbf{i},\mathbf{j}} \binom{\mathbf{n}}{\mathbf{i}}\binom{\mathbf{m}}{\mathbf{j}} \mathcal{G}_{\mathbf{i},\mathbf{j}},$$

with

$$\mathcal{G}_{\mathbf{i},\mathbf{j}} = \mathbb{E}\left[\frac{\left(\prod_{h=1}^k\prod_{r=1}^c \mu_1^{i_r^h}(A_h\times B_r)\mu_2^{j_h^r}(A_h\times B_r)\right)\prod_{h=1}^k \mu_1^{i_{c+1}^h}(A_h\times B_{c+1})\prod_{r=1}^c \mu_2^{j_{k+1}^r}(A_{k+1}\times B_r)}{\mu_1(\mathbb{X}\times\mathbb{X})^n\mu_2(\mathbb{X}\times\mathbb{X})^m}\right].$$

Moreover, we have

$$\frac{1}{\mu_1(\mathbb{X}\times\mathbb{X})^n\mu_2(\mathbb{X}\times\mathbb{X})^m} = \frac{1}{\Gamma(n)\Gamma(m)}\int_{\mathbb{R}_+^2} u_1^{n-1}u_2^{m-1}e^{-u_1\mu_1-u_2\mu_2}\,\mathrm{d}u,$$

with $u=(u_1,u_2)$.

By Fubini's Theorem

$$\begin{aligned}\mathcal{G}_{\mathbf{i},\mathbf{j}} = \int_{\mathbb{R}_+^2}\frac{u_1^{n-1}u_2^{m-1}}{\Gamma(n)\Gamma(m)}\mathbb{E}\Bigg[e^{-u_1\mu_1-u_2\mu_2}\left(\prod_{h=1}^{k}\prod_{r=1}^{c}\mu_1^{i_r^h}(A_h\times B_r)\mu_2^{j_h^r}(A_h\times B_r)\right)\\ \prod_{h=1}^{k}\mu_1^{i_{c+1}^h}(A_h\times B_{c+1})\prod_{r=1}^{c}\mu_2^{j_{k+1}^r}(A_{k+1}\times B_r)\Bigg]\,\mathrm{d}u =\\ = \int_{\mathbb{R}_+^2}\frac{u_1^{n-1}u_2^{m-1}}{\Gamma(n)\Gamma(m)}\rho_{\mathbf{i},\mathbf{j}}(u)\,\mathrm{d}u.\end{aligned}$$

By independence of the increments we have

$$\begin{aligned}\rho_{\mathbf{i},\mathbf{j}}(u) = \mathbb{E}\Bigg[\left(\prod_{h=1}^{k}\prod_{r=1}^{c}e^{-u_1\mu_1(A_h\times B_r)-u_2\mu_2(A_h\times B_r)}\mu_1^{i_r^h}(A_h\times B_r)\mu_2^{j_h^r}(A_h\times B_r)\right)\\ \prod_{h=1}^{k}e^{-u_1\mu_1(A_h\times B_{c+1})-u_2\mu_2(A_h\times B_{c+1})}\mu_1^{i_{c+1}^h}(A_h\times B_{c+1})\\ \prod_{r=1}^{c}e^{-u_1\mu_1(A_{k+1}\times B_r)-u_2\mu_2(A_{k+1}\times B_r)}\mu_2^{j_{k+1}^r}(A_{k+1}\times B_r)\Bigg]\end{aligned}$$

Thus

$$\begin{aligned}\rho_{\mathbf{i},\mathbf{j}}(u) = \prod_{h=1}^{k}\prod_{r=1}^{c}\mathbb{E}\left[e^{-u_1\mu_1(A_h\times B_r)-u_2\mu_2(A_h\times B_r)}\mu_1^{i_r^h}(A_h\times B_r)\mu_2^{j_h^r}(A_h\times B_r)\right]\\ \prod_{h=1}^{k}\mathbb{E}\left[e^{-u_1\mu_1(A_h\times B_{c+1})-u_2\mu_2(A_h\times B_{c+1})}\mu_1^{i_{c+1}^h}(A_h\times B_{c+1})\right]\\ \prod_{r=1}^{c}\mathbb{E}\left[e^{-u_1\mu_1(A_{k+1}\times B_r)-u_2\mu_2(A_{k+1}\times B_r)}\mu_2^{j_{k+1}^r}(A_{k+1}\times B_r)\right].\end{aligned}$$

Considering each element separately we have

$$\begin{aligned}
&\mathbb{E}\left[e^{-u_1\mu_1(A_h\times B_r)-u_2\mu_2(A_h\times B_r)}\mu_1^i(A_h\times B_r)\mu_2^j(A_h\times B_r)\right] \\
&\qquad = \mathbb{E}\left[(-1)^{i+j}\frac{\partial^{i+j}}{\partial u_1^i\partial u_2^j}e^{-u_1\mu_1(A_h\times B_r)-u_2\mu_2(A_h\times B_r)}\right] \\
&\qquad = (-1)^{i+j}\frac{\partial^{i+j}}{\partial u_1^i\partial u_2^j}\mathbb{E}\left[e^{-u_1\mu_1(A_h\times B_r)-u_2\mu_2(A_h\times B_r)}\right] \\
&\qquad = (-1)^{i+j}\frac{\partial^{i+j}}{\partial u_1^i\partial u_2^j}\left\{e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(ds)G_0(x)}\right\}.
\end{aligned}$$

Recall that we are interested in the limit as $\epsilon\to 0$, so that

$$\begin{aligned}
&\frac{\partial^{i+j}}{\partial u_1^i\partial u_2^j}\left\{e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}\right\} \\
&\qquad \sim e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)} \\
&\qquad \frac{\partial^{i+j}}{\partial u_1^i\partial u_2^j}\left\{\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\},
\end{aligned} \tag{3.26}$$

where we say $f\sim g$ if $\lim_{\epsilon\to 0} f(x)/g(x)=1$.

It comes from simple computations

$$\begin{aligned}
&\frac{\partial^{i+j}}{\partial u_1^i\partial u_2^j}\left\{e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}\right\} = \\
&\frac{\partial^{i+j-1}}{\partial u_1^{i-1}\partial u_2^j}\left\{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}e^{-u_1s_1-u_2s_2}s_1\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}\right\} = \\
&\frac{\partial^{i+j-2}}{\partial u_1^{i-2}\partial u_2^j}\left\{\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}e^{-u_1s_1-u_2s_2}s_1^2\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}+\right. \\
&\left.\left(\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}e^{-u_1s_1-u_2s_2}s_1\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right)^2 e^{-\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}\right\},
\end{aligned}$$

and

$$\lim_{\epsilon\to 0}\frac{\left(\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}e^{-u_1s_1-u_2s_2}s_1\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right)^2}{\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}e^{-u_1s_1-u_2s_2}s_1^2\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}=0.$$

Repeating the argument we get (3.26).

Thus, denoting $\rho(u) = \sum_{\mathbf{i},\mathbf{j}} \binom{\mathbf{n}}{\mathbf{i}}\binom{\mathbf{m}}{\mathbf{j}}\rho_{\mathbf{i},\mathbf{j}}(u)$, aggregating the terms

$$\begin{aligned}\rho(u) \sim \sum_{\mathbf{i},\mathbf{j}} \binom{\mathbf{n}}{\mathbf{i}}\binom{\mathbf{m}}{\mathbf{j}}(-1)^{n+m}e^{-\psi_b(u)} \prod_{h=1}^{k}\prod_{r=1}^{c}\left\{\frac{\partial^{i_r^h+j_h^r}}{\partial u_1^{i_r^h}\partial u_2^{j_h^r}}\int_{A_h\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\} \\ \prod_{h=1}^{k}\left\{\frac{\partial^{i_{c+1}^h}}{\partial u_1^{i_{c+1}^h}}\int_{A_h\times B_{c+1}}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\} \\ \prod_{r=1}^{c}\left\{\frac{\partial^{j_{k+1}^r}}{\partial u_2^{i_{k+1}^r}}\int_{A_{k+1}\times B_r}\int_{\mathbb{R}_+^2}(1-e^{-u_1s_1-u_2s_2})\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\} \\ = \sum_{\mathbf{i},\mathbf{j}}\binom{\mathbf{n}}{\mathbf{i}}\binom{\mathbf{m}}{\mathbf{j}}(-1)^{n+m}V(\mathbf{i},\mathbf{j}).\end{aligned}$$

The following three Lemmas characterizes the set of indices $(\mathbf{i},\mathbf{j})$.

Lemma 3.8. *Consider $(\boldsymbol{i},\boldsymbol{j})$ such that $0 < i_r^h, i_l^h < n_h$, with $r > l$ and $1 \le h \le k$. Then $\exists(\tilde{\boldsymbol{i}},\tilde{\boldsymbol{j}})$ such that $\lim_{\epsilon\to 0} V(\boldsymbol{i},\boldsymbol{j})/V(\tilde{\boldsymbol{i}},\tilde{\boldsymbol{j}}) \to 0$.*

Proof. For ease of notation denote $i^h = (i_1^h,\dots,i_{c+1}^h)$. Then

- If $r = c+1$, set $\tilde{i}^h = (i_1^h,\dots,i_l^h + i_{c+1}^h,\dots,0)$.
- If $j_h^r = 0$, set $\tilde{i}^h = (i_1^h,\dots,i_l^h + i_r^h,\dots,0,\dots)$.
- If $j_h^l = 0$, set $\tilde{i}^h = (i_1^h,\dots,0,\dots,i_r^h + i_l^h,\dots)$.
- If $j_h^l > 0$ and $j_h^r > 0$, set $\tilde{j}^r = (j_1^r,\dots,0,\dots,j_{k+1}^r + j_h^r)$ and $\tilde{i}^h = (i_1^h,\dots,i_l^h + i_r^h,\dots,0,\dots)$.

Indeed, considering the last case

$$\lim_{\epsilon\to 0}\frac{V(\mathbf{i},\mathbf{j})}{V(\tilde{\mathbf{i}},\tilde{\mathbf{j}})} = \lim_{\epsilon\to 0}\frac{\int_{A_h\times B_r}\int_{\mathbb{R}_+^2} e^{-u_1s_1-u_2s_2}s_1^{i_r^h}s_2^{j_h^r}\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)}{\int_{A_{c+1}\times B_r}\int_{\mathbb{R}_+^2} e^{-u_1s_1-u_2s_2}s_2^{j_h^r+j_{c+1}^r}\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)} = 0.$$

□

Lemma 3.8 guarantees that i^h has exactly one element different from 0, that is equal to n_h.

Lemma 3.9. *Consider $(\boldsymbol{i},\boldsymbol{j})$ such that $i_r^h = n_h$ and $j_h^r = 0$. Then $\exists(\tilde{\boldsymbol{i}},\tilde{\boldsymbol{j}})$ such that $\lim_{\epsilon\to 0} V(\boldsymbol{i},\boldsymbol{j})/V(\tilde{\boldsymbol{i}},\tilde{\boldsymbol{j}}) \to 0$.*

Proof. Set $(\tilde{\mathbf{i}},\tilde{\mathbf{j}})$ equal to $(\mathbf{i},\mathbf{j})$, apart from $\tilde{i}_r^h = 0$ and $\tilde{i}_{c+1}^h = n_h$. □

Lemma 3.10. *Consider $(\boldsymbol{i},\boldsymbol{j})$ such that $i_{c+1}^h = n_h$ and $j_h^r > 0$. Then $\exists(\tilde{\boldsymbol{i}},\tilde{\boldsymbol{j}})$ such that $\lim_{\epsilon\to 0} V(\boldsymbol{i},\boldsymbol{j})/V(\tilde{\boldsymbol{i}},\tilde{\boldsymbol{j}}) \to 0$.*

Proof. Set $(\tilde{\mathbf{i}},\tilde{\mathbf{j}})$ equal to $(\mathbf{i},\mathbf{j})$, apart from $\tilde{j}_h^r = 0$ and $\tilde{j}_{k+1}^r = m_r$. □

The three lemmas tell that each relevant $(\mathbf{i}, \mathbf{j})$ corresponds to a CLS, i.e.

$$\rho(u) \sim \sum_{\mathbf{p}\in\mathscr{P}} (-1)^{n+m} e^{-\psi_b(u)} \prod_{(i,j)\in\Delta_{\mathbf{p}}} \left\{ \frac{\partial^{n_i+m_j}}{\partial u_1^{n_i} \partial u_2^{m_j}} \int_{A_i\times B_j} \int_{\mathbb{R}_+^2} (1 - e^{-u_1 s_1 - u_2 s_2})\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^1} \left\{ \frac{\partial^{n_i}}{\partial u_1^{n_i}} \int_{A_i\times B_{c+1}} \int_{\mathbb{R}_+^2} (1 - e^{-u_1 s_1 - u_2 s_2})\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^2} \left\{ \frac{\partial^{m_j}}{\partial u_2^{m_j}} \int_{A_{k+1}\times B_j} \int_{\mathbb{R}_+^2} (1 - e^{-u_1 s_1 - u_2 s_2})\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}.$$

Evaluating the derivatives we have

$$\rho(u) \sim \sum_{\mathbf{p}\in\mathscr{P}} e^{-\psi_b(u)} \prod_{(i,j)\in\Delta_{\mathbf{p}}} \left\{ \int_{A_i\times B_j} \int_{\mathbb{R}_+^2} e^{-u_1 s_1 - u_2 s_2} s_1^{n_i} s_2^{m_j}\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^1} \left\{ \int_{A_i\times B_{c+1}} \int_{\mathbb{R}_+^2} e^{-u_1 s_1 - u_2 s_2} s_1^{n_i}\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^2} \left\{ \int_{A_{k+1}\times B_j} \int_{\mathbb{R}_+^2} e^{-u_1 s_1 - u_2 s_2} s_2^{m_j}\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}.$$

Finally, we get

$$\mathcal{J} \sim \sum_{\mathbf{p}\in\mathscr{P}} \int_{\mathbb{R}_+^2} \frac{u_1^{n-1} u_2^{m-1}}{\Gamma(n)\Gamma(m)} e^{-\psi_b(u)} \prod_{(i,j)\in\Delta_{\mathbf{p}}} \left\{ \int_{A_i\times B_j} \int_{\mathbb{R}_+^2} e^{-u_1 s_1 - u_2 s_2} s_1^{n_i} s_2^{m_j}\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^1} \left\{ \int_{A_i\times B_{c+1}} \int_{\mathbb{R}_+^2} e^{-u_1 s_1 - u_2 s_2} s_1^{n_i}\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^2} \left\{ \int_{A_{k+1}\times B_j} \int_{\mathbb{R}_+^2} e^{-u_1 s_1 - u_2 s_2} s_2^{m_j}\, \rho(\mathrm{d}s) G_0(\mathrm{d}x) \right\} \mathrm{d}u.$$

Evaluating the numerator of (3.25) with the same reasoning yields a formula asymptotic to

$$\sum_{\mathbf{p}\in\mathscr{P}}\int_{\mathbb{R}_+^2}\frac{u_1^{n-1}u_2^{m-1}}{\Gamma(n)\Gamma(m)}e^{-\psi_h(u)}\prod_{(i,j)\in\Delta_{\mathbf{p}}}\left\{\int_{A_i\times B_j}\int_{\mathbb{R}_+^2}e^{-(h_1(x)+u_1)s_1-(h_2(x)+u_2)s_2}s_1^{n_i}s_2^{m_j}\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^1}\left\{\int_{A_i\times B_{c+1}}\int_{\mathbb{R}_+^2}e^{-(h_1(x)+u_1)s_1-(h_2(x)+u_2)s_2}s_1^{n_i}\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\}$$
$$\prod_{(i,j)\in\Delta_{\mathbf{p}}^2}\left\{\int_{A_{k+1}\times B_j}\int_{\mathbb{R}_+^2}e^{-(h_1(x)+u_1)s_1-(h_2(x)+u_2)s_2}s_2^{m_j}\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)\right\}\mathrm{d}u.$$

where $\psi_h(u)=\int_{\mathbb{X}^2}\int_{\mathbb{R}_+^2}\left(1-e^{-(h_1(x)+u_1)s_1-(h_2(x)+u_2)s_2}\right)\,\rho(\mathrm{d}s)G_0(\mathrm{d}x)$.

Notice that $1-e^{-(h_1(x)+u_1)s_1-(h_2(x)+u_2)s_2}=\left[1-e^{-u_1s_1-u_2s_2}\right]+\left[1-e^{-h_1(x)s_1-h_2(x)s_2}\right]$ so that

$$\begin{aligned}e^{-\psi_h(u)}&=e^{-\psi_b(u)}e^{-\int_{\mathbb{X}^2}\int_{\mathbb{R}_+^2}\left[1-e^{-h_1(x)s_1-h_2(x)s_2}\right]\rho(\mathrm{d}s)G_0(\mathrm{d}x)}\\&=e^{-\psi_b(u)}\mathbb{E}\left[e^{-\int_{\mathbb{X}^2}h_1(x)\,\hat{\mu}_1(\mathrm{d}x)-\int_{\mathbb{X}^2}h_2(x)\,\hat{\mu}_2(\mathrm{d}x)}\right].\end{aligned}$$

Notice that $G_0(A_h\times B_r)=\epsilon\frac{G_0(A_h\times B_r)}{\epsilon}\sim\epsilon g_{h,r}$, for $1\le i\le c, 1\le j\le k$, $G_0(A_h\times\mathrm{d}x)\sim\epsilon g_{h,c+1}Q_{X_h^*}(\mathrm{d}x)$ and $G_0(\mathrm{d}x\times B_r)\sim\epsilon g_{k+1,r}P_{Y_r^*}(\mathrm{d}x)$.

Thus, evaluating the limit in (3.25) we get

$$\begin{aligned}\mathbb{E}\left[e^{-\int_{\mathbb{X}^2}h_1(x)\,\mu_1(\mathrm{d}x)-\int_{\mathbb{X}^2}h_2(x)\,\mu_2(\mathrm{d}x)}\mid\boldsymbol{X}_{1:n},\boldsymbol{Y}_{1:m}\right]&=\sum_{\mathbf{p}\in\mathscr{P}}\int_{\mathbb{R}_+^2}\mathbb{E}\left[e^{-\int_{\mathbb{X}^2}h_1(x)\,\hat{\mu}_1(\mathrm{d}x)-\int_{\mathbb{X}^2}h_2(x)\,\hat{\mu}_2(\mathrm{d}x)}\right]\\&\times\prod_{(i,j)\in\Delta_{\mathbf{p}}}\int_{\mathbb{R}_+^2}e^{-h_1(X_i^*,Y_j^*)s_1-h_2(X_i^*,Y_j^*)s_2}\frac{s_1^{n_i}s_2^{m_j}e^{-u_1s_1-u_2s_2}\rho(\mathrm{d}s)}{\tau_{n_i,m_j}(u)}\\&\times\prod_{(i,j)\in\Delta_{\mathbf{p}}^1}\int_{\mathbb{X}}\int_{\mathbb{R}_+^2}e^{-h_1(X_i^*,x_2)s_1-h_2(X_i^*,x_2)s_2}\frac{s_1^{n_i}e^{-u_1s_1-u_2s_2}\rho(\mathrm{d}s)}{\tau_{n_i,0}(u)}Q_{X_i^*}(\mathrm{d}x_2)\\&\times\prod_{(i,j)\in\Delta_{\mathbf{p}}^2}\int_{\mathbb{X}}\int_{\mathbb{R}_+^2}e^{-h_1(x_1,Y_2^*)s_1-h_2(x_1,Y_2^*)s_2}\frac{s_2^{m_j}e^{-u_1s_1-u_2s_2}\rho(\mathrm{d}s)}{\tau_{0,m_j}(u)}P_{Y_j^*}(\mathrm{d}x_1)\\&\times\left(\frac{\int_{\mathbb{R}_+^2}u_1^{n-1}u_2^{m-1}\prod_{(i,j)\in\mathbf{p}}g_{i,j}\tau_{n_i,m_j}(u)e^{-\psi_b(u)}\,\mathrm{d}u}{\sum_{\mathbf{q}\in\mathscr{P}}\int_{\mathbb{R}_+^2}u_1^{n-1}u_2^{m-1}\prod_{(i,j)\in\mathbf{q}}g_{i,j}\tau_{n_i,m_j}(u)e^{-\psi_b(u)}\mathrm{d}u}\right)\\&\frac{u_1^{n-1}u_2^{m-1}\prod_{(i,j)\in\mathbf{p}}\tau_{n_i,m_j}(u)e^{-\psi_b(u)}\,\mathrm{d}u}{\int_{\mathbb{R}_+^2}u_1^{n-1}u_2^{m-1}\prod_{(i,j)\in\mathbf{p}}\tau_{n_i,m_j}(u)e^{-\psi_b(u)}\,\mathrm{d}u},\end{aligned}$$

as desired. □

Conditional on the latent variables, the structure is very natural: the posterior is given by a CRV with modified intensity and fixed locations, given by the pairs formed by the hyper-ties. Notice that the result is reminiscent of the analogous for the exchangeable case (James et al., 2009); however, a significant novelty is given by the role played by hyper-ties.
The distribution of the latent variables appears to be very complex, nonetheless it yields a nice interpretation. For instance, the mass function of the latent structure $\mathbf{p}$ is the product of two terms: the probability of observing the number of hyper-ties specified by $\mathbf{p}$ times the likelihood that exactly those pairs are formed, through the density function g_0. Thus, thanks to the homogeneity of the original CRV, we observe a separate effect for jumps and locations on this hidden clustering structure. The next Corollary shows how the posterior distribution of the normalized measures can be deducted from Theorem 3.2.

Corollary 3.2. *Consider the same setting of Theorem 3.2. Then the conditional distribution of p_1 in (3.4), given $\boldsymbol{X}_{1:n}$, $\boldsymbol{Y}_{1:m}$ and the appropriate latent variables is given by*

$$w_1 \frac{\hat{\mu}_1}{T_1} + w_2 \frac{\sum_{(i,j)\in\Delta_p} J^1_{i,j}\delta_{(X^*_i,Y^*_j)}}{\sum_{(i,j)\in\Delta_p} J^1_{i,j}}$$

$$+ w_3 \frac{\sum_{(i,j)\in\Delta^1_p} J^1_{i,c+1}\delta_{(X^*_i,Z^x_i)}}{\sum_{(i,j)\in\Delta^1_p} J^1_{i,c+1}} + w_4 \frac{\sum_{(i,j)\in\Delta^2_p} J^1_{k+1,j}\delta_{(Z^y_j,Y^*_j)}}{\sum_{(i,j)\in\Delta^2_p} J^1_{k+1,j}},$$

where $T_1 = \hat{\mu}_1(\mathbb{X}\times\mathbb{X})$, while

$$w_1 \propto T_1, \quad w_2 \propto \sum_{(i,j)\in\Delta_p} J^1_{i,j}, \quad w_3 \propto \sum_{(i,j)\in\Delta^1_p} J^1_{i,c+1}, \quad w_4 \propto \sum_{(k+1,j)\in\Delta^2_p} J^1_{i,j},$$

with the constraint $\sum_{i=1}^4 w_i = 1$. The expressions for $\hat{\mu}_1$, the jumps $J^1_{i,j}$ and the latent variables can be found in the statement of Theorem 3.2.

Proof. We use the short notation $\mu_1(f) = \int_{\mathbb{X}} f(x)\,\mu_1(\mathrm{d}x)$ for any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $\mu_1(|f|) < \infty$. Denoting $\mathbf{U}$ the set of latent variables of Theorem 3.2, i.e. $\mathbf{U} = (\mathbf{p}, U_1, U_2, Z^x, Z^y)$ for any $y_1,\dots,y_n \in (0,1)$ and $A_1,\dots,A_n \in \mathcal{X}^2$ we get

$$\begin{aligned} &\mathbb{P}\left[p_1(A_1) \le y_1,\dots,p_n(A_n) \le y_n \mid \mathbf{U}, \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}\right] = \\ &= \mathbb{P}\left[\mu_1(\mathbb{1}_{A_1} - y_1) \le 0,\dots,\mu_n(\mathbb{1}_{A_n} - y_n) \le 0 \mid \mathbf{U}, \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}\right]. \end{aligned} \tag{3.27}$$

The result follows since the finite dimensional distributions of p_1 given $\mathbf{U}, \boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$ coincide with the ones of the normalized posterior distribution of μ_1, given $\mathbf{U}, \boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$. □

Of course an analogous representation holds for p_2.

3.6.1 Predictive structure

While dealing with Bayesian models, it is often of great interest the *predictive distribution*, that is the distribution of new observations given past data. On the one hand, it gives more intuition on how the model behaves and learns, on the other hand, it coincides with the estimate of the distribution under a square loss function. For this reason, it can be used to develop marginal algorithms that avoid the direct sampling of $\tilde{p}_1$ and $\tilde{p}_2$, which are infinite-dimensional objects. In Lemma 3.4 we saw how to sample the first pair of observations. Next theorem illustrates the general case.

Theorem 3.3. *Consider samples $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$ from model (3.8), with the same setting of Theorem 3.2. Then there exist probability weights ξ_0, $\{\xi_i^x\}$ and $\{\xi_j^y\}$ such that for any $C \in \mathcal{X}$ it holds*

$$\mathbb{P}\big(X_{n+1} \in C \mid \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}\big) = \xi_0 P_0(C) + \sum_{i=1}^{k} \xi_i^x \delta_{X_i^*}(C) + \sum_{j=1}^{c} \xi_j^y P_{Y_j^*}(C)\,.$$

Analogously, there exist probability weights η_0, $\{\eta_i^x\}$ and $\{\eta_j^y\}$ such that for any $C \in \mathcal{X}$ it holds

$$\mathbb{P}\big(Y_{m+1} \in C \mid \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}\big) = \eta_0 P_0(C) + \sum_{j=1}^{c} \eta_j^y \delta_{Y_j^*}(C) + \sum_{i=1}^{k} \eta_i^x P_{X_i^*}(C)\,.$$

Explicit formulae for the weights are available here below in the proof.

Proof. Denote $\mathbf{U} = (\mathbf{p}, U_1, U_2)$ with domain D. Then

$$\mathbb{P}(X_{n+1} \in \mathrm{d}x \mid \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}) = \mathbb{E}[\tilde{p}_1(\mathrm{d}x) \mid \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}] = \int_D \mathbb{E}[\tilde{p}_1(\mathrm{d}x) \mid \mathbf{U} = \mathbf{u}, \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}]\, F(\mathrm{d}\mathbf{u}),$$

where $F(\cdot)$ is the distribution of $\mathbf{U}$ a posteriori with $\mathbf{u} = (\mathbf{p}, u_1, u_2)$. Recalling the notation in Corollary 2 we have

$$\begin{aligned}
\mathbb{E}[\tilde{p}_1(\mathrm{d}x) \mid \mathbf{U} = \mathbf{u}, \boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}] =& \mathbb{E}\left[\frac{\hat{\mu}_1(\mathrm{d}x \times \mathbb{X})}{R}\right] + \mathbb{E}\left[\frac{\sum_{(i,j)\in\Delta_{\mathbf{p}}} J_{i,j}^1 \delta_{X_i^*}}{R}\right] \\
&+ \mathbb{E}\left[\frac{\sum_{(i,j)\in\Delta_{\mathbf{p}}^1} J_{i,c+1}^1 \delta_{X_i^*}}{R}\right] + \mathbb{E}\left[\frac{\sum_{(i,j)\in\Delta_{\mathbf{p}}^2} J_{k+1,j}^1 \delta_{Z_j^y}}{R}\right] \\
&= \sum_{k=1}^{4} I_k,
\end{aligned}$$

where $R = T_1 + \sum_{(i,j)\in\Delta_{\mathbf{p}}} J_{i,j}^1 + \sum_{(i,j)\in\Delta_{\mathbf{p}}^1} J_{i,c+1}^1 + \sum_{(i,j)\in\Delta_{\mathbf{p}}^2} J_{k+1,j}^1$. Denoting $S = \sum_{(i,j)\in\Delta_{\mathbf{p}}} J_{i,j}^1 + \sum_{(i,j)\in\Delta_{\mathbf{p}}^1} J_{i,c+1}^1 + \sum_{(i,j)\in\Delta_{\mathbf{p}}^2} J_{k+1,j}^1$ and exploiting the conditional independence between J_{ij}^1

and $\hat{\mu}_1$ we have

$$
\begin{aligned}
I_1 &= \int_{\mathbb{R}_+} \mathbb{E}\left[e^{-vS}\right] \mathbb{E}\left[\hat{\mu}_1(\mathrm{d}x \times \mathbb{X}) e^{-vT_1}\right] \mathrm{d}v \\
&= \theta P_0(\mathrm{d}x) \int_{\mathbb{R}_+} \left(\prod_{(i,j)\in\mathbf{p}} \frac{\tau_{n_i,m_j}(u_1+v,u_2)}{\tau_{n_i,m_j}(u_1,u_2)} \right) \tau_{1,0}(u_1+v,u_2) e^{-\psi_b^{\mathbf{u}}(v,0)} \,\mathrm{d}v,
\end{aligned}
$$

where $\psi_b^{\mathbf{u}}(\lambda_1,\lambda_2)$ is the Laplace exponent of $(\hat{\mu}_1,\hat{\mu}_2)$ in Theorem 3.2. Noticing that $\psi_b^{\mathbf{u}}(v,0) + \psi(u_1,u_2) = \psi(u_1+v,u_2)$ we obtain

$$
\begin{aligned}
\xi_0 &= \int_D I_1 \, F(\mathrm{d}\mathbf{u}) \\
&= \theta P_0(\mathrm{d}x) \int\int_{\mathbb{R}_+^3} u_1^{n-1} u_2^{m-1} \left(\prod_{(i,j)\in\mathbf{p}} \tau_{n_i,m_j}(u_1+v,u_2) \right) \tau_{1,0}(u_1+v,u_2) e^{-\psi(u_1+v,u_2)} \,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v L(\mathrm{d}\mathbf{p}) \\
&= \frac{\theta P_0(\mathrm{d}x)}{n} \int\int_{\mathbb{R}_+^2} u_1^{n} u_2^{m-1} \left(\prod_{(i,j)\in\mathbf{p}} \tau_{n_i,m_j}(u_1,u_2) \right) \tau_{1,0}(u_1,u_2) e^{-\psi(u_1,u_2)} \,\mathrm{d}u_1\mathrm{d}u_2 L(\mathrm{d}\mathbf{p}) \\
&= \frac{\theta P_0(\mathrm{d}x)}{n} \int_D u_1 \tau_{1,0}(u_1,u_2) \, F(\mathrm{d}\mathbf{u}),
\end{aligned}
$$

where $L(\cdot)$ is the distribution of $\mathbf{p}$ and the second equality follows from the change of variables $(w,z) = (u_1+v,u_1)$. The proof for the remaining weights follows the same lines and leads to

$$
\xi_i^x = \frac{1}{n} \int_D u_1 \left[\frac{\tau_{n_i+1,m_j}(u_1,u_2)}{\tau_{n_i,m_j}(u_1,u_2)} + \frac{\tau_{n_i+1,0}(u_1,u_2)}{\tau_{n_i,0}(u_1,u_2)} \right] F(\mathrm{d}\mathbf{u})
$$

and

$$
\xi_i^y = \frac{1}{n} \int_D u_1 \frac{\tau_{1,m_j}(u_1,u_2)}{\tau_{0,m_j}(u_1,u_2)} \, F(\mathrm{d}\mathbf{u}).
$$

The weights for Y_{m+1} can be analogously computed. □

The predictive distributions have a quite intuitive form, since they are linear combinations of the centering distribution P_0, a weighted version of the empirical distribution and a last term that depends on the other sample. It is somewhat similar to the structure of the exchangeable case (James et al., 2009), with the addition of the last term; this shows very clearly how posterior inference changes incorporating heterogeneous information and performing borrowing of information.

Example 3.7 (n-FuRBI with equal atoms)**.** *If the joint distribution G_0 is degenerate such that the atoms are completely shared between $\tilde{p}_1$ and $\tilde{p}_2$, then $P_Z(\cdot) = \delta_Z(\cdot)$. Therefore, the last term in Theorem 3.3 becomes a weighted version of the empirical distribution relative to the other sample.*

3.7 Sampling methods for FuRBIs

The problem of sampling from the posterior distributions in models that involve infinite dimensional parameters have been extensively studied and can in general be approached in two different ways. Thanks to *conditional* algorithms, it is possible to sample approximations of the infinite dimensional object exploiting its series representation (see, for instance Ishwaran & James, 2001; Arbel & Prünster, 2017). Alternatively, one can use *marginal* algorithms, that integrate out the random measures and sample sequentially from the predictive distributions (see, for instance, Neal, 2000).

3.7.1 Conditional samplers

To develop a conditional algorithm, we can sample from the distribution of $(\tilde{\mu}_1, \tilde{\mu}_2)$ and then normalize each draw to get an approximate realization of the random probabilities. We develop below a general conditional sampler based on this approach that can be tailored to specific choices of the intensity in the prior.
Alternatively, a second strategy is to sample approximate draws from the posterior distribution of the random probabilities $(\tilde{p}_1, \tilde{p}_2)$. We provide an example for Gamma FuRBI CRMs with equal jumps.

Conditional sampler based on the law of the CRV

By Theorem 3.2 we know that a posteriori $\mu = (\mu_1, \mu_2)$ is the sum of two components, that we call μ_{obs} and $\hat{\mu}$ and are such that

$$\mu_{obs} = \sum_{(i,j)\in\Delta_p} J_{i,j}\delta_{(X_i^*, Y_j^*)} + \sum_{(i,j)\in\Delta_p^1} J_{i,c+1}\delta_{(X_i^*, Z_i^x)} + \sum_{(i,j)\in\Delta_p^2} J_{k+1,j}\delta_{(Z_j^y, Y_j^*)}.$$

where $J_{i,j} = (J_{i,j}^1, J_{i,j}^2)$, and

$$\hat{\mu} = \left(\sum_{h=1}^{+\infty} S_h^1 \delta_{(V_h, W_h)}, \sum_{h=1}^{+\infty} S_h^2 \delta_{(V_h, W_h)}\right)$$

where $\hat{\mu}$ is a CRV with Lévy intensity $e^{-U_1 s_1 - U_2 s_2}\rho(\mathrm{d}s_1, \mathrm{d}s_2)G_0(\mathrm{d}x)$. Define the marginal and joint tail integrals of $\hat{\mu}$ as

$$N_1(s) = \int_s^{+\infty}\int_0^{+\infty} e^{-U_1 s_1 - U_2 s_2}\rho(\mathrm{d}u_1, \mathrm{d}u_2)$$

$$N_2(s) = \int_0^{+\infty}\int_s^{+\infty} e^{-U_1 s_1 - U_2 s_2}\rho(\mathrm{d}u_1, \mathrm{d}u_2)$$

and

$$N(s_1, s_2) = \int_{s_1}^{+\infty} \int_{s_2}^{+\infty} e^{-U_1 s_1 - U_2 s_2} \rho(\mathrm{d}u_1, \mathrm{d}u_2)$$

Lastly, define the correspondent Lévy copula as

$$F(x, y) = N(N_1^{-1}(x), N_2^{-1}(y))$$

If $F(x, y)$ is continuous on $[0, +\infty]^2$, the iterative conditional sampler based on Ferguson and Klass algorithm (Ferguson & Klass, 1972) reads

(a) Generate μ_{obs} as follows

(a1) Generate $(U_1, U_2, \mathbf{p})$ from the distributions specified in Theorem 3.2;

(a2) Generate $J_{i,j} = (J^1_{i,j}, J^2_{i,j})$ from the distributions specified in Theorem 3.2;

(a3) Generate Z^x_i and Z^y_j from the distributions specified in Theorem 3.2.

(b) Generate an approximation of $\hat{\mu}$, given by $\left(\sum_{h=1}^{M} S^1_h \delta_{(V_h, W_h)}, \sum_{h=1}^{M} S^2_h \delta_{(V_h, W_h)} \right)$ as follows

(b1) Generate $\xi^x_1, \ldots, \xi^x_M$ from a Poisson process with unit rate;

(b2) Generate $\xi^y_1, \ldots, \xi^y_M$ from $\xi^y_h \sim \frac{\partial}{\partial x} F(x, \xi) \Big|_{x=\xi^x_h}$

(b3) Determine (S^1_h, S^2_h) solving

$$\xi^x_h = N_1(S^1_h) \qquad \xi^y_h = N_2(S^2_h)$$

(b4) Generate (V_h, W_h) from G_0.

(c) Obtain a draw from $\tilde{p}_1$ as follows

$$\tilde{p}_1 \approx \frac{\sum_{h=1}^{M} S^1_h \delta_{V_h} + \sum_{(i,j)\in\Delta_p} J^1_{i,j} \delta_{X^*_i} + \sum_{(i,j)\in\Delta^1_p} J^1_{i,c+1} \delta_{X^*_i} + \sum_{(i,j)\in\Delta^2_p} J^1_{k+1,j} \delta_{Z^y_j}}{\sum_{h=1}^{M} S^1_h + \sum_{(i,j)\in\Delta_p} J^1_{i,j} + \sum_{(i,j)\in\Delta^1_p} J^1_{i,c+1} + \sum_{(i,j)\in\Delta^2_p} J^1_{k+1,j}}.$$

An analogous approximation can be computed for $\tilde{p}_2$.

Gamma process with equal jumps

In the case of a process with equal jumps, we know from the definition that the measures in the product space are $p_1 = p_2 = p$. Therefore, posterior inference can be conducted without

loss of generality on

$$p = \sum_{k\geq 1} \bar{W}_k \delta_{(\theta_k,\phi_k)}, \quad \text{with } (\theta_k, \phi_k) \overset{\text{i.i.d.}}{\sim} G_0(\cdot),$$

where $\{\bar{W}_k\}_k$ are the weights of a Dirichlet process, which can defined through the celebrated stick-breaking construction (Sethuraman, 1994). In this context, Ishwaran & James (2001) developed a conditional algorithm for hierarchical mixture models, called *blocked Gibbs* sampler, based on the approximation

$$p \approx \sum_{k=1}^{N} \bar{W}_k \delta_{(\theta_k,\phi_k)}, \quad \text{with } N \text{ large}.$$

Exploiting the nice analytical properties of the Dirichlet process, it is possible to devise simple formulae for the posterior distribution of the N jumps and N locations: see Section 5 of Ishwaran & James (2001) for more details.

3.7.2 Marginal algorithms

Given $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$ and using the results in Theorem 3.3, we can sample iteratively new observations from $\tilde{p}_1$ as follows.

(a) Compute weights ξ_0, $\{\xi_i^x\}$ and $\{\xi_j^y\}$ from $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$.

(b) Draw X from

$$m(\mathrm{d}x) = \xi_0 P_0(\mathrm{d}x) + \sum_{i=1}^{k} \xi_i^x \delta_{X_i^*}(\mathrm{d}x) + \sum_{j=1}^{c} \xi_j^y P_{Y_j^*}(\mathrm{d}x)\,.$$

(c) Add X to $\boldsymbol{X}_{1:n}$.

The algorithm is straightforward, but relies on the computation of the weights at point (a): this is not optimal, since in general the explicit evaluation can be demanding. Nonetheless, Theorem 3.2 and Corollary 3.2 show that, conditional on a suitable set of latent variables, the posterior representation simplifies greatly. Indeed, given $(\boldsymbol{X}_{1:n}, \boldsymbol{Y}_{1:m}, U_1, U_2, \mathbf{p})$, the predictive distribution of the first sample is given by

$$\begin{aligned} m(\mathrm{d}x) \propto & \theta \tau_{1,0}(U_1,U_2) P_0(\mathrm{d}x) + \sum_{(i,j)\in\Delta_p} \frac{\tau_{n_i+1,m_j}(U_1,U_2)}{\tau_{n_i,m_j}(U_1,U_2)} \delta_{X_i^*}(\mathrm{d}x) \\ & + \sum_{(i,j)\in\Delta_p^1} \frac{\tau_{n_i+1,0}(U_1,U_2)}{\tau_{n_i,0}(U_1,U_2)} \delta_{X_i^*}(\mathrm{d}x) + \sum_{(i,j)\in\Delta_p^2} \frac{\tau_{1,m_j}(U_1,U_2)}{\tau_{0,m_j}(U_1,U_2)} P_{Y_j^*}(\mathrm{d}x). \end{aligned} \tag{3.28}$$

Those new weights are often easier to compute, as the next Example shows.

Example 3.8 (Inverse Gaussian n-FuRBIs with equal jumps)**.** *In this case* $\tau_{n,m}(u_1, u_2) = \int_{\mathbb{R}} s^{n+m} e^{-(u_1+u_2)s} \rho(\mathrm{d}s)$*, where* $\rho(\mathrm{d}s)$ *is the common marginal jump intensity. If the Lévy intensity is given by*

$$v(\mathrm{d}s, \mathrm{d}x) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2}s}}{s^{\frac{3}{2}}} \mathrm{d}s\, \alpha(\mathrm{d}x)$$

we call the resulting NRMI normalized inverse gaussian *process (whose application on Bayesian nonparametrics has been analyzed in Lijoi et al. (2005)). It holds*

$$\tau_j(u_1, u_2) = \frac{2^{j-1}\Gamma\left(j - \frac{1}{2}\right)}{\sqrt{\pi}(2u+1)^{j-\frac{1}{2}}}, \quad u = u_1 + u_2.$$

Thus, conditional on the usual latent variables we have

$$\begin{aligned} m(\mathrm{d}x) \propto & \theta P_0(\mathrm{d}x) + \frac{2}{\sqrt{2U+1}} \sum_{(i,j)\in\Delta_p} \left(n_i + m_j - \frac{1}{2}\right) \delta_{X_i^*}(\mathrm{d}x) \\ & + \frac{2}{\sqrt{2U+1}} \sum_{(i,j)\in\Delta_p^1} \left(n_i - \frac{1}{2}\right) \delta_{X_i^*}(\mathrm{d}x) + \frac{2}{\sqrt{2U+1}} \sum_{(i,j)\in\Delta_p^2} \left(m_j - \frac{1}{2}\right) P_{Y_j^*}(\mathrm{d}x), \end{aligned}$$

where $U = U_1 + U_2$*. Sampling from this mixture is straightforward.*

Thus we can derive a new marginal algorithm that reads:

(a) Draw $(U_1, U_2, \mathbf{p})$ given $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$, from the distributions specified in Theorem 3.2.

(b) Draw X from $m(\mathrm{d}x)$ in (3.28).

(c) Add X to $\boldsymbol{X}_{1:n}$.

However, even the full conditional distribution of $\boldsymbol{p}$ may not always be available in closed form, and it may be computationally intensive to evaluate, since it may have a very large support. When this is the case, we may encode the latent clustering structure in a more convenient way introducing two arrays of latent variables $\boldsymbol{c_x} = (c_{i,x})_{i\geq 1}$ and $\boldsymbol{c_y} = (c_{j,y})_{j\geq 1}$ such that: $c_{i,x} = c_{i',x}$ denotes a tie between X_i and $X_{i'}$, $c_{j,y} = c_{j',y}$ denotes a tie between Y_j and $Y_{j'}$, while $c_{i,x} = c_{j,y}$ denotes a hyper-tie between X_i and Y_j. Moreover, we reorder the unique values in $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$, so that

$$X_c^* = X_i \quad \text{iff} \quad c_{i,x} = c \qquad \text{and} \qquad Y_c^* = Y_j \quad \text{iff} \quad c_{j,y} = c$$

So that

$$\mathbb{P}[c_{n+1,x} = c \,|\, \boldsymbol{c_x}, \boldsymbol{c_y}, \boldsymbol{X}^*, \boldsymbol{Y}^*] = \begin{cases} \mathbb{P}[X_{n+1} = X^*_c \,|\, \boldsymbol{c_x}, \boldsymbol{c_y}, \boldsymbol{X}^*, \boldsymbol{Y}^*], & \text{for } c \in \boldsymbol{c}_{\boldsymbol{x}}^{-(i)} \\ \displaystyle\int \mathbb{P}[X_{n+1} = x \mid \boldsymbol{c_y}, \boldsymbol{Y}^*]\, p_{Y^*_c}(x)\mathrm{d}x, & \text{for } c \in \boldsymbol{c_y} \setminus \boldsymbol{c}_{\boldsymbol{x}}^{-(i)} \\ \displaystyle\int \mathbb{P}[X_{n+1} = x]\, p_0(x)\mathrm{d}x, & \text{otherwise} \end{cases}$$

Finally notice that the distribution of $\boldsymbol{p}$, given $\boldsymbol{c_x}$ and $\boldsymbol{c_y}$, is degenerate. Moreover, the posterior distribution of (U_1, U_2) given $\boldsymbol{p}$ is equal to the posterior distribution of (U_1, U_2) given $(\boldsymbol{c_x}, \boldsymbol{c_y})$. Therefore, we may build a marginal sampler sampling $\boldsymbol{c_x}$, $\boldsymbol{c_y}$ instead of $\boldsymbol{p}$, without modifying the full conditional distribution for U_1 and U_2. The final marginal algorithm reads:

(a) Draw (U_1, U_2), $\boldsymbol{c_x}$ and $\boldsymbol{c_y}$.

(b) Draw $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$.

The advantage of such approach is twofold. Firstly, we do not need to sample directly for the full conditional distribution of $\boldsymbol{p}$. Secondly, sampling $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$, instead of $\boldsymbol{X}$ and $\boldsymbol{Y}$, improves the mixing of the algorithm (cf. Neal, 2000).

3.8 Illustration

3.8.1 Bayesian mixture models

Discrete Bayesian models, as the one specified in (3.8), are usually not employed directly on the data, but as a building block in hierarchical mixture models: in this setting $\boldsymbol{X}$ and $\boldsymbol{Y}$ are hidden values that describes the clustering structure within the data, as already mentioned in Section 1.2.2. Such models have been introduced by Ferguson (1983) and Lo (1984) for the Dirichlet processes and gained popularity thanks also to the availability of MCMC methods for posterior sampling (Escobar & West, 1995; Ishwaran & James, 2001; Neal, 2000). Suppose $\{f(\cdot \mid x) : x \in \mathbb{X}\}$ is a family of non-negative kernels on a Polish space $\mathbb{W}$, such that $\int f(w \mid x)\, \lambda(\mathrm{d}w) = 1$ for a suitable dominating measure λ. Then the model can be formulated as

$$\begin{matrix} W_i \,|X_i \overset{\text{ind}}{\sim} f(\cdot \mid X_i) \\ X_i \mid \tilde{p}_1 \overset{\text{i.i.d.}}{\sim} p_1 \end{matrix}, \quad \begin{matrix} V_j \,|Y_j \overset{\text{ind}}{\sim} f(\cdot \mid Y_j) \\ Y_j \mid \tilde{p}_2 \overset{\text{i.i.d.}}{\sim} p_2 \end{matrix}, \quad (\tilde{p}_1, \tilde{p}_2) \sim \text{n-FuRBI}. \tag{3.29}$$

where $(W_i)_{i=1}^n$ and $(V_j)_{j=1}^m$ are the observable samples. Integrating out the latent variables X_i and Y_j, the data are exchangeable draws from suitable countable mixtures, i.e.

$$W_i \mid \tilde{p}_1 \overset{iid}{\sim} \int f(\cdot \mid x)\, \tilde{p}_1(\mathrm{d}x), \quad V_j \mid \tilde{p}_2 \overset{iid}{\sim} \int f(\cdot \mid y)\, \tilde{p}_2(\mathrm{d}y).$$

Example 3.9 (Gaussian mixtures). *We assume $f(\cdot \mid x) := N(\cdot \mid x, \sigma^2)$, with σ^2 positive known constant, to be the normal density. Thus, the latent parameter is given by the mean, i.e. $\mathbb{X} = \mathbb{R}$. In this case*

$$Cov(X_i, Y_j) = Cov(W_i, V_j),$$

so that the joint behavior of the latent means is reflected on the observations: this shows the importance of the correlation structure given by Proposition 3.1 also for hierarchical models. As a generalization, the latent parameters could specify both the mean and the variance, with $\mathbb{X} = \mathbb{R} \times \mathbb{R}_+$.

Clearly the posterior distribution given samples $\mathbf{W}_{1:n} = (W_1, \ldots, W_n)$ and $\mathbf{V}_{1:m} = (V_1, \ldots, V_m)$ is of interest: however it requires to integrate out all the possible partitions of the $n+m$ latent variables. Luckily, the sampling schemes developed in the end of the previous section allow us to set up a Gibbs sampler for drawing from the posterior distribution of $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$.
For instance, denoting $\boldsymbol{X}^t = (X_1^t, \ldots, X_n^t)$ and $\boldsymbol{Y}^t = (Y_1^t, \ldots, Y_n^t)$ the vectors sampled at step t, the marginal algorithm reads

1. Initialize at random $\boldsymbol{X}^0$ and $\boldsymbol{Y}^0$.

2. For any $t \geq 1$ do:

(a) Draw $(U_1, U_2, \mathbf{p})$ given $\boldsymbol{X}^{t-1}$ and $\boldsymbol{Y}^{t-1}$, from the distributions specified in Theorem 3.2.

(b) Draw $\boldsymbol{X}^t$ as follows: for any i sample X_i^t from

$$q(\mathrm{d}x \mid \boldsymbol{X}_{-i}^t) = q_{i,0}(U_1, U_2) P_0(\mathrm{d}x) + \sum_{(i,j) \in \Delta_{\mathbf{p}}} q_{i,j}(U_1, U_2) \delta_{X_i^*} + \sum_{(i,j) \in \Delta_{\mathbf{p}}^1} q_{i,j}^1(U_1, U_2) \delta_{X_i^*}(\mathrm{d}x) + \sum_{(i,j) \in \Delta_{\mathbf{p}}} q_{i,j}^2(U_1, U_2) P_{Y_j^*}(\mathrm{d}x),$$

where $\boldsymbol{X}_{-i}^t = \left(X_1^t, \ldots, X_{i-1}^t, X_{i+1}^{t-1}, \ldots X_n^{t-1}\right)$, with unique values $(X_1^*, \ldots, X_k^*)$ and multiplicities $(n_1, \ldots, n_k)$. Analogously, $(Y_1^*, \ldots, Y_c^*)$ denotes the unique values in $\boldsymbol{Y}^{t-1}$ with multiplicities $(m_1, \ldots, m_c)$. The mixing proportions are given by

$$\begin{aligned}
q_{i,0}(U_1, U_2) &\propto \theta \tau_{1,0}(U_1, U_2) \int_{\mathbb{X}} f(W_i \mid x) P_0(\mathrm{d}x), \\
q_{i,j}(U_1, U_2) &\propto \frac{\tau_{n_i+1, m_j}(U_1, U_2)}{\tau_{n_i, m_j}(U_1, U_2)} f(W_i \mid X_i^*), \\
q_{i,j}^1(U_1, U_2) &\propto \frac{\tau_{n_i+1, 0}(U_1, U_2)}{\tau_{n_i, 0}(U_1, U_2)} f(W_i \mid X_i^*), \\
q_{i,j}^2(U_1, U_2) &\propto \frac{\tau_{1, m_j}(U_1, U_2)}{\tau_{0, m_j}(U_1, U_2)} \int_{\mathbb{X}} f(W_i \mid x) P_{Y_j^*}(\mathrm{d}x)
\end{aligned}$$

(c) Sample $\boldsymbol{Y}^t$ similarly to point (b).

Once a posterior sample for $\boldsymbol{X}_{1:n}$ and $\boldsymbol{Y}_{1:m}$ is collected, relevant quantities of interest can be approximated, exploiting the independence within $\mathbf{W}$ and $\mathbf{V}$ given the latent variables.

3.8.2 Simulation study

We consider a simple application with simulated data, in order to understand how inference changes taking into account heterogeneous sources of information. Assume the following generating mechanism for two independent samples $\mathbf{W}_{1:n}$ and $\mathbf{V}_{1:m}$

$$\begin{aligned} W_i &\overset{\text{i.i.d.}}{\sim} N(\cdot \mid 10, 1), \quad i = 1, \dots, 20, \\ V_j &\overset{\text{i.i.d.}}{\sim} N(\cdot \mid -10, 1), \quad j = 1, \dots, 100. \end{aligned} \tag{3.30}$$

Supposing only the phenomenon associated to the first sample is of interest, hierarchical mixtures are considered to make prediction on the unknown density of W_i. The kernel considered is the one specified in Example 3.9, with known $\sigma^2 = 1$ and latent mean μ. Four different approaches for modelling dependence between $\mathbf{W}$ and $\mathbf{V}$ are devised

1. *Exchangeable* approach: observations in $\mathbf{W}$ and $\mathbf{V}$ are supposed to be exchangeable, inducing the highest positive correlation between W_i and V_j.

2. *Independent* approach: the sample $\mathbf{V}$ is disregarded entirely, that is $\mathbf{W}$ and $\mathbf{V}$ are assumed to be independent.

3. *Hierarchical* approach: the dependence between $\mathbf{W}$ and $\mathbf{V}$ is described by a hierarchical Dirichlet process (see Example 3.1). This approach corresponds to a classical borrowing of information.

4. *FuRBI* approach: the underlying random probability measures $\tilde{p}_1$ and $\tilde{p}_2$ have a joint distribution as specified in Definition 3.1. In particular, we consider the case of equal weights with distribution on the atoms given as

$$G_0(\cdot \mid \rho_0) = N_2\left(\cdot \mid \underline{0}, 1, \rho_0\right), \quad \rho_0 \sim Unif([-1, 1]), \tag{3.31}$$

   where $N_2(\cdot \mid \underline{m}, \sigma_0^2, \rho_0)$ denotes the bivariate normal distribution with mean vector $\underline{m}$, common variance σ_0^2 and correlation ρ_0. It can be proven that under this specification $\text{Corr}(W_i, V_j) = 0$, so that a priori $\mathbf{W}$ and $\mathbf{V}$ are marginally uncorrelated.

For the first two cases and the n-FuRBI, the marginal distribution is given by a Dirichlet process with $\theta = 1$ and $P_0(\cdot) = N(\cdot \mid 0, 1)$; instead for the hierarchical process the concentration parameters are fixed in order to match the expected number of different clusters with the other methods, for a fair comparison. As it was highlighted in Example 3.6, n-FuRBI with equal jumps lead to the most general setting in terms of achievable correlation

between samples; moreover, choosing the marginal processes to derive from a Gamma process, we can achieve any value in the interval $(-1, 1)$, tuning appropriately the concentration parameter θ. Thus, it can be a competitive choice for modelling purposes.

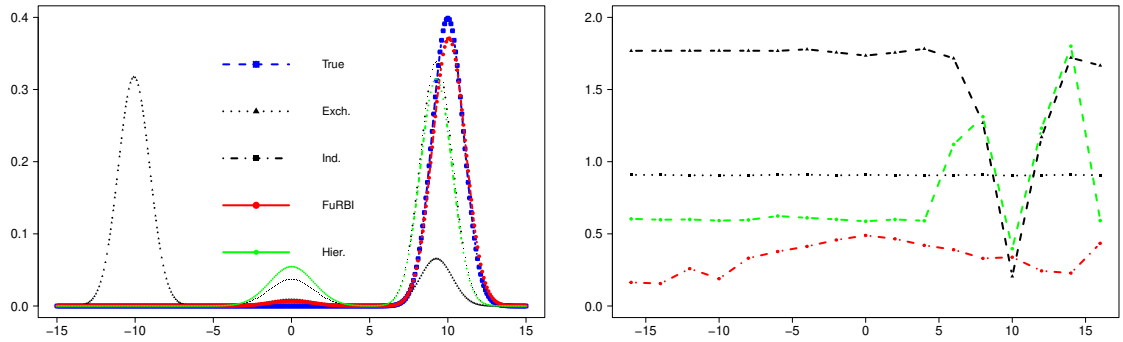


Figure 3.3: Left: estimated densities for the case with opposite true means. Right: mean integrated error (computed on a grid) for the four estimates, varying the true mean of $\mathbf{V}$.

The left panel of Figure 3.3 shows the performances of the four methods, after the application of the Blocked Gibbs sampler. The exchangeable approach behaves very badly, as expected, because the two samples have clearly a different distribution. The independent choice leads to a reasonable estimate, even if it still overestimates the probability mass around the prior mean (because of the small sample size of the first sample). The hierarchical estimate is quite good, but our proposal, instead, fits almost perfectly the target density and seems to exploit the opposite behaviour of the two phenomena: this is clearly highlighted by the posterior distribution of ρ_0 in (3.31), whose approximated mean is close to -0.9.

One may wonder whether these superior performances follow from the precise specification above, with opposite true means. Therefore, we repeated the experiment with the same formulation of (3.30), with the true mean of $\mathbf{V}$ ranging from -16 to 16: the mean integrated absolute error is depicted in the right panel of Figure 3.3.

It is apparent that the FuRBI approach almost always yields the smallest error, regardless of the true value. Its performance is close to the exchangeable case only when the two true means are equal, that is when exchangeability actually holds. The hierarchical process captures the right dependence when the two means coincide, but can be misled when they are close; finally, when the second sample is very far from the first one it performs better than the independent sampler, probably thanks to the different inner clustering structure. The results are also summarized in Table 3.1.

Mean of **V**	Exch.	Ind.	FuRBI	Hier.
-15	1.769	0.995	**0.225**	0.730
-10	1.769	0.995	**0.119**	0.759
0	1.737	0.995	**0.637**	0.756
10	**0.267**	0.995	0.288	0.412
15	1.649	0.995	**0.299**	0.740

Table 3.1: Mean integrated absolute error associated to the four methods for some values of the mean of **V**. The values in bold are the smallest ones for each row.

Thus, FuRBI models seem to be always capable of combining heterogeneous information in the right way; in particular, at least in this example, they recognize the most useful type of borrowing of information. Finally, we consider a similar application with three groups, in order to see whether the n-FuRBIs are able to discern more complex types of dependence. We assume to observe

$$\begin{aligned} W_i &\overset{\text{i.i.d.}}{\sim} N(\cdot \mid 10, 1), \quad i = 1, \dots, 20, \\ V_j &\overset{\text{i.i.d.}}{\sim} N(\cdot \mid -10, 1), \quad j = 1, \dots, 20, \\ R_j &\overset{\text{i.i.d.}}{\sim} N(\cdot \mid x, 1), \quad j = 1, \dots, 20, \end{aligned} \tag{3.32}$$

with $x \in \{-10, -9, \dots, 10\}$. Then, for each value of x we apply n-FuRBIs with same weights, as before, but where the atoms are distributed as follows

$$G_0\left(\cdot|\rho_{12}, \rho_{13}, \rho_{23}\right)) = N_3\left(\cdot \middle| \underline{0}, 1, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}\right), \tag{3.33}$$

$$\rho_{12}, \rho_{13}, \rho_{23} \overset{\text{i.i.d.}}{\sim} Unif([-1, 1])$$

where $N_3(\cdot \mid \mu_0, \sigma^2, \Psi)$ denotes a multivariate normal distribution with mean μ_0, all the variances equal to σ^2 and correlation matrix Ψ. The posterior medians of ρ_{12}, ρ_{13} and ρ_{23} are depicted in figure 3.4, for any value of x.
The results are in line with our intuition: the correlation between the first and second component is always close to -1 (indeed they have opposite behaviour), while ρ_{13} and ρ_{23} vary linearly with x, being positive when the means have the same sign.

3.8.3 Stocks and commodities returns

Data

We collected monthly returns of January 2021 for a sample of 49 stocks portfolios from the Kenneth R. French's Data Library (data available at `http://mba.tuck.dartmouth.`

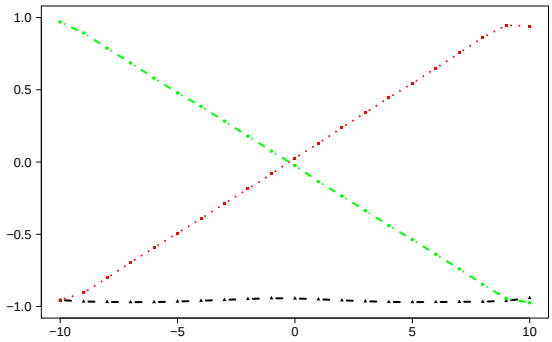


Figure 3.4: Posterior median of the correlation (obtained through 100 samples) between the three unknown means. In black: correlation between the first and second component. In red: correlation between the first and third component. In green: correlation between the second and third component.

edu/pages/faculty/ken.french/data_library.html) and for a sample of 55 commodities from the Primary Commodity Prices Database of the International Monetary Fund (data available at https://www.imf.org/en/Research/commodity-prices).

Stocks and commodities exhibit correlation that largely varies over time ranging from positive to negative values (see, for instance, Bhardwaj & Dunsby, 2013, and Figure 3.5). Inference on their densities is therefore an interesting application of our model. Indeed, commodities returns contain useful information to make inference over the distribution of stocks portfolios; however, in periods of negative correlation, the classical borrowing of information may not be appropriate.

The intuitive idea is the following: if the observed commodities returns outperform our prior guess and we are in a period of negative correlation, an appropriate model should increase the probability of observing stocks returns which underperform with respect the prior guess and vice versa.

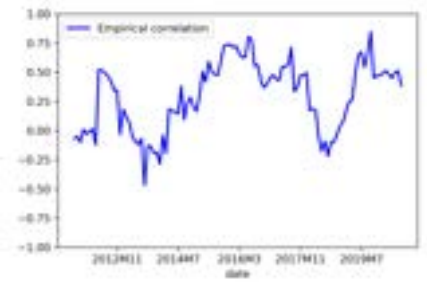


Figure 3.5: Empirical correlation between average stock return and average commodity return computed on a moving window of 12 months using data from March 2011 to January 2021.

Additive FuRBI mixture model

Coherently with Bayesian mixture models, we assume that data come from two mixtures of normals with unknown mean and variance. Denoting with W_i and V_j the monthly returns

of respectively stocks and commodities, we have

$$\begin{aligned} W_i \mid \tilde{p}_1 &\overset{iid}{\sim} \int N(\cdot \mid x, \sigma_w^2)\, \tilde{p}_1(\mathrm{d}x, \mathrm{d}\sigma_w^2) \\ V_j \mid \tilde{p}_2 &\overset{iid}{\sim} \int N(\cdot \mid y, \sigma_v^2)\, \tilde{p}_2(\mathrm{d}y, \mathrm{d}\sigma_v^2) \\ (\tilde{p}_1, \tilde{p}_2) \mid \theta, z, G_0 &\sim \text{n-FuRBI}(\theta, \rho(\mathrm{d}s_1, \mathrm{d}s_2), G_0) \\ \theta &\sim \text{Gamma}(\alpha, \beta) \end{aligned} \tag{3.34}$$

The base measure G_0 is chosen so that marginal distribution are given by NRMIs with conjugate Normal-InverseGamma base measure, i.e.

$$\begin{aligned} G_0(\mathrm{d}_x, \mathrm{d}_y, \mathrm{d}\sigma_w^2, \mathrm{d}\sigma_v^2 \mid \rho_0) = N_2(\mathrm{d}x, \mathrm{d}y \mid \underline{m}, \Sigma(\lambda_1, \lambda_2, \sigma_w^2, \sigma_v^2\rho_0)) \\ \times \text{InvGamma}(\mathrm{d}\sigma_w^2 \mid \alpha_1, \beta_1) \times \text{InvGamma}(\mathrm{d}\sigma_v^2 \mid \alpha_2, \beta_2) \end{aligned} \tag{3.35}$$

with

$$\underline{m} = (m_1, m_2)' \qquad \text{and} \qquad \Sigma = \begin{bmatrix} \frac{\sigma_w^2}{\lambda_1} & \rho_0 \frac{\sigma_w}{\lambda_1^{1/2}} \frac{\sigma_v}{\lambda_2^{1/2}} \\ \rho_0 \frac{\sigma_w}{\lambda_1^{1/2}} \frac{\sigma_v}{\lambda_2^{1/2}} & \frac{\sigma_v^2}{\lambda_2} \end{bmatrix}$$

and we use the following joint underlying Lévy intensity

$$v(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x_2) = \{z\,[\rho(\mathrm{d}s_1)\delta_0(\mathrm{d}s_2) + \rho(\mathrm{d}s_2)\delta_0(\mathrm{d}s_1)] + (1-z)\,\rho(\mathrm{d}s_1)\delta_{s_1}(\mathrm{d}s_2)\}\,\theta\, G_0(\mathrm{d}x_1, \mathrm{d}x_2),$$

with

$$z \sim \text{U}(0,\, 1).$$

We name the resulting n-FuRBI *additive n-FuRBI*, since the series representation of the corresponding FuRBI CRMs is given by

$$\tilde{\mu}_1(\cdot) \overset{a.s.}{=} \sum_{k\geq 1} W_k \delta_{\theta_{0,k}} + \sum_{k\geq 1} J_k \delta_{\theta_{1,k}} \qquad\qquad \tilde{\mu}_2(\cdot) \overset{a.s.}{=} \sum_{k\geq 1} W_k \delta_{\phi_{0,k}} + \sum_{k\geq 1} V_k \delta_{\phi_{2,k}},$$

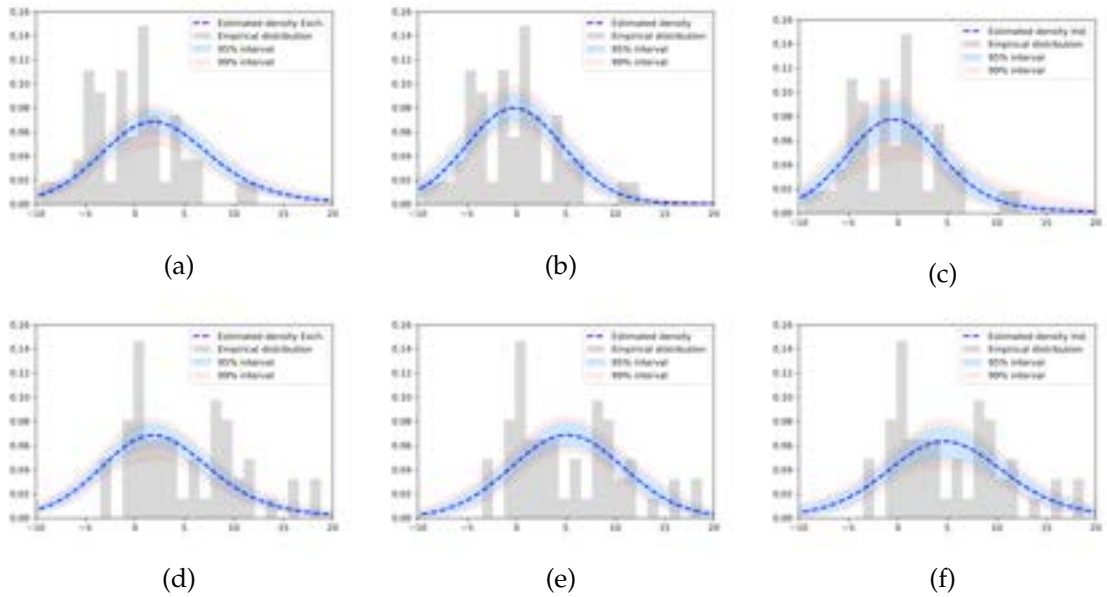
where $(\theta_{0,k}, \phi_{0,k}) \overset{i.i.d}{\sim} G_0$, $\theta_{1,k} \overset{i.i.d}{\sim} P_0$ and $\phi_{2,k} \overset{i.i.d}{\sim} P_0$.
When G_0 is degenerate on the main diagonal (i.e. $\rho_0 = 1$), one retrieves GM-dependent completely random measure (Lijoi et al., 2014a,b; Lijoi & Nipoti, 2014). In order to obtain two Dirichlet processes marginally we set $\rho(s) = s^{-1}e^{-s}$, so that

$$\beta = \frac{1}{1+\theta}$$

and

$$\gamma = (1-z)\frac{\theta}{(1+\theta)^2}\, {}_3F_2(\theta - \theta\, z + 2,\, 1,\, 1;\, \theta + 2,\, \theta + 2;\, 1)$$

where ${}_3F_2$ is the generalized hypergeometric function. In the choice of the hyperparame-



(a) (b) (c)

(d) (e) (f)

Figure 3.6: Panels (a-c) contains density estimates for stocks returns, Panels (d-f) contains density estimates for commodities returns. Panels (a) and (d) are obtained through the exchangeable approach, Panels (b) and (e) are obtained using additive FuRBI, Panels (c) and (f) are obtained with the independent approach.

ters, we try to mimick the situation in which prior knowledge on the marginal distribution can be elicited, for instance through outlooks of the markets. We use returns in the previous month as proxy of prior information and, therefore, we set m_1 and m_2 to the empirical average returns of the two groups in December 2020, which are respectively 5.8591 and 3.9731. Moreover, we assume that no information is available on the correlation and use a uniform prior over $[-1, 1]$ for ρ_0. However, note that eventual information about the correlation could be incorporated through a more-informative prior on the parameter ρ_0. Moreover after standardizing the data, we set the remaining hyperparameters to $\lambda_1 = \lambda_2 = 1$, $\alpha_1 = \alpha_2 = 2$ and $\beta_1 = \beta_2 = 4$. We perform 10 000 iterations of the marginal sampler algorithm and discard the first half as burnin. On the same data we also estimate densities using the exchangeable and independent approach described in the previous section.

	Exch	FuRBI	Ind
ALCPO	-1.4553	**-1.3168**	-1.3752
MLCPO	-1.1672	**-1.1496**	-1.2331

Table 3.2: ALCPO and MLCPO under the three models. The values in bold are the highest for each row.

The results of the analysis are displayed in Figure 3.6. The model employing n-FuRBIs produces the density estimates that most resemble the empirical distributions (see, panel (b) and (e) of Figure 3.6). Since we use a non-informative prior over the correlation ρ_0, intensity and direction of the borrowing of information is inferred from the data, leading to a reinforcement of the in-

formation coming from each sample. To compare the performances of the three models we resort to the conditional predictive ordinates (CPOs) statistics (see, for example, Gelfand et al., 1992; Barrios et al., 2013). Essentially, for each value i, we train the model without the i-th observation and compute the predictive density at the observed point.

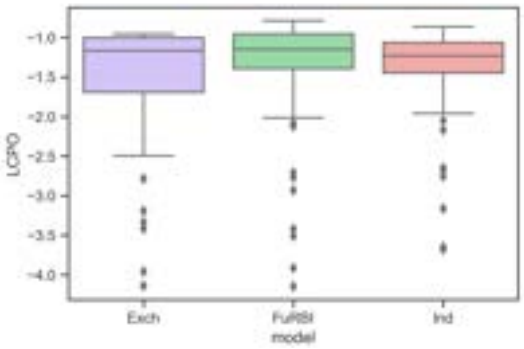


Figure 3.7: Boxplots of the logarithmic CPOs.

For the first sample it reads $\text{CPO}_i^w = \tilde{f}(w_i \mid \boldsymbol{w}^{-i}, \boldsymbol{v})$ for $i = 1, \ldots, n$ and analogously for the second sample, where $\boldsymbol{w}$ and $\boldsymbol{v}$ denote the observed returns for respectively stocks and commodities. Table 3.2 contains the average logarithmic CPO (ALCPO) and the median logarithmic CPO (MLCPO) in the overall sample. The distribution of the logarithmic CPO is depicted in Figure 3.7. Clearly higher values correspond to a better performance and the n-FuRBI turns out to be the best approach.

Chapter 4

Invariant Dependent Processes for Model Selection

In this fourth chapter, we propose a new class of dependent processes, namely *invariant dependent processes*, that as n-FuRBIs in Chapter 3, are not mSSPs. This chapter has a more applied flavor than the previous ones and a specific motivating application. Invariant dependent processes are introduced here in order to define a flexible prior distribution for the error terms of the model considered here. The ultimate aim is to perform model selection to detect the relationship between hypertensive disorders diagnoses and cardiac dysfunctions in pregnant women. The outline of the chapter is as follows. In the first two sections we introduce the motivating application, the framework, and the main ideas on which this chapter is based. In Section 4.3 we introduce the model, which makes use of invariant dependent processes obtained through a hierarchical prior structure for symmetric distributions (Section 4.3.2). In Section 4.4 we derive the prior law of the random partitions induced by the model, key ingredient for the Gibbs sampling scheme devised in Section 4.5. In Section 4.6, we present two alternative priors for the upper clustering of hypertensive disorders, which will be used for comparisons. In Section 4.7, we first present a series of simulation studies that highlight the behaviour of the model before applying it to obtain our results on cardiac dysfunction in hypertensive disorders. Section 4.8 contains some concluding remarks.

4.1 Motivating application

Hypertensive disorders of pregnancy are a class of high blood pressure disorders that occur during the second half of pregnancy, which include gestational hypertension, preeclampsia and severe preeclampsia. They are characterized by a diastolic blood pressure higher than 90 mm Hg and/or a systolic blood pressure higher than 140 mm Hg and they are often accompanied by proteinuria. These disorders affect about 10% of pregnant women around the world, with preeclampsia occurring in 2–8% of all pregnancies (Timokhina et al., 2019).

These disorders represent one of the leading causes of maternal and fetal morbidity and mortality, contributing to 7–8% of maternal death worldwide (Dolea & AbouZahr, 2003; Shah et al., 2009; McClure et al., 2009). The World Health Organization estimates that the incidence of preeclampsia is seven times higher in developing countries than in developed countries. However, the occurrence of these diseases appears under-reported in low and middle income countries, implying that the true incidence is unknown (Igberase & Ebeigbe, 2006; Malik et al., 2019). While there is evidence that hypertensive disorders of pregnancy are related with the development of cardiac dysfunctions both in the mother and in the child (Bellamy et al., 2007; Davis et al., 2012; Ambrožic et al., 2020; Garcia-Gonzalez et al., 2020; Aksu et al., 2021; deMartelly et al., 2021), there is no common agreement on the relation between the severity of hypertension and cardiac dysfunction (Tatapudi & Pasumarthy, 2017b) and echocardiography is not included in baseline evaluation of hypertensive disorders of pregnancy. Further investigations on these disorders are needed, especially for developing countries, where women often give birth at a younger age with respect to developed countries.
The goal of this chapter is to detect which cardiac function is altered and under which hypertensive disorders by relying on a principled Bayesian nonparametric approach. An interesting case-control study to explore the relation between cardiac dysfunction and hypertensive disorders is provided by Tatapudi & Pasumarthy (2017a), where the measures of ten different cardiac function indexes were recorded in four groups of pregnant women in India. Groups of women are characterized by different hypertensive disorder diagnoses, that are naturally ordered based on the severity of the diagnosed disorder: healthy (C), gestational hypertension (G), mild preeclampsia (M) and severe preeclampsia (S). Hypertensive diagnoses are used as identifiers for what we call populations of patients and we refer to cardiac function indexes also with the term response variables. For each response variable we want to determine a partition of the four populations of patients. This amounts to identifying similarities between different hypertensive disorders, with respect to each cardiac index. Supposing, for instance, that the selected partition assigns all the populations to the same cluster, one can conclude that no alteration is shown for the corresponding cardiac index across different hypertensive diseases.

4.2 Challenges, main idea and related works

Our goal of identifying a partition of the four patients' populations for each of the ten responses can be rephrased as a problem of multiple model selection: we want to select the most plausible partition for each cardiac index. Frequentist hypothesis testing does not allow to deal with more than two populations in a straightforward way, since pairwise comparisons may lead to conflicting conclusions. Conversely, a Bayesian approach yields the posterior distribution on the space of partitions, which can be used for simultaneous comparisons. Moreover, the presence of $M = 10$ jointly tested cardiac indexes requires to

perform model selection repeatedly ten times. Once again, a Bayesian approach seems to be preferred, because, as observed for instance by Scott & Berger (2006), it does not require the introduction of a penalty term for multiple comparison, thanks to the prior distribution build-in penalty.

Here we design a Bayesian nonparametric model, that is tailored to deal with both a collection of populations and the multivariate information of the response variables, while preserving the typical flexibility of nonparametric models and producing easily interpretable results. When applied to the dataset on transthoracic echocardiography results for a cohort of Indian pregnant women in Section 4.7, our model effectively identifies modified cardiac functions in hypertensive patients compared to healthy subjects and progressively increased alterations with the severity of the disorder, in addition to other more subtle findings.

The observed data $X_{i,j,m}$ represent the measurement of the m–th response variable (cardiac index) on the i–th individual (pregnant woman) in the j–th population (hypertensive disorder) and, as in standard univariate ANOVA models, they are assumed to be partially exchangeable across disorders. As explained in Section 1.4, this means that for every $m \in \{1, \dots, M\}$, the law of $((X_{i,1,m})_{i\geq 1}, \dots, (X_{i,J,m})_{i\geq 1})$ is invariant with respect to permutations within each sequence of random variables, namely for any positive integers $n_1, \dots, n_J$

$$((X_{i,1,m})_{i=1}^{n_1}, \dots, (X_{i,J,m})_{i=1}^{n_J}) \stackrel{\mathrm{d}}{=} ((X_{\sigma_1(i),1,m})_{i=1}^{n_1}, \dots, (X_{\sigma_J(i),J,m})_{i=1}^{n_J})$$

for all permutations σ_j of $(1, \dots, n_j)$, with $j = 1, \dots, J$. This is a natural generalization of exchangeability to tackle heterogeneous data and, by de Finetti's representation theorem (Theorem 1.10), it amounts to assuming the existence of a collection of (possibly dependent) random probability measures $\{\pi_{j,m} : \ j = 1, \dots, J \ \ m = 1, \dots, M\}$ such that

$$X_{i,j,m} \mid \pi_{j,m} \overset{\text{i.i.d.}}{\sim} \pi_{j,m} \qquad i = 1, \dots, n_j$$

Hence, for any two populations $j \neq j'$, homogeneity corresponds to $\pi_{j,m} = \pi_{j',m}$ (almost surely). However, a reliable assessment of this type of homogeneity is troublesome when having just few patients per diagnosis, as it happens in the mild preeclampsia subsample. Not rely on simplifying parametric assumptions, in fact, a small sub-sample size may not be sufficiently informative to infer equality of entire unknown distributions. To overcome this issue, without introducing parametric assumptions, we resort to an alternative weaker notion of homogeneity between populations j and j': we only require the conditional means of the two populations to (almost surely) coincide

$$\mathbb{E}(X_{i,j,m} \mid \pi_{j,m}) = \mathbb{E}(X_{i,j',m} \mid \pi_{j',m}). \tag{4.1}$$

According to this definition, the detection of heterogeneities in cardiac function indexes

amounts to inferring which cardiac indexes have means that differ across diagnoses, as it is done in standard parametric ANOVA models. Besides clustering populations according to (4.1), it is also of interest to cluster patients, both within and across different groups, once the effect of the specific hypertensive disorder is taken into account. This task may be achieved by assuming a model that decomposes the observations as

$$X_{i,j,m} = \theta_{j,m} + \varepsilon_{i,j,m} \qquad \varepsilon_{i,j,m}|(\xi_{i,j,m}, \sigma^2_{i,j,m}) \overset{\text{ind}}{\sim} \text{N}(\xi_{i,j,m}, \sigma^2_{i,j,m}) \tag{4.2}$$

and the $\xi_{i,j,m}$ have a symmetric distribution around the origin, in order to ensure $E(\xi_{i,j,m}) = 0$. In view of this decomposition, we will let $\theta_{j,m}$ govern the clustering of populations while the $(\xi_{i,j,m}, \sigma^2_{i,j,m})$'s determine the clustering of individuals, namely patients, after removing the effect of the specific hypertensive disorder. In order to pursue this, for each cardiac index m, we will specify a hierarchical process prior for $(\xi_{i,j,m}, \sigma^2_{i,j,m})$ that is suited to infer the clustering structure both within and across different hypertensive disorders for a specific cardiac index. In particular, we will deploy a novel instance of hierarchical Dirichlet process, introduced in Teh et al. (2006), that we name *symmetric*, to highlight its centering in 0.

Early examples of Bayesian nonparametric models for ANOVA can be found in Cifarelli & Regazzini (1978) and Muliere & Petrone (1978), while the first popular proposal is due to De Iorio et al. (2004), uses the dependent Dirichlet process (DDP) of (MacEachern, 2000) and is therefore termed ANOVA-DDP. This model is mainly tailored to estimate populations' probability distributions, while we draw inferences over clusters of populations' means and obtain estimates of the unknown distributions as a by-product. Moreover, the ANOVA-DDP of De Iorio et al. (2004) was not introduced as a model selection procedure. A popular Bayesian nonparametric model, that does cluster populations and can be used for model selection, is the nested Dirichlet process of Rodriguez et al. (2008). As shown in Camerlenghi et al. (2019a), such a prior is biased towards homogeneity, in the sense that even a single tie between populations j and j', namely $X_{i,j,m} = X_{i',j',m}$ for some i and i', entails $\pi_{j,m} = \pi_{j'm}$ (almost surely). In order to overcome such a drawback, a novel class of nested, and more flexible, priors has been proposed in Camerlenghi et al. (2019a). See also Soriano & Ma (2017) for related work. Interesting alternatives that extend the analysis to more than two populations can be found in Christensen & Ma (2020), Lijoi et al. (2020) and in Beraha et al. (2021). Another similar proposal is the one by Gutiérrez et al. (2019), whose model identifies differences over cases' distributions and the control group. These models imply that two populations belong to the same cluster if they share the entire distribution. However, as already mentioned, distribution-based clustering is not ideal when dealing with scenarios as the one of hypertensive dataset. Further evidence will be provided in Section 4.7.1, through simulation studies. In addition, note that all these contributions deal with only one response variable and would need to be suitably generalized to fit the setup of this chapter. As far as the contributions treating multiple response variables are concerned, uses of nonparametric priors for multiple testing can be found, for instance, in

Gopalan & Berry (1998), Do et al. (2005), Dahl & Newton (2007), Guindani et al. (2009), Martin & Tokdar (2012) and more recently in Cipolli et al. (2016), who propose an approximate finite Pólya tree multiple testing procedure to compare two-samples' locations, and in Denti et al. (2020). However, in all these contributions, models are developed directly over summaries of the original data (e.g. averages, z-scores) and, as such, do not allow to draw any inference on the entire distributions and clusters of subjects.

4.3 The Bayesian nonparametric model

The use of discrete nonparametric priors for Bayesian model-based clustering has become standard practice. The DP (Ferguson, 1973) is the most popular instance, and clustering is typically addressed by resorting to the mixture model described in Section 1.2.2, which with our data structure amounts to

$$X_{i,j,m}|\psi_{i,j,m} \overset{\text{ind}}{\sim} k(X_{i,j,m};\psi_{i,j,m}), \qquad \psi_{i,j,m}|\tilde{p}_{j,m} \overset{\text{ind}}{\sim} \tilde{p}_{j,m}$$

for $m = 1,\dots,M$, $j = 1,\dots,J$ and $i = 1,\dots,n_j$. Here $k(\,\cdot\,;\,\cdot\,)$ is some kernel and the $\tilde{p}_{j,m}$'s are discrete random probability measures. Hence, the $\psi_{i,j,m}$'s may exhibit ties. The model specification for $\tilde{p}_{j,m}$ will be tailored to address the following goals: (i) cluster the J probability distributions based on their means; (ii) cluster the observations $X_{i,j,m}$ according to the ties induced on the $\psi_{i,j,m}$'s by the $\tilde{p}_{j,m}$'s for a given fixed j and across different j's. These two issues will be targeted separately: we first design a clustering scheme for the populations, through the specification of a DP on the means of the $X_{i,j,m}$'s and, then, we cluster the data using a hierarchical DP having a specific invariance structure that ideally suited to the application at hand.

4.3.1 The prior on disease–specific locations

As a model for the observations we consider a nonparametric mixture of Gaussian distributions specified as

$$X_{i,j,m} \,|\, (\boldsymbol{\theta}_m, \boldsymbol{\xi}_m, \boldsymbol{\sigma}^2_m) \overset{\text{ind}}{\sim} \text{N}(\theta_{j,m} + \xi_{i,j,m}, \sigma^2_{i,j,m}) \tag{4.3}$$

where $\boldsymbol{\theta}_m = (\theta_{1,m},\dots,\theta_{J,m})$, $\boldsymbol{\xi}_m = (\xi_{1,1,m},\dots,\xi_{1,n_1,m},\xi_{2,1,m},\dots,\xi_{n_J,J,m})$, with a similar definition for the vector $\boldsymbol{\sigma}^2_m$, and $\text{N}(\,\mu,\,\sigma^2\,)$ denotes a normal distribution with mean μ and variance σ^2. The assumption in (4.3) clearly reflects (4.2). Moreover, in order to account for the two levels of clustering we are interested in, we will assume that

$$(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M) \sim P, \qquad (\xi_{i,j,m},\sigma^2_{i,j,m}) \,|\, \tilde{q}_{j,m} \overset{\text{i.i.d.}}{\sim} \tilde{q}_{j,m} \quad (i = 1,\dots,n_j) \tag{4.4}$$

where $\tilde{q}_{1,m}, \ldots, \tilde{q}_{J,m}$ are discrete random probability measures independent from $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$. Thus, the likelihood corresponds to

$$\prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{i=1}^{n_j} \frac{1}{\sigma_{i,j,m}} \varphi\Big(\frac{x_{i,j,m} - \theta_{j,m} - \xi_{i,j,m}}{\sigma_{i,j,m}}\Big) \tilde{q}_{j,m}(\mathrm{d}\xi_{i,j,m}, \mathrm{d}\sigma_{i,j,m}) \tag{4.5}$$

with φ denoting the standard Gaussian density. Relevant inferences can be carried out if one is able to marginalize this expression with respect to both $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ and $(\tilde{q}_{1,m}, \ldots, \tilde{q}_{J,m})$ for each $m = 1, \ldots, M$.

This specification allows to address the model selection problem in the following way. If $\mathcal{M}^m$ stands for the space of all partitions of the J populations for the m-th cardiac function index, then $\mathcal{M}^m = \{M_b^m : b = 1, \ldots, \text{card}(\mathscr{P}_J)\}$ where $\mathscr{P}_J$ is the collection of all possible partitions of $[J] = \{1, \ldots, J\}$. In our specific case, $J = 4$ and $\text{card}(\mathscr{P}_J) = 15$, thus we have 15 competing models per cardiac index. Each competing model corresponds to a specific partition in $\mathcal{M}^m$. In particular, the partition arises from ties between the population specific means in $\boldsymbol{\theta}_m$ and, hence, the distribution P in (4.4) needs to associate positive probabilities to ties between the parameters within the vector $\boldsymbol{\theta}_m$, for each $m = 1, \ldots, M$.

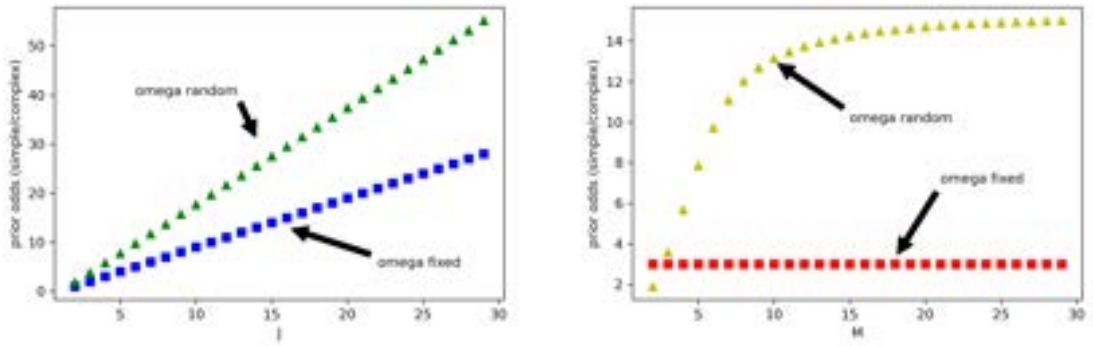
Let us start considering as distribution P a well-known effective clustering prior, i.e. a mixture of DPs in the spirit of Antoniak (1974), namely

$$\begin{aligned} \theta_{j,m} \mid \tilde{p}_m &\overset{\text{i.i.d.}}{\sim} \tilde{p}_m && j = 1, \ldots, J \\ \tilde{p}_m \mid \omega &\overset{\text{i.i.d.}}{\sim} \text{DP}(\omega, G_m) && m = 1, \ldots, M \\ \omega &\sim p_\omega \end{aligned} \tag{4.6}$$

where $\text{DP}(\omega, G_m)$ denotes the DP with concentration parameter ω and non-atomic baseline probability measure G_m and p_ω is a probability measure on $\mathbb{R}^+$. The discreteness of the DP implies the presence (with positive probability) of ties within the vector of locations $\boldsymbol{\theta}_m$ associated to a certain cardiac index m, as desired. The ties give rise to a random partition: as shown in Antoniak (1974), the probability of observing a specific partition of the elements in $\boldsymbol{\theta}_m$ consisting of $k \leq J$ distinct values with respective frequencies $n_1, \ldots, n_k$ coincides with

$$\Pi_k^{(J)}(n_1, \ldots, n_k) = \frac{\omega^k}{(\omega)_J} \prod_{i=1}^{k} (n_i - 1)! \tag{4.7}$$

where $(\omega)_J = \Gamma(\omega + J)/\Gamma(\omega)$. The use of a shared concentration parameter over (4.7) to address multiple model selection has been also explored in Moser et al. (2021), where they cluster parameters in a probit model. When there is no prior information available on competing partitions, the advantages of using (4.7) as prior for model selection are that, firstly, it induces borrowing of strength across diagnoses and, secondly, being ω random, it generates borrowing of information across cardiac indexes, improving the Bayesian learning mechanism. Moreover, these two features can also be interpreted from a frequentist point



(a) (b)

Figure 4.1: Changes in the prior odds for varying J and M, for fixed ω and random ω. The simple model is the one corresponding to no modification of cardiac indexes under any diagnosis, the complex is the one of one modification in corresponding of one diagnosis for all cardiac indexes and the prior used for ω is a Gamma prior.

of view in terms of desirable penalties for the model selection problem. In fact, for any fixed $\omega \in [0, 1]$, J and M, the DP penalizes more complex models since, the prior probability of a model in $\mathcal{M}^m$ decreases while the associated k in (4.7) increases, inducing a so-called Ockham's-razor penalty. At the same time, the procedure penalizes for the multiplicity of the model selections that are performed. This second type of penalty has to be intended in the following way: while J and/or M increase, the prior odds change in favor of less complex models (see Figure 4.1). For more details on this, see Scott & Berger (2010). Summing up, the mixture of DPs automatically induces a prior distribution on $\{\mathcal{M}^m : m = 1, \ldots, M\}$, provided by (4.7) and the prior on ω, and it presents desirable properties for model selection purposes that can be interpreted either in terms of borrowing of information or as penalties.

However, in the analysis of hypertensive disorders, some prior information on competing models is available, and this is not yet incorporated in (4.7). In fact, as already mentioned, there is a natural order of the diagnoses, which is given by the severity of the disorders, i.e. C, G, M, S. Partitions that do not respect this order, e.g. $\{\{C, S\}\{G\}, \{M\}\}$, should reasonably be excluded from the support of the prior. Thus, we consider a prior over $\mathcal{M}^m$ that associates zero probability to partitions that do not respect the natural order of the diagnoses and a probability proportional to that in (4.7) to the remaining partitions, i.e.

$$\mathbb{P}(M_b^m \mid \omega) \propto \begin{cases} \Pi_k^{(J)}(n_1, \ldots, n_k) & \text{if } M_b^m \text{ is compatible with the natural order} \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

This amounts to a distribution P for $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ given by

$$\begin{aligned} (\theta_{1,m}, \ldots, \theta_{J,m}) \mid \omega &\overset{\text{ind}}{\sim} P_{\omega, G_m} \qquad m = 1, \ldots, M \\ \omega &\sim p_\omega \end{aligned} \tag{4.9}$$

where P_{ω,G_m} is the distribution obtained sampling a partition accordingly to (4.8) and associating to each cluster a unique value sampled from G_m. Using (4.9) as prior for the disease-specific locations, we preserve the desirable properties of the mixture of DPs mentioned before, while incorporating prior information on the severity of the diseases.
As detailed in the next section, we further consider random probability measures $\tilde{q}_{j,m}$ that satisfy the symmetry condition

$$\tilde{q}_{j,m}(A \times B) = \tilde{q}_{j,m}((-A) \times B) \qquad \text{a.s.} \tag{4.10}$$

for any A and B. This condition is crucial as it ensures that the parameters $\theta_{j,m}$, for $j = 1, \ldots, J$ and $m = 1, \ldots, M$, in (4.3) are identified, namely $\mathbb{E}(X_{i,j,m} \mid \boldsymbol{\theta}_m, \tilde{q}_{j,m}) = \theta_{j,m}$ with probability one. This identifiability property is crucial to make inference over the location parameters $\boldsymbol{\theta}_m$'s. Similar model specifications for discrete exchangeable data have been proposed and studied in Dalal (1979b), Doss (1984), Diaconis & Freedman (1986) and Ghosal et al. (1999), of which (4.5) represents a generalization to density functions and partially exchangeable data.

4.3.2 The prior for the error terms

While the clustering of populations is governed by (4.8), we use a mixture of hierarchical discrete processes for the error terms. This has the advantage of modeling the clustering of the observations, both within and across different samples, once one has taken into account the disease-specific effects. This clustering structure allows to model heterogeneity across patients in a much more realistic way with respect to standard ANOVA models based on assumption of normality. Cardiac indexes may be influenced by a number of factors that are not directly observed in the study, such as pre-existing conditions (Hall et al., 2011) and psychosocial factors (Pedersen et al., 2017). These unobserved relevant factors may be shared across patients with the same or a different diagnosis and may also result in outliers. To take into account this latent heterogeneity of the data, we introduce the hierarchical symmetric DP that satisfies the symmetry condition in (4.10) and, moreover, allows to model heterogeneous data similarly to the hugely popular hierarchical DP (Teh et al., 2006).
The basic building block of the proposed prior is the invariant Dirichlet process, which was introduced for a single population ($J = 1$) in an exchangeable framework by Dalal (1979a). Such a modification of the DP satisfies a symmetry condition, in the sense that it is a random probability measure that is invariant with respect to a chosen group of trans-

formations $\mathcal{G}$. A more formal definition and detailed description of the invariant DP can be found in Section 1.2.3. For our purposes it is enough to consider the specific case given by the symmetric Dirichlet process, that can be constructed through a symmetrization of a Dirichlet process. Consider a non-atomic probability measure P_0 on $\mathbb{R}$ and assume, without loss of generality, that P_0 is symmetric about 0. Let $\tilde{Q}_0 \sim \text{DP}(\alpha, P_0)$. If

$$\tilde{Q}(A) = \frac{\tilde{Q}_0(A) + \tilde{Q}_0(-A)}{2} \qquad \forall A \in \mathcal{B}(\mathbb{R}) \tag{4.11}$$

where $-A = \{x \in \mathbb{R} : -x \in A\}$, then $\tilde{Q}$ is symmetric about 0 (almost surely) and termed symmetric DP, in symbols $\tilde{Q} \sim \text{s-DP}(\alpha\,,\, P_0)$. Notice that P_0 does not need to be symmetric, however requiring P_0 to be symmetric is without lost of generality and with the advantage of P_0 being an interpretable parameter of the prior: P_0, if symmetric, is the expected value of $\tilde{Q}$. The random probability measure $\tilde{Q}$ is the basic building block of the hierarchical process that we use to model the heterogeneity of the error terms across different populations, $j = 1, \ldots, J$, in such a way that clusters identified by the unique values can be shared within and across populations. Such a prior is termed *symmetric hierarchical Dirichlet process* (s-HDP) and is described as

$$\begin{aligned} \tilde{q}_{j,m} \mid \gamma_{j,m},\, \tilde{q}_{0,m} &\overset{\text{ind}}{\sim} \text{s-DP}(\gamma_{j,m}, \tilde{q}_{0,m}) \\ \tilde{q}_{0,m} \mid \alpha_m &\overset{\text{ind}}{\sim} \text{s-DP}(\alpha_m, P_{0,m}) \end{aligned} \tag{4.12}$$

where $\gamma_{j,m}$ and α_m are positive parameters and $P_{0,m}$ is a non-atomic probability distribution symmetric about 0. We use the notation $(\tilde{q}_{1,m}, \ldots, \tilde{q}_{J,m}) \sim \text{s-HDP}(\boldsymbol{\gamma}_m, \alpha_m, P_{0,m})$, where $\boldsymbol{\gamma}_m = (\gamma_{1,m}, \ldots, \gamma_{j,m})$. This definition clearly ensures the validity of (4.10). A graphical model representation of the over-all proposed model is displayed in Figure 4.2.

Still referring to the decomposition of the observations into disease-specific locations and an error term, i.e. $X_{i,j,m} = \theta_{j,m} + \varepsilon_{i,j,m}$, it turns out that the $\varepsilon_{i,j,m}$'s are from a symmetric hierarchical DP mixture (s-HDP mixture) with a normal kernel. Hence, the patient's clusters are identified through the $\varepsilon_{i,j,m}$, which, according to (4.3), are conditionally independent from a $\text{N}(\xi_{i,j,m}, \sigma^2_{i,j,m})$ given $(\xi_{i,j,m}, \sigma^2_{i,j,m})$. The choice of the specific invariant DP is aimed at ensuring that $\mathbb{E}(\varepsilon_{i,j,m}|\tilde{q}_{j,m}) = 0$. The clusters identified by the s-HDP mixture can be interpreted as representing common unobserved factors across patients, once the disease-specific locations have been accounted for. Indeed, for any pair of patients, we may consider the decomposition $X_{i,j,m} - X_{i',j',m} = \Delta^{(m)}_\theta + \Delta^{(m)}_\xi + (e_{i,j,m} - e_{i',j',m})$ where $\Delta^{(m)}_\theta = \theta_{j,m} - \theta_{j',m}$, $\Delta^{(m)}_\xi = \xi_{i,j,m} - \xi_{i',j',m}$ and $e_{i,j,m}$ and $e_{i',j',m}$ are independent and normally distributed random variables with zero mean and variances $\sigma^2_{i,j,m}$ and $\sigma^2_{i',j',m}$, respectively.

Hence, patients' clustering reflects the residual heterogeneity that is not captured by the disease-specific component $\Delta^{(m)}_\theta$ and are related to the subject-specific locations $\Delta^{(m)}_\xi$ and to the zero-mean error component $(e_{i,j,m} - e_{i',j',m})$. In view of this interpretation, using a s-

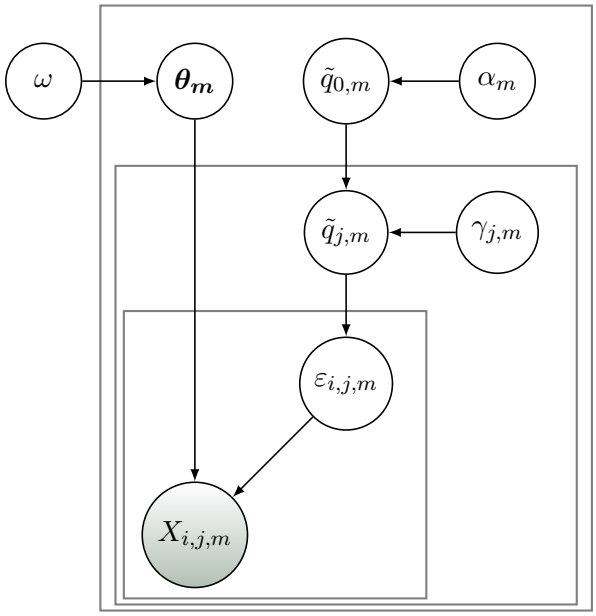


Figure 4.2: Graphical representation of the model. Each node represents a random variable and each rectangle denotes conditional i.i.d. replications of the model within the rectangle.

HDP mixture over error terms offers a three-fold advantage. Firstly, the presence of clearly separated clusters of patients within and across populations will indicate the presence of unobserved relevant factors which affect the cardiac response variables. Secondly, single patients with very low probabilities of co-clustering with all other subjects will have to be interpreted as outliers. Finally, the estimated clustering structure can also be used to check whether the relative effect of a certain disease (with respect to another) is fully explained by the corresponding $\Delta_\theta^{(m)}$. To clarify this last point consider two diseases: if the posterior co-clustering probabilities among patients sharing the same disease are different between the two populations, this will indicate that different diagnoses not only have an influence on disease-specific locations (which is measured by $\Delta_\theta^{(m)}$), but they also have an impact on the shape of the distribution of the corresponding cardiac index. More details on this can be found in Section 4.7.1.

4.4 Marginal distributions and random partitions

As emphasized in the previous sections, ties among the $\theta_{j,m}$'s and the $(\xi_{i,j,m}, \sigma^2_{i,j,m})$'s are relevant for inferring the clustering structure both among the populations (hypertensive diseases) and among the individual units (patients). Indeed, for each m (cardiac index) they induce a random partition that emerges as a composition of two partitions generated respectively by the prior in (4.9) and the s-HDP. The laws of these random partitions are not only crucial to understand the clustering mechanism, but also necessary in order to derive

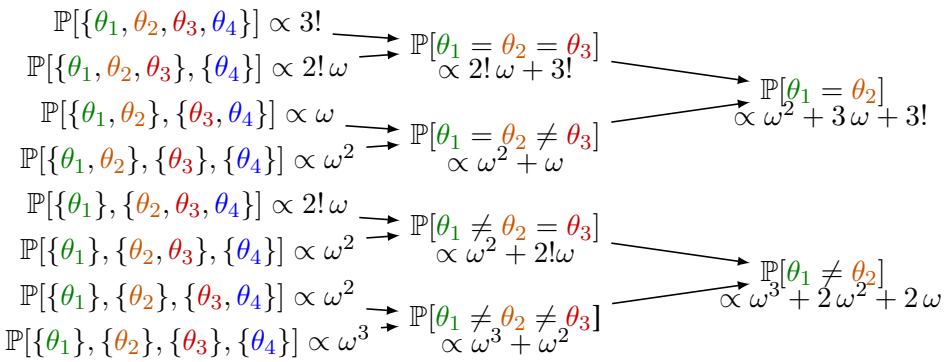


Figure 4.3: A priori partitions' probabilities joint (on the left) and marginals

posterior sampling schemes. In this section the law of the partitions are derived and used to compute the predictive distributions that, jointly with the likelihood, determine the full conditionals of the Gibbs sampler in the next section. To reduce the notational burden, in this and the following section, we remove the dependence of observations and parameters on the specific response variable m and denote with $\phi_{i,j}$ the pair $(\xi_{i,j}, \sigma^2_{i,j})$ and with $\boldsymbol{\phi}$ the collection $(\phi_{1,1}, \ldots, \phi_{1,n_1}, \phi_{2,1}, \ldots \phi_{n_J,J})$.

Conditionally on ω, the law of the partition in (4.8) leads to the following predictive distribution for the disease-specific locations

$$\theta_j \,|\, \omega, \theta_1, \ldots, \theta_{j-1} \;\sim\; a_j(\omega, \theta_1, \ldots, \theta_{j-1})\, \delta_{\theta_{j-1}} + (1 - a_j(\omega, \theta_1, \ldots, \theta_{j-1}))G$$

where

$$a_j(\omega, \theta_1, \ldots, \theta_{j-1}) = \frac{\sum_{(*_j)} \Pi_k^{(J)}(n_1, \ldots, n_k)}{\sum_{(*)} \Pi_k^{(J)}(n_1, \ldots, n_k)} \tag{4.13}$$

where the sum at the denominator runs over the set of partitions compatible to the ties in $(\theta_1, \ldots, \theta_{j-1})$ and the one at the numerator runs over a subset of the same partitions where also $\theta_j = \theta_{j-1}$.

We recall that the prior over partitions is provided by

$$\mathbb{P}(M_b^m \mid \omega) \propto \begin{cases} \omega^{k-1} \prod_{i=1}^k (n_i - 1)! & \text{if } M_b^m \text{ is compatible with the natural order} \\ 0 & \text{otherwise} \end{cases}$$

where k is the number of distinct clusters accordingly to the partition M_b^m and $n_1, \ldots, n_k$ are the clusters' respective frequencies. Thus, being $J = 4$, one obtains the probabilities in Figure 4.3, starting from which it is possible to compute the joint distribution for $(\theta_{1,m}, \ldots, \theta_{4,m})$ conditional on ω

$$\theta_{1,m} \mid \omega \sim G_m$$

$$\theta_{2,m} \mid \theta_{1,m}, \omega \sim \frac{\omega^2+3\omega+6}{(\omega+2)(\omega^2+\omega+3)}\delta_{\theta_{1,m}} + \frac{\omega^3+2\omega^2+2\omega}{(\omega+2)(\omega^2+\omega+3)}G_m$$

$$\theta_{3,m} \mid \theta_{1,m}, \theta_{2,m}, \omega \sim \begin{cases} \frac{2\omega+6}{\omega^2+3\omega+6}\delta_{\theta_{2,m}} + \frac{\omega^2+\omega}{\omega^2+3\omega+6}G_m & \text{if } \theta_{1,m} = \theta_{2,m} \\ \frac{\omega+2}{\omega^2+2\omega+2}\delta_{\theta_{2,m}} + \frac{\omega^2+\omega}{\omega^2+2\omega+2}G_m & \text{if } \theta_{1,m} \neq \theta_{2,m} \end{cases}$$

$$\theta_{4,m} \mid \theta_{1,m}, \theta_{2,m}, \theta_{3,m}, \omega \begin{cases} \frac{3}{\omega+3}\delta_{\theta_{3,m}} + \frac{\omega}{\omega+3}G_m & \text{if } \theta_{1,m} = \theta_{2,m} = \theta_{3,m} \\ \frac{2}{\omega+2}\delta_{\theta_{3,m}} + \frac{\omega}{\omega+2}G_m & \text{if } \theta_{1,m} \neq \theta_{2,m} = \theta_{3,m} \\ \frac{1}{\omega+1}\delta_{\theta_{3,m}} + \frac{\omega}{\omega+1}G_m & \text{otherwise} \end{cases}$$

Moving to second-level partitions induced by the s-HDP, we recall that the key concept for studying random partitions on multi-sample data is the *partially exchangeable partition probability function* (pEPPF). See, e.g., Lijoi et al. (2014a) and Camerlenghi et al. (2019b). The pEPPF returns the probability of a specific multi-sample partition and represents the appropriate generalization of the well-known single-sample EPPF, which in the DP case corresponds to (4.7). Discreteness of the s-HDP $(\tilde{q}_1, \ldots, \tilde{q}_m)$ in (4.12) induces a partition of the elements of $\boldsymbol{\phi}$ into equivalence classes identified by the distinct values. Taking into account the underlying partially exchangeable structure, such a random partition is characterized by the pEPPF

$$\tilde{\Pi}_k^{(N)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J) = \mathbb{E}\left(\int_{\Phi^k} \prod_{j=1}^{J}\prod_{h=1}^{k} \tilde{q}_{j,m}^{n_{j,h}}(\mathrm{d}\phi_i)\right) \tag{4.14}$$

where $\boldsymbol{n}_j = (n_{j,1}, \ldots, n_{j,k})$ are non-negative integers, for any $j = 1, \ldots, J$, such that $n_{j,h}$ is the number of elements in $\boldsymbol{\phi}$ corresponding to population j and belonging to cluster h. Thus $\sum_{j=1}^{J} n_{j,h} \geq 1$ for any $h = 1, \ldots, k$, $\sum_{h=1}^{k} n_{j,h} = n_j$ and $\sum_{h=1}^{k}\sum_{j=1}^{J} n_{j,h} = N$. The determination of probability distributions of this type is challenging and only recently the first explicit instances have appeared in the literature. See e.g., Lijoi et al. (2014a), Camerlenghi et al. (2019a) and Camerlenghi et al. (2019b). With respect to the hierarchical case considered in Camerlenghi et al. (2019b), the main difference is that here we have to take into account the specific structure (4.11) of the $\tilde{q}_{j,m}$. The almost sure symmetry of the process generates a natural random matching between sets in the induced partition. Therefore, instead of studying the marginal law in (4.14), we derive the joint law of the partition and of the random matching. Formally, consider a partition $\{A_1^+, A_1^-, \ldots, A_k^+, A_k^-\}$ of $\boldsymbol{\phi}$, such that, for $h = 1, \ldots, k$, all the elements in A_h^+ belong to $\mathbb{R}^+ \times \mathbb{R}^+$, all the elements in A_h^- belong to $\mathbb{R}^- \times \mathbb{R}^+$ and, if $\phi_{i,j} \in A_h^+$ and $\phi_{i',j'} \in A_h^-$, then the element-wise absolute

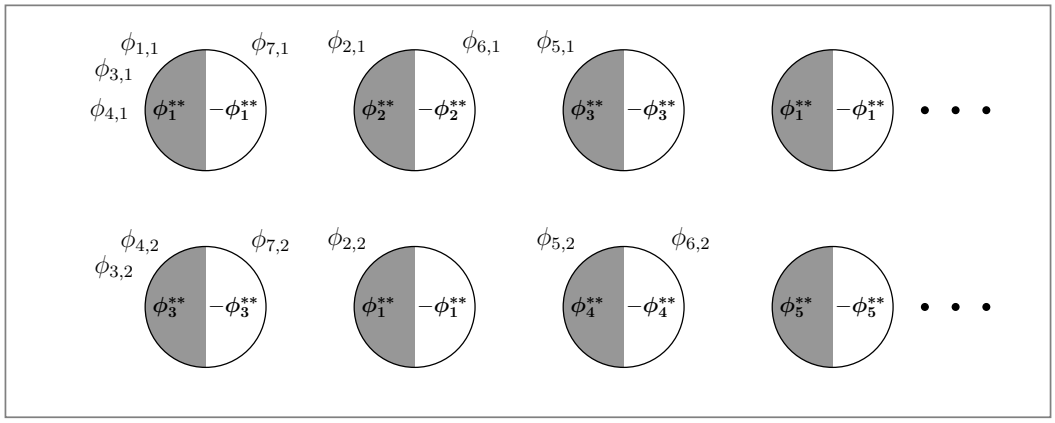


Figure 4.4: Chinese restaurant franchise representation of the symmetric hierarchical DP for $J = 2$ populations. Each circle represents a table.

values of $\phi_{i,j}$ and $\phi_{i',j'}$ are equal. Denote with $n^+_{j,h}$ the number of elements in $A^+_h \cap \{\phi_{i,j}, i = 1, \ldots, n_j\}$ and with $n^-_{j,h}$ the number of elements in $A^-_h \cap \{\phi_{i,j}, i = 1, \ldots, n_j\}$. The probability of observing $\{A^+_1, A^-_1, \ldots, A^+_k, A^-_k\}$ is

$$\tilde{\tilde{\Pi}}^{(N)}_k(\boldsymbol{n}^+_{\mathbf{1}}, \boldsymbol{n}^-_{\mathbf{1}}, \ldots, \boldsymbol{n}^+_{\boldsymbol{J}}, \boldsymbol{n}^-_{\boldsymbol{J}}) = \mathbb{E}\left(\int_{\Phi^k} \prod_{j=1}^{J} \prod_{h=1}^{k} \tilde{q}_{j,m}{}^{n^+_{j,h}+n^-_{j,h}}(d\phi)\right) \tag{4.15}$$

with $\boldsymbol{n}^+_{\boldsymbol{j}} = (n^+_{j,1}, \ldots, n^+_{j,k})$. As for the determination of (4.15), a more intuitive understanding may be gained if one considers its corresponding Chinese restaurant franchise (CRF) metaphor, which displays a variation of both the standard Chinese restaurant franchise of Teh et al. (2006) and the skewed Chinese restaurant process of Iglesias, Orellana & Quintana (2009). Figure 4.4 provides a graphical representation. The scheme is as follows: there are J restaurants sharing the same menu and the customers are identified by their choice of $\phi_{i,j}$ but, unlike in the usual CRF, at each table two *symmetric dishes* are served. Denote with $\phi^*_{t,j} = (\xi^*_{t,j}, \sigma^{2*}_{t,j})$ and $-\phi^*_{t,j} = (-\xi^*_{t,j}, \sigma^{2*}_{t,j})$ the two dishes served at table t in restaurant j, with $\phi^{**}_h = (\xi^{**}_h, \sigma^{**2}_h)$ and $-\phi^{**}_h = (-\xi_h, \sigma^{**2}_h)$ the h-th pair of dishes in the menu, with $n^+_{j,h}$ the number of customers in restaurant j eating dish ϕ^{**}_h, and with $n^-_{j,h}$ the number of customers in restaurant j eating dish $-\phi^{**}_h$. This means that two options are available to a customer entering restaurant j: she/he will either sit at an already occupied table, with probability proportional to the number of customers at that table or will sit at a new table with probability proportional to the concentration parameter γ_j. In the former case, the customer will choose the dish $\phi^*_{t,j}$ with probability $1/2$ and $-\phi^*_{t,j}$ otherwise. In the latter case, the customer will eat a dish served at another table of the franchise with probability proportional to half the number of tables that serve that dish, or will make a new order with probability proportional to the concentration parameter α. In view of this scheme, the

probability in (4.15) turns out to be

$$\tilde{\bar{\Pi}}_k^{(N)}(\boldsymbol{n}_1^+, \dots, \boldsymbol{n}_J^-) = 2^{-N} \bar{\Pi}_k^{(N)}(\boldsymbol{n}_1^+ + \boldsymbol{n}_1^-, \dots, \boldsymbol{n}_J^+ + \boldsymbol{n}_J^-)$$

and $\bar{\Pi}_k^{(N)}$ on the right-hand-side is the pEPPF of the hierarchical DP derived in Camerlenghi et al. (2019b), namely

$$\bar{\Pi}_k^{(N)}(\boldsymbol{n}_1, \dots, \boldsymbol{n}_k) = \left(\prod_{j=1}^{J} \frac{\prod_{i=1}^{k} (\gamma_j)_{n_{j,h}}}{(\gamma_j)_{n_j}} \right) \sum_{\boldsymbol{\ell}} \frac{\alpha^k}{(\alpha)_{|\boldsymbol{\ell}|}} \prod_{h=1}^{k} (\ell_{\bullet,h} - 1)! \prod_{j=1}^{J} P(K_{n_{j,h}} = \ell_{j,h})$$

where each sums runs over all $\ell_{j,h}$ in $\{1, \dots, n_{j,h}\}$, if $n_{j,h} \geq 1$, and equals 1 if $n_{j,h} = 0$, whereas $\ell_{\bullet,h} = \sum_{j=1}^{J} \ell_{j,h}$ and $|\boldsymbol{\ell}| = \sum_{j=1}^{J} \sum_{h=1}^{k} \ell_{j,h}$. Note that the latent variable $\ell_{j,h}$ is the number of tables in restaurant j serving the h-th pair of dishes. Moreover, $K_{n_{j,h}}$ is a random variable denoting the number of distinct clusters, out of $n_{j,h}$ observations generated by a DP with parameter γ_j and diffuse baseline P_0 and it is well-known that

$$\mathbb{P}(K_{n_{j,h}} = \ell_{j,h}) = \frac{\gamma_j^{\ell_{j,h}}}{(\gamma_j)_{n_{j,h}}} |\mathfrak{s}(n_{j,h}, \ell_{j,h})|$$

where $|\mathfrak{s}(n_{j,h}, \ell_{j,h})|$ is the signless Stilring number of the first kind. In view of this, one can deduce the predictive distribution

$$\begin{aligned} \mathbb{P}(\phi_{n_j+1,j} \in \cdot \,|\, \boldsymbol{\phi}) = {} & \frac{\gamma_j}{i - 1 + \gamma_j} \sum_{\boldsymbol{\ell}} \frac{\alpha}{|\boldsymbol{\ell}| + \alpha} \pi(\boldsymbol{\ell} \,|\, \boldsymbol{\phi}) P_0(\cdot) \\ & + \sum_{h=1}^{k} \left[\frac{n_{j,h}^+ + n_{j,h}^-}{n_j + \gamma_j} + \frac{\gamma_j}{n_j + \gamma_j} \sum_{\boldsymbol{\ell}} \frac{\ell_{\bullet,h}}{|\boldsymbol{\ell}| + \alpha} \pi(\boldsymbol{\ell} \,|\, \boldsymbol{\phi}) \right] \left(\frac{\delta_{\phi_h^{**}}(\cdot) + \delta_{-\phi_h^{**}}(\cdot)}{2} \right) \end{aligned}$$

where $\pi(\boldsymbol{\ell} \,|\, \boldsymbol{\phi})$ is the posterior distribution of the latent variables $\ell_{j,h}$'s and reads

$$\pi(\boldsymbol{\ell} \,|\, \boldsymbol{\phi}) \propto \frac{\alpha^k}{(\alpha)_{|\boldsymbol{\ell}|}} \prod_{h=1}^{k} (\ell_{\bullet,h} - 1)! \prod_{j=1}^{J} \frac{\gamma_j^{\ell_{j,h}}}{(\gamma_j)_{n_{j,h}^+ + n_{j,h}^-}} |\mathfrak{s}(n_{j,h}^+ + n_{j,h}^-, \ell_{j,h})| \mathbb{1}_{\{1, \dots, n_{j,h}^+ + n_{j,h}^-\}}(\ell_{j,h})$$

where $\mathbb{1}$ is the indicator function.

4.5 Posterior inference

The findings of the previous section are the key ingredients to perform posterior inference with a marginal Gibbs sampling scheme. The output of the sampler is structured into three levels: the first produces posterior probabilities on partitions of disease-specific locations; the second generates density estimates; the third provides clusters of patients. For notational simplicity we keep omitting the dependence on m, except when the sampling of the

concentration ω is concerned. Recall that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$ and $\boldsymbol{\phi} = \{(\phi_{1,j}, \ldots, \phi_{n_j,j}) : j = 1, \ldots, J\}$, with $\phi_{i,j} = (\xi_{i,j}, \sigma^2_{i,j})$. The target distribution of the sampler is the joint distribution of $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and ω conditionally on the observed data $\boldsymbol{X}$.

Sampling $\boldsymbol{\phi}$. In view of the CRF representation of the s-HDP, $t_{i,j}$ stands for the label of the table where the i-th customer in restaurant j sits and $h_{t,j}$ for the dish label served at table t in restaurant j and with $\boldsymbol{t}$ and $\boldsymbol{h}$ the corresponding vectors for varying i, j and t. Moreover, define the assignment variable $s_{i,j} = \mathbb{1}(\phi_{i,j} = \phi^*_{t_{i,j},j}) - \mathbb{1}(\phi_{i,j} = -\phi^*_{t_{i,j},j})$ and $\boldsymbol{s}$ is the corresponding vector. In order to generate $\boldsymbol{\phi}$, we need to sample

(i) $(t_{i,j}, s_{i,j})$ for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$;

(ii) $h_{t,j}$ for $t \in \boldsymbol{t}$ and $j = 1, \ldots, J$;

(ii) ϕ^{**}_h for $h \in \boldsymbol{h}$.

Note that, using the latent allocation indicators in $\boldsymbol{t}$ and $\boldsymbol{h}$, the sampling scheme is more efficient than sampling directly from the full conditional of each $\phi_{i,j}$, since the algorithm can change more than one parameter simultaneously (Neal, 2000). Define $\varepsilon_{i,j} = X_{i,j} - \theta_j$ and denote the conditional normal density of $\varepsilon_{i,j}$ associated with the parameter $\phi^* = (\xi^*, \sigma^{2*})$ with $h(\varepsilon_{i,j,m}|\phi^*)$ and the marginal density of $\varepsilon_{i,j}$ as

$$\bar{h}(\varepsilon_{i,j}) = \int h(\varepsilon_{i,j}|\phi) P_0(d\phi)$$

To sample $(t_{i,j}, s_{i,j})$ from their joint full conditional, we first sample $t_{i,j}$ from

$$P(t_{i,j} = t \mid \boldsymbol{t}^{-(i,j)}, \boldsymbol{h}^{-(i,j)}, \boldsymbol{\phi}^{*-(i,j)}, \boldsymbol{\phi}^{**-(i,j)}, \varepsilon_{i,j}) \propto \begin{cases} n^{-(i,j)}_{t,j}\, p_{\text{old}}(\varepsilon_{i,j}|\phi^*_{t,j}) & \text{if } t \in \boldsymbol{t}^{-(i,j)} \\ \gamma_j\, p_{\text{new}}(\varepsilon_{i,j}|\boldsymbol{\phi}^{**-(i,j)}) & \text{if } t = t^{\text{new}} \end{cases}$$

where $\boldsymbol{t}^{-(i,j)}$, $\boldsymbol{h}^{-(i,j)}$ $\boldsymbol{\phi}^{*-(i,j)}$,$\boldsymbol{\phi}^{**-(i,j)}$ coincide with the vectors $\boldsymbol{t}$, $\boldsymbol{h}$ $\boldsymbol{\phi}^*$,$\boldsymbol{\phi}^{**}$ after having removed the entries corresponding to the i-th customer in restaurant j. Moreover

$$p_{\text{old}}(\varepsilon_{i,j}|\phi^*_{t,j}) = \frac{1}{2} h(\varepsilon_{i,j}|\phi^*_{t,j}) + \frac{1}{2} h(\varepsilon_{i,j}| - \phi^*_{t,j})$$

and

$$p_{\text{new}}(\varepsilon_{i,j}|\boldsymbol{\phi}^{**-(i,j)}) = \sum_{h=1}^{k^{-(i,j)}} \frac{\ell_{\bullet,h}}{|\boldsymbol{\ell}| + \alpha} \left\{ \frac{1}{2} h(\varepsilon_{i,j}|\phi^{**}_h) + \frac{1}{2} h(\varepsilon_{i,j}| - \phi^{**}_h) \right\} + \frac{\alpha}{|\boldsymbol{\ell}| + \alpha} \bar{h}(\varepsilon_{i,j})$$

Then we sample $s_{i,j}$ from its full conditional

$$p(s_{i,j} = s \mid \boldsymbol{\phi}^*, t_{i,j}, \epsilon_{i,j}) \propto \begin{cases} h(\varepsilon_{i,j}|\phi^*_{t_{i,j}}) & \text{if } s = 1 \\ h(\varepsilon_{i,j}| - \phi^*_{t_{i,j}}) & \text{if } s = -1 \end{cases}$$

The conditional distribution of $h_{t,j}$ is

$$p(h_{t,j} = h \mid \boldsymbol{t}, \boldsymbol{h}^{-(t,j)}, \boldsymbol{\phi}^{**-(t,j)}, \boldsymbol{s}, \boldsymbol{\varepsilon}) \propto \begin{cases} \ell_{\bullet,h}^{-(t,j)} \displaystyle\prod_{\{(i,j):\, t_{i,j}=t\}} h(s_{i,j}\, \varepsilon_{i,j} | \phi_h) & \text{if } h \in \boldsymbol{h}^{-(t,j)} \\ \alpha \displaystyle\int \prod_{\{(i,j):\, t_{i,j}=t\}} h(s_{i,j}\, \varepsilon_{i,j} | \phi) P_0(d\phi) & \text{if } h = h^{new} \end{cases}$$

Finally, when P_0 is conjugate with respect to the Gaussian kernel, the full conditional distribution of ϕ_h^{**} is obtained in closed form as posterior distribution of a Gaussian model, using as observations the collection $\{\, (s_{i,j}\, \epsilon_{i,j}) :\ h_{t_{i,j},j} = h\}$.

Sampling $\boldsymbol{\theta}$. For sampling the disease-specific location parameters, one can rely on the classical Chinese restaurant metaphor corrected for taking into account only the partitions that have positive prior probability. Thus, in order to generate $\boldsymbol{\theta}$, we first sample the labels $\boldsymbol{t}_\theta = \{t_1, \ldots, t_J\}$, where t_j is the label of the table where the j-th customer sits. Then, we sample the dish θ_t^* associated to table t for all $t \in \boldsymbol{t}_\theta$. If $z_{i,j} = X_{i,j} - \xi_{i,j}$, the conditional density of $\boldsymbol{z}_j = (z_{1,j}, \ldots, z_{n_j,j})$ associated to the location parameter θ^*, given $\boldsymbol{\sigma}_j = (\sigma_{1,j}, \ldots, \sigma_{n_j,j})$, is

$$f_{\theta^*}(\boldsymbol{z}_j | \boldsymbol{\sigma}_j) = \frac{1}{\sqrt{2\pi} \displaystyle\prod_{i=1}^{n_j} \sigma_{i,j}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n_j} \frac{(z_{i,j} - \theta^*)^2}{\sigma_{i,j}^2} \right\}$$

Under the prior in (4.9), the full conditional distribution of $\boldsymbol{t}_\theta$ is provided by

$$\begin{aligned} &p(t_j = t \mid t_1, \ldots, t_{j-1}, \theta_{j-1}, \boldsymbol{z}_j, \boldsymbol{\sigma}_j) \\ &\qquad \propto \begin{cases} a(\omega, \theta_1, \ldots, \theta_{j-1})\, f_{\theta_{j-1}}(\boldsymbol{z}_j | \boldsymbol{\sigma}_j) & \text{if } t = t_j \\ [1 - a(\omega, \theta_1, \ldots, \theta_{j-1})] \displaystyle\int f_\theta(\boldsymbol{z}_j | \boldsymbol{\sigma}_j)\, G(d\theta) & \text{if } t = t^{\text{new}} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Finally, when G is conjugate with respect to the Gaussian kernel, the full conditional distribution of θ_t^*, given $\{\boldsymbol{z}_j :\ t_j = t\}$, is obtained in closed form using conjugacy of the Normal-Normal model.

Sampling the concentration parameter. Finally, the concentration parameter ω can be sampled through an importance sampling step using as importance distribution the prior p_ω over ω. Denoting with M_m the selected partition for $\boldsymbol{\theta}_m$ and with T_m the number of clusters in M_m, we have

$$p(\omega \mid M_m : m = 1, \ldots, M) \propto p_\omega(\omega)\, \frac{\omega^{\sum_{m=1}^M T_m - M}}{(\omega + 2)^M (\omega^2 + \omega + 3)^M}.$$

4.6 Alternative priors over disorder-specific locations

For comparison purposes and prior sensitivity analysis we consider also two alternative priors over the disorder-specific locations: a uniform prior, which does not penalize multiplicity but incorporates the prior information on the severity of disorders, and a mixture of Dirichlet processes (DPs), which penalizes for multiplicity but does not reflect prior information.

4.6.1 Uniform prior

The uniform prior is obtained associating zero-probability to nonsensical partitions and a uniform prior over the remaining, i.e.

$$\mathbb{P}(M_b^m) \propto \begin{cases} \frac{1}{8} & \text{if } M_b^m \text{ is compatible with the natural order} \\ 0 & \text{otherwise} \end{cases}$$

The predictive distributions are

$$\theta_j \,|\theta_1, \ldots, \theta_{j-1} \;\sim\; \frac{1}{2}\,\delta_{\theta_{j-1}} + \frac{1}{2} G$$

and the full conditional distribution of t_j is

$$p(t_j = t \mid \boldsymbol{t}_\theta^{(-j)}, \boldsymbol{\theta}^{(-j)}, \boldsymbol{z}_j, \boldsymbol{\sigma}_j) \propto \begin{cases} f_{\theta_{j-1}}(\boldsymbol{z}_j|\boldsymbol{\sigma}_j) & \text{if } t = t_{j-1} \\ \displaystyle\int f_\theta(\boldsymbol{z}_j|\boldsymbol{\sigma}_j)\, G(d\theta) & \text{if } t = t^{\text{new}} \end{cases}$$

Notice that with this prior there is not a common concentration parameter and therefore there is no borrowing of information across cardiac indexes as well as no Ockham's-razor effect.

4.6.2 Mixture of DPs prior

Using as prior the mixtures of DPs, the locations $(\theta_1, \ldots, \theta_J)$, conditionally on ω, are from a DP and the law of the partition in (4.7) yields the well-known predictive distributions

$$\theta_j \,|\, \omega, \theta_1, \ldots, \theta_{j-1} \;\sim\; \sum_{t=1}^{T_{j-1}} \frac{n_t}{j-1+\omega}\,\delta_{\theta_t^*} + \frac{\omega}{j-1+\omega} G$$

with T_{j-1} the number of distinct values θ_t^* in $(\theta_1, \ldots, \theta_{j-1})$ and $n_t = \text{card}\{i \in \{1, \ldots, j-1\} : \theta_i = \theta_t^*\}$. From this, one easily deduces that the conditional prior odds against two

populations sharing the same location is

$$\frac{\mathbb{P}(\theta_{j,m} \neq \theta_{j'm} \mid \omega)}{\mathbb{P}(\theta_{j,m} = \theta_{j'm} \mid \omega)} = \frac{\Pi_2^{(2)}(1,1)}{\Pi_1^{(2)}(2)} = \omega$$

Under the mixture of DP prior, the full conditional distribution of t_j is

$$p(t_j = t \mid \boldsymbol{t}_\theta^{(-j)}, \boldsymbol{\theta}^{*(-j)}, \boldsymbol{z}_j, \boldsymbol{\sigma}_j) \propto \begin{cases} n_t^{-j}\, f_{\theta_t^*}(\boldsymbol{z}_j|\boldsymbol{\sigma}_j) & \text{if } t \in \boldsymbol{t}^{(-j)} \\ \omega \int f_\theta(\boldsymbol{z}_j|\boldsymbol{\sigma}_j)\, G(d\theta) & \text{if } t = t^{\text{new}} \end{cases}$$

where $\boldsymbol{t}^{(-j)} = \{t_{j'} : j' \neq j\}$, $\boldsymbol{\theta}^{*(-j)} = \{\theta_t^* : t \in \boldsymbol{t}^{(-j)}\}$ and n_t^{-j} denotes the number of customers already allocated to table t, after removing the j-th customer.

Moreover, if the prior p_ω for the concentration parameter is chosen to be gamma with shape a and rate b, the full conditional for the parameter ω can be obtained by generalizing the result for a single mixture of DPs in Escobar (1994), as follows. Denote with T_m the number of distinct values of $\boldsymbol{\theta}_m = \{\theta_{1,m}, \ldots, \theta_{d,m}\}$, for $m = 1, \ldots, M$ and note that ω depends on the data only through $T_1, \ldots, T_M$. The full conditional distribution of ω is:

$$\begin{aligned} p(\omega \mid T_1, \ldots, T_m) &\propto p_\omega(\omega) \cdot \prod_{m=1}^{M} p(T_m \mid \omega) \\ &\propto p_\omega(\omega) \cdot \prod_{m=1}^{M} \left[c_d(T_m)\, d!\, \omega^{T_m}\, \frac{\Gamma(\omega)}{\Gamma(\omega + d)} \right] \end{aligned}$$

where $p_\omega(\omega)$ is the prior density of ω and $c_d(T_m) = p(T_m \mid \omega = 1)$. Therefore

$$p(\omega \mid T_1, \ldots, T_m) \propto p_\omega(\omega) \cdot \omega^{\sum_m T_m - M}\, (\omega + d)^M \prod_{m=1}^{M} \left[\int_0^1 u^\omega (1-u)^{d-1} du \right]$$

Defining M auxiliary random variables u_m for $m = 1, \ldots, M$ such that $u_m \mid \omega \overset{iid}{\sim} Beta(\omega + 1,\, d)$, if $p_\omega \equiv \text{Gamma}(a,\, b)$, then

$$p(\omega \mid u_1, \ldots, u_M, T_1, \ldots, T_m) \propto \omega^{a + \sum_m T_m - M - 1}\, (\omega + d)^M\, \exp\left\{ -\omega (b - \sum_{m=1}^{M} \log(u_m)) \right\}$$

$$\propto \sum_{v=0}^{M} \binom{M}{v} \frac{d^v\, \Gamma\left(a + \sum_{m=1}^{M} T_m - v\right)}{\left(b - \sum_{m=1}^{M} \log(u_m)\right)^{a + \sum_{m=1}^{M} T_m - v}} \times \text{Gamma}\left(a + \sum_{m=1}^{M} T_m - v,\, b - \sum_{m=1}^{M} \log(u_m)\right)$$

So that the conditional distribution of ω is a mixture of $M+1$ Gamma distributions and the sampling of ω becomes

(i) Sample u_m, for $m = 1, \ldots, M$, independently from Beta$(\omega + 1,\, J)$, where J is the number of populations.

(ii) Sample v_ω from

$$p(v_\omega = v \,|\, u_1, \ldots, u_m) = \binom{M}{v} d^v \, \Gamma\Big(a + \sum_{m=1}^{M} T_m - v\Big) \, \Big(b - \sum_{m=1}^{M} \log(u_m)\Big)^v$$

for $v \in \{0, \ldots, M\}$, where T_m is the number of distinct values in $\boldsymbol{\theta}_m$, for $m = 1, \ldots, M$.

(iii) Sample ω from Gamma$\Big(a + \sum_{m=1}^{M} T_m - v,\; b - \sum_{m=1}^{M} \log(u_m)\Big)$.

4.7 Results

4.7.1 Simulation studies

Generating mechanism with underlying relevant factor

We perform here a series of simulation studies with two main goals. First, we aim to highlight the drawbacks of clustering based on the entire distribution, if compared to our proposal, when applied to small sample sizes. Second, we check the model's ability of detecting the presence of underlying relevant factors in the sense described in Section 4.3.2. To accomplish the first goal, we compare the results obtained using our model against the nested Dirichlet process (NDP) of Rodriguez et al. (2008), which is probably the most widely used Bayesian model to cluster populations. Mimicking the real hypertensive dataset, we simulate data for 4 samples, ideally corresponding to four diseases, with respective sample sizes of 50, 19, 9 and 22, which correspond to the sample sizes of the real data investigated in Section 4.7.2. Since the NDP does not allow to treat jointly multiple response variables, we consider only one response variable to ensure a fair comparison. The observations are sampled from the following distributions and 100 simulation studies are performed.

$$\begin{aligned}
X_{i,1} &\overset{iid}{\sim} 0.5\,\mathrm{N}(0,\, 0.5) + 0.5\,\mathrm{N}(2,\, 0.5) && \text{for } i = 1, \ldots, n_1 \\
X_{i,2} &\overset{iid}{\sim} 0.5\,\mathrm{N}(2,\, 0.5) + 0.5\,\mathrm{N}(4,\, 0.5) && \text{for } i = 1, \ldots, n_2 \\
X_{i,3} &\overset{iid}{\sim} 0.5\,\mathrm{N}(4,\, 0.5) + 0.5\,\mathrm{N}(6,\, 0.5) && \text{for } i = 1, \ldots, n_3 \\
X_{i,4} &\overset{iid}{\sim} 0.5\,\mathrm{N}(6,\, 0.5) + 0.5\,\mathrm{N}(8,\, 0.5) && \text{for } i = 1, \ldots, n_4
\end{aligned}$$

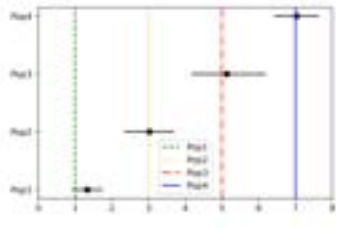
Note that here the true data generating process corresponds to samples from distinct distributions with pairwise sharing of a mixture component. Alternative scenarios are considered in the additional simulation studies in the following sections.

The implementation of the NDP was carried out through the marginal sampling scheme proposed in Zuanetti et al. (2018) extended in order to accomodate hyperpriors on the concentration parameters of the NDP. To simplify the choice of the hyperparameters, as suggested by Gelman et al. (2013, p. 535 and p. 551–554) we estimate both models over standardized data. For our model, we set $G_m = \text{N}(0, 1)$ and $P_{0,m} = \text{NIG}(\mu = 0, \tau = 1, \alpha = 2, \beta = 4)$. Here, $\text{NIG}(\mu, \tau, \alpha, \beta)$ indicates a normal inverse gamma distribution. The base distribution for the NDP is $\text{NIG}(\mu = 0, \tau = 0.01, \alpha = 3, \beta = 3)$, as in Rodriguez et al. (2008). Finally, we use Gamma priors with shape 3 and rate 3 for all concentration parameters, which is a common choice. For each simulation study, we perform 10,000 iterations of the MCMC algorithms with the first 5,000 used as burn-in.

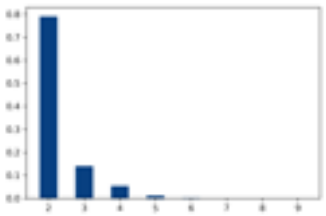
Table 4.1: Simulation studies summaries.

	sHDP			**NDP**		
Partitions	MAP count	Average post. prob.	Median post. prob.	MAP count	Average post. prob.	Median post. prob.
{1,2,3,4}	0	0.000	0.000	0	0.000	0.000
{1}{2,3,4}	0	0.000	0.000	2	0.020	0.000
{1,2}{3,4}	0	0.000	0.000	**72**	**0.695**	**0.860**
{1,3,4}{2}	0	0.000	0.000	0	0.000	0.000
{1}{2}{3,4}	0	0.027	0.007	3	0.035	0.000
{1,2,3}{4}	0	0.000	0.000	5	0.061	0.000
{1,4}{2,3}	0	0.000	0.000	0	0.000	0.000
{1}{2,3}{4}	1	0.054	0.015	0	0.014	0.000
{1,3}{2,4}	0	0.000	0.000	0	0.000	0.000
{1,2,4}{3}	0	0.000	0.000	0	0.000	0.000
{1}{2,4}{3}	0	0.000	0.000	0	0.000	0.000
{1,2}{3}{4}	0	0.004	0.000	18	0.175	0.032
{1,3}{2}{4}	0	0.000	0.000	0	0.000	0.000
{1,4}{2}{3}	0	0.000	0.000	0	0.000	0.000
{1}{2}{3}{4}	**99**	**0.915**	**0.954**	0	0.000	0.000

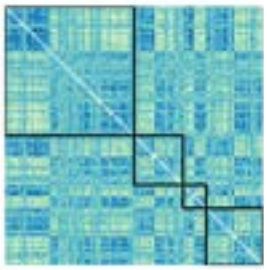
Table 4.1 displays summaries of the results on population clustering, darker rows correspond to nonsensical partitions. The true clustering structure is given by the finest partition. As already observed in Rodriguez et al. (2008), the NDP tends to identify fewer, rather than more clusters, due to the presence of small sample sizes. Using the *maximum a posteriori* estimate, our model correctly identifies the partition in 99 out of 100 simulation studies and a partition with three elements or more in 100 out of 100 simulation studies.



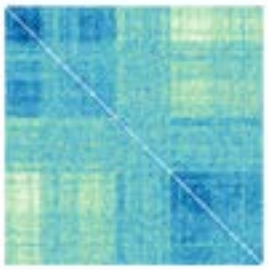
(a) 95% credible intervals for population-specific locations



(b) Number of second-level clusters.



(c) Co-clustering.

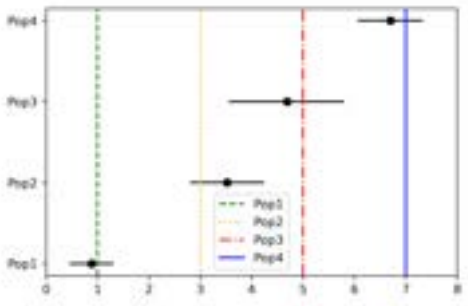


(d) Co-clustering.

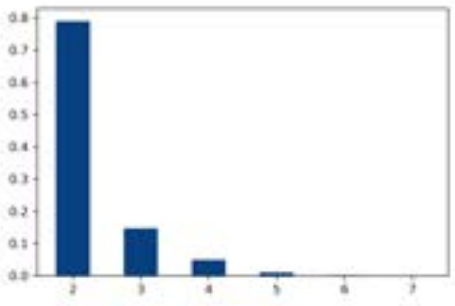
Figure 4.5: Panel (a): Mean point estimates and 95% credible intervals for the four populations, vertical lines correspond to true values. Panel (b): Posterior distribution on the number of second-level clusters. Panels (c) and (d): heatmaps of second level clustering, darker colors correspond to higher probability of co-clustering; in (c) patients are ordered based on the diagnosis and the four black squares highlight the within-sample probabilities and in (d) patients are reordered based on co-clustering probabilities.

The same counts for the NDP are 0 out of 100 and 21 out of 100. Analogous conclusions can be drawn looking at posterior probability averages and medians across the 100 simulation studies (see Table 4.1) leaving no doubt about the model to be preferred under this scenario. Finally, we randomly select three simulation studies among the 100 to achieve a better understanding the performance in the estimation of the other model parameters. Here we comment on one of the studies, the other two leading to similar results are reported in Figures 4.6 and 4.7. Figure 4.5a shows point estimates and credible intervals for the population-specific location parameters $\theta_1, \theta_2, \theta_3, \theta_4$. The true means belong to the 95% credible intervals. Moreover, it turns out that the model is able to detect the presence of two clusters of subjects leading to a posterior distribution for the number of clusters that is rather concentrated on the true value, see Figure 4.5b–4.5d. Moreover, the point estimate for the subject partition, obtained minimizing the Binder loss function, also contains two clusters, proving the ability of the model in detecting the underlying relevant factor. In the following, a number of additional simulation studies are conducted, both using alternative specifications over the disorder-specific parameters and different data generating mecha-

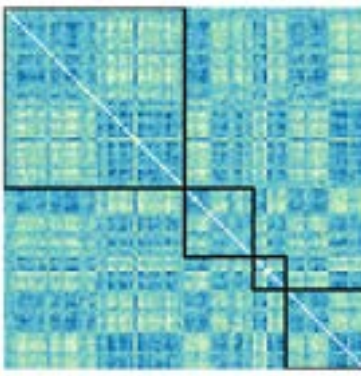
nisms, the results highlight a good performance of the model, which appears also able to detect outliers, to highlight non-location effects of the disorders and to produce reliable outputs even under deviation from symmetry.
In Figures 4.6 and 4.7 below, we display the plots regarding the inference for two additional randomly selected simulation studies among the 100. Like for the simulation study already discussed, the true means belong to the 95% credible intervals and the model correctly identifies the two clusters.



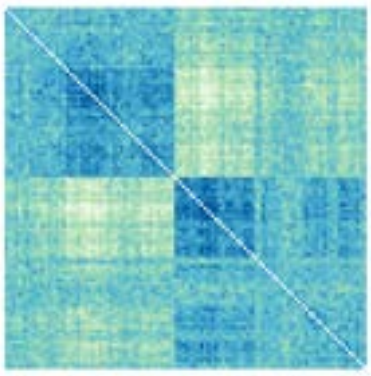
(a) Inference on location parameters



(b) Number of second-level clusters.

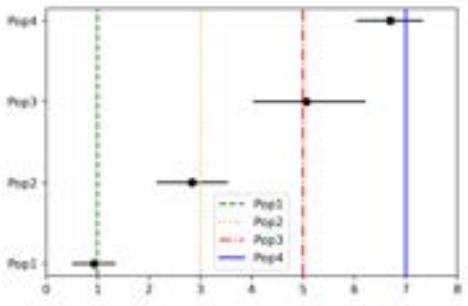


(c) Co-clustering.

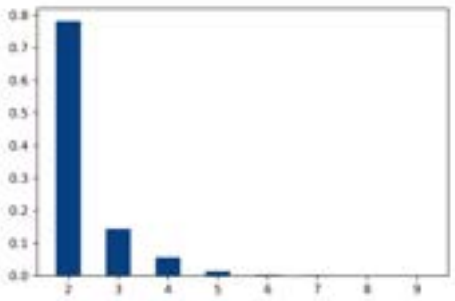


(d) Co-clustering.

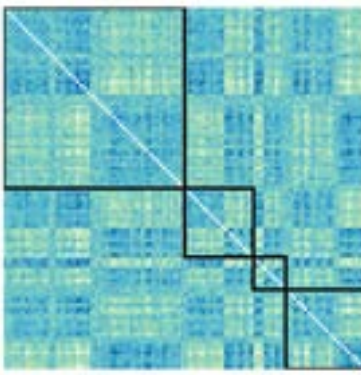
Figure 4.6: Panel (a): Results of the 37th simulation study. Mean point estimates and 95% credible intervals for the four populations, vertical lines correspond to true values. Panel (b): Posterior distribution on the number of second-level clusters. Panels (c) and (d): heatmaps of second level clustering, darker colors correspond to higher probability of co-clustering; in (c) patients are ordered based on the diagnosis and the four black squares highlight the within-sample probabilities and in (d) patients are reordered based on co-clustering probabilities.



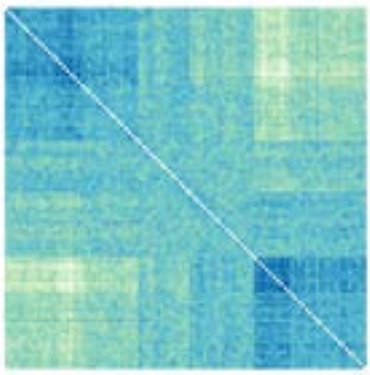
(a) Inference on location parameters



(b) Number of second-level clusters.



(c) Co-clustering.



(d) Co-clustering.

Figure 4.7: Results of the 9th simulation study. Panel (a): Mean point estimates and 95% credible intervals for the four populations, vertical lines correspond to true values. Panel (b): Posterior distribution on the number of second-level clusters. Panels (c) and (d): heatmaps of second level clustering, darker colors correspond to higher probability of co-clustering; in (c) patients are ordered based on the diagnosis and the four black squares highlight the within-sample probabilities and in (d) patients are reordered based on co-clustering probabilities.

Lastly, we display the results obtained over the same simulated data using the alternative priors described in Section 4.6.

Table 4.2: Simulation studies summaries.

	sHDP-with mixture of DPs			**sHDP-with unifor prior**		
Partitions	MAP count	Average post. prob.	Median post. prob.	MAP count	Average post. prob.	Median post. prob.
{1,2,3,4}	0	0.000	0.000	0	0.000	0.000
{1}{2,3,4}	0	0.000	0.000	0	0.000	0.000
{1,2}{3,4}	0	0.000	0.000	0	0.000	0.000
{1,3,4}{2}	0	0.000	0.000	0	0.000	0.000
{1}{2}{3,4}	5	0.083	0.022	0	0.030	0.009
{1,2,3}{4}	0	0.000	0.000	0	0.001	0.000
{1,4}{2,3}	0	0.000	0.000	0	0.000	0.000
{1}{2,3}{4}	2	0.056	0.012	1	0.051	0.014
{1,3}{2,4}	0	0.000	0.000	0	0.000	0.000
{1,2,4}{3}	0	0.000	0.000	0	0.000	0.000
{1}{2,4}{3}	0	0.000	0.000	0	0.000	0.000
{1,2}{3}{4}	0	0.002	0.000	0	0.003	0.000
{1,3}{2}{4}	0	0.000	0.000	0	0.000	0.000
{1,4}{2}{3}	0	0.000	0.000	0	0.000	0.000
{1}{2}{3}{4}	**93**	**0.859**	**0.918**	**99**	**0.916**	**0.956**

Both models perform better than the NDP, whose results are in Table 4.1, confirming the advantages of location-based clustering in presence of small sample sizes, when compared to distribution-based clustering. Moreover, sHDP-with mixture of DPs has a slightly worst performance with respect to our main proposal, as expected, since the corresponding prior incorporates less information and ignores the natural order of the four populations.

Generating mechanism with outliers

We present here a simulation study with a twofold goal: (1) compare again the location–based clustering approach of our proposal with the distribution–based clustering approach of the nested Dirichlet process (NDP) under a different DGP; (2) study the performance of our model in presence of outliers. The simulated data have been sampled according to the following DGP

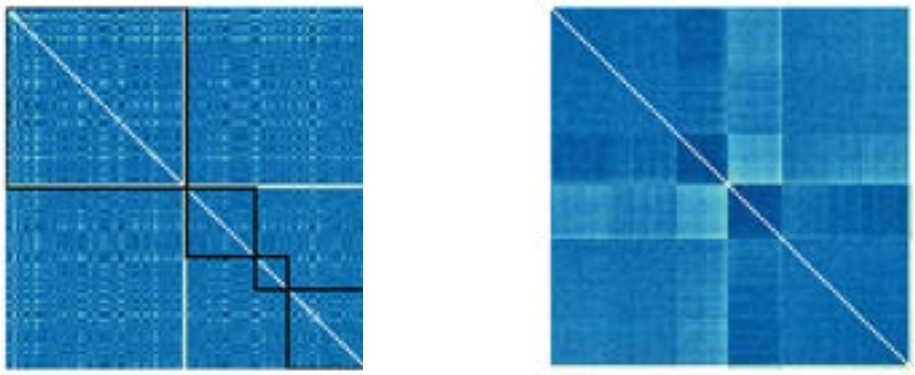
$$
\begin{aligned}
\textbf{DGP 1:} \quad & X_{i,1} \overset{iid}{\sim} \mathrm{N}(\,0,\,0.5\,) && \text{for } \; i = 1, \ldots, n_1 - 1 \\
& X_{n_1,1} \sim \mathrm{N}(\,4,\,0.5\,) && \\
& X_{i,2} \overset{iid}{\sim} \mathrm{N}(\,1,\,0.5\,) && \text{for } \; i = 1, \ldots, n_2 \\
& X_{i,3} \overset{iid}{\sim} \mathrm{N}(\,1,\,0.5\,) && \text{for } \; i = 1, \ldots, n_3 \\
& X_{i,4} \overset{iid}{\sim} \mathrm{N}(\,2,\,0.5\,) && \text{for } \; i = 1, \ldots, n_4
\end{aligned}
$$

Thus, the true partition is $\{1\}, \{2, 3\}, \{4\}$. Moreover, there is one outlier in the first sample.

Table 4.3: Posterior probabilities over the space of partitions.

	s-HDP	NDP
{1,2,3,4}	0	0
{1}{2,3,4}	0	0
{1,2}{3,4}	0	0
{1,3,4}{2}	0	0
{1}{2}{3,4}	0	0
{1,2,3}{4}	0.013	**0.980**
{1,4}{2,3}	0	0
{1}{2,3}{4}	**0.771**	0.010
{1,3}{2,4}	0	0
{1,2,4}{3}	0	0
{1}{2,4}{3}	0	0
{1,2}{3}{4}	0.006	0.020
{1,3}{2}{4}	0	0
{1,4}{2}{3}	0	0
{1}{2}{3}{4}	0.210	0

Table 4.3 displays the posterior probabilities obtained using our model (s-HDP) and the NDP. Our model largely outperforms the competitor.



(a) Co-clustering. (b) Co-clustering.

Figure 4.8: Posterior similarity matrices for the simulation study under DGP 1. In (a) patients are ordered based on the diagnosis; in (b) patients are reordered based on co-clustering probabilities.

Figure 4.8 shows the posterior co-clustering probabilities obtained in the simulation study. Our proposal is able to correctly identify the outlier.

Generating mechanism with non-location effects

We present here a simulation study with a twofold goal: (1) compare again the location–based clustering approach of our proposal with the distribution–based clustering approach of the nested Dirichlet process (NDP) under a different DGP; (2) study the performance of our model in the case in which heterogeneity between populations is not fully explained by shift in locations. The simulated data have been sampled according to the following DGP.

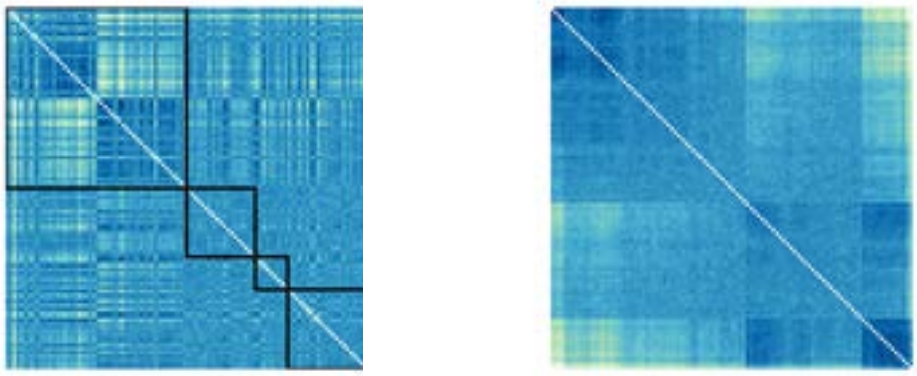
$$
\begin{aligned}
\textbf{DGP 2:} \quad & X_{i,1} \overset{iid}{\sim} 0.5\,\mathrm{N}(-1,\,0.5) + 0.5\,\mathrm{N}(1,\,0.5) && \text{for } i = 1, \dots, n_1 \\
& X_{i,2} \overset{iid}{\sim} \mathrm{N}(1,\,0.5) && \text{for } i = 1, \dots, n_2 \\
& X_{i,3} \overset{iid}{\sim} \mathrm{N}(1,\,0.5) && \text{for } i = 1, \dots, n_3 \\
& X_{i,4} \overset{iid}{\sim} \mathrm{N}(2,\,0.5) && \text{for } i = 1, \dots, n_4
\end{aligned}
$$

Thus, the true partition is $\{1\}, \{2, 3\}, \{4\}$. Moreover, the relative effect of the first population w.r.t. the others is not fully explained by the shift of the location, since the whole distribution is different and not only the mean.

Table 4.4: Posterior probabilities over the space of partitions.

	s-HDP	NDP
{1,2,3,4}	0	0
{1}{2,3,4}	0.001	0
{1,2}{3,4}	0	0
{1,3,4}{2}	0	0
{1}{2}{3,4}	0.001	0
{1,2,3}{4}	0.058	**0.98**
{1,4}{2,3}	0	0
{1}{2,3}{4}	**0.706**	0.01
{1,3}{2,4}	0	0
{1,2,4}{3}	0	0
{1}{2,4}{3}	0	0
{1,2}{3}{4}	0.019	0.02
{1,3}{2}{4}	0	0
{1,4}{2}{3}	0	0
{1}{2}{3}{4}	0.214	0

Table 4.4 displays the posterior probabilities obtained using our model (s-HDP) and the NDP. Our model largely outperforms the competitor.



(a) Co-clustering. (b) Co-clustering.

Figure 4.9: Posterior similarity matrices for the simulation study under DGP 2. In (a) patients are ordered based on the diagnosis and the four black squares highlight the within-sample probabilities; in (b) patients are reordered based on co-clustering probabilities.

Figure 4.9 shows the posterior co-clustering probabilities. Our proposal is able to correctly identify the non–location effect (see Figure 4.9(a)).

Simulation studies under non-symmetric data generating process

We present here three simulation studies to check the performance of the model under deviations from symmetry. The simulated data have been sampled according to the following DGPs.

$$
\begin{aligned}
\textbf{DGP 3:}\quad & X_{i,1} \overset{iid}{\sim} \mathrm{N}(0,\, 0.5\,) && \text{for } i = 1, \dots, n_1 \\
& X_{i,2} \overset{iid}{\sim} \mathrm{Gamma}(\,3,\, 3) && \text{for } i = 1, \dots, n_2 \\
& X_{i,3} \overset{iid}{\sim} \mathrm{N}(\,1,\, 0.5\,) && \text{for } i = 1, \dots, n_3 \\
& X_{i,4} \overset{iid}{\sim} \mathrm{N}(\,2,\, 0.5\,) && \text{for } i = 1, \dots, n_4
\end{aligned}
$$

$$
\begin{aligned}
\textbf{DGP 4:}\quad & X_{i,1} \overset{iid}{\sim} 0.7\,\mathrm{N}(\,-1,\, 0.5\,) + 0.3\,\mathrm{N}(\,1,\, 0.5\,) && \text{for } i = 1, \dots, n_1 \\
& X_{i,2} \overset{iid}{\sim} \mathrm{N}(\,1,\, 0.5\,) && \text{for } i = 1, \dots, n_2 \\
& X_{i,3} \overset{iid}{\sim} \mathrm{N}(\,1,\, 0.5\,) && \text{for } i = 1, \dots, n_3 \\
& X_{i,4} \overset{iid}{\sim} \mathrm{N}(\,2,\, 0.5\,) && \text{for } i = 1, \dots, n_4
\end{aligned}
$$

$$
\begin{aligned}
\textbf{DGP 5:}\quad & X_{i,1} \overset{iid}{\sim} \mathrm{Gamma}(\,10,\, 10) && \text{for } i = 1, \dots, n_1 \\
& X_{i,2} \overset{iid}{\sim} \mathrm{Gamma}(\,10,\, 10) && \text{for } i = 1, \dots, n_2 \\
& X_{i,3} \overset{iid}{\sim} \mathrm{Gamma}(\,10,\, 10) && \text{for } i = 1, \dots, n_3 \\
& X_{i,4} \overset{iid}{\sim} 0.5\,\mathrm{N}(\,0,\, 0.5\,) + 0.5\,\mathrm{N}(\,2,\, 0.5\,) && \text{for } i = 1, \dots, n_4
\end{aligned}
$$

Under all DGPs the model is misspecified due to lack of symmetry in one or more populations. Under DGP 3 and DGP 4 the true partition is $\{1\}, \{2, 3\}, \{4\}$, while under DGP 5 it is $\{1, 2, 3, 4\}$. In DGP 3 the second population differs from the others also in distribution (what we called non-location effect), the same is true for the first and the fourth populations respectively under DGP 4 and DGP 5. Table 4.5 shows that the model is able to detect the right clustering of the population-specific locations under all three DGPs. Moreover, Figure 4.10 shows co-clustering probabilities that differ in correspondence of the populations affected by non-location effects, more or less evidently based on the DGP used to generate the data. This results are reassuring: under misspecification, not only the model appears robust in estimating locations' partitions, but also, the different within-population patterns of co-clustering probabilities still highlighting heterogeneities different than shifts in population-specific locations.

Table 4.5: Posterior probabilities over the space of partitions.

	DGP 3	DGP 4	DGP 5
{1,2,3,4}	0	0.001	**0.494**
{1}{2,3,4}	0	0	0.023
{1,2}{3,4}	0	0	0.014
{1,3,4}{2}	0	0	0
{1}{2}{3,4}	0	0	0.004
{1,2,3}{4}	0.016	0	0.375
{1,4}{2,3}	0	0	0
{1}{2,3}{4}	**0.736**	**0.788**	0.047
{1,3}{2,4}	0	0	0
{1,2,4}{3}	0	0	0
{1}{2,4}{3}	0	0	0
{1,2}{3}{4}	0.015	0	0.030
{1,3}{2}{4}	0	0	0
{1,4}{2}{3}	0	0	0
{1}{2}{3}{4}	0.232	0.211	0.012

(a) Co-clustering DGP 3. (b) Co-clustering DGP 4. (c) Co-clustering DGP 5.

Figure 4.10: Posterior similarity matrices under DGP 3-4-5. Patients are ordered based on the diagnosis and the four black squares highlight the within-sample probabilities.

Table 4.6: Posterior probabilities over partitions of means. Maximum a posteriori probabilities are in **bold**.

partitions	CI	CWI	LVMI	IVST	LVPW	EF	FS	EW	AW	E/A
{C,G,M,S}	0.021	0.000	0.000	0.000	0.000	**0.365**	**0.303**	0.096	0.000	0.000
{C}{G,M,S}	0.002	**0.546**	0.001	0.083	0.016	0.078	0.190	0.021	0.036	0.000
{C,G}{M,S}	0.002	0.000	0.001	0.000	0.000	0.037	0.038	0.072	0.076	0.049
{C,M,S}{G}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M,S}	0.001	0.139	0.001	0.019	0.024	0.028	0.078	0.042	0.232	0.055
{C,G,M}{S}	**0.463**	0.000	**0.595**	0.000	0.000	0.276	0.045	**0.498**	0.020	0.002
{C,S}{G,M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,M}{S}	0.146	0.099	0.188	**0.551**	**0.672**	0.074	0.164	0.092	0.260	0.033
{C,M}{G,S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G}{M}{S}	0.233	0.000	0.107	0.000	0.000	0.083	0.062	0.114	0.091	0.371
{C,M}{G}{S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,S}{G}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M}{S}	0.133	0.216	0.108	0.347	0.288	0.060	0.121	0.065	**0.287**	**0.491**
$\sum \log_{15}\left(p_i^{-p_i}\right)$	0.501	0.430	0.415	0.361	0.289	0.632	0.688	0.598	0.613	0.424

4.7.2 Impact of hypertensive disorders on maternal cardiac dysfunction

Our analysis is based on the dataset of Tatapudi & Pasumarthy (2017a), which can be obtained from https://data.mendeley.com/datasets/d72zr4xggx/1. The dataset contains observations for 10 cardiac function measurements collected through a prospective case-control study on women in the third semester of pregnancy divided in $n_1 = 50$ control cases (C), $n_2 = 19$ patients with gestational hypertension (G), $n_3 = 9$ patients with mild preeclampsia (M) and $n_4 = 22$ patients with severe preeclampsia (S). The cases are women admitted to King George Hospital Visakhapatnam India from 2012 to 2014. The healthy sample is composed by normotensive pregnant women. All women with hypertension were on antihypertensive treatment with oral Labetalol or Nifedipine. Women with severe hypertension were treated with either oral nifedipine and parenteral labetalol or a combination. For more details on the dataset, we refer to Tatapudi & Pasumarthy (2017b). The prior specification is the same as in the previous section. Section 4.7.3 contains a prior-sensitivity analysis and shows rather robust results w.r.t. different prior specifications. Inference is based on 10,000 MCMC iterations with the first half used as burn-in.

Table 4.6 displays the posterior distributions for the partitions of unknown disease-specific means along with the corresponding entropy measurements, that can be used as measures of uncertainty. First of all, we notice that, if one takes also the ordering among distinct disease-specific locations into account: the posterior partition probabilities are, as desired,

Table 4.7: Posterior probabilities over ordered partitions of means.

cardiac index	ordered partition with highest posterior probability	posterior prob
CI	{C,G,M}>{S}	0.463
CWI	{C}<{G,M,S}	0.546
LVMI	{C,G,M}<{S}	0.595
IVST	{C}<{G,M}<{S}	0.548
LVPW	{C}<{G,M}<{S}	0.671
EF	{C,G,M,S}	0.365
FS	{C,G,M,S}	0.303
EW	{C,G,M}>{S}	0.497
AW	{C}<{G,M}<{S}	0.256
E/A	{C}>{G}>{M}>{S}	0.466

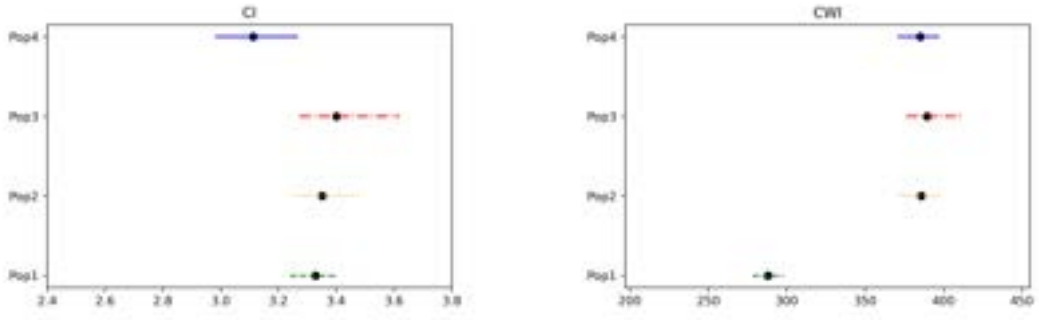


Figure 4.11: 95% credible intervals for population-specific locations for CI and CWI

concentrated on specific orders of the associated unique values for all ten cardiac indexes. For instance, we have $\mathbb{P}(\{\theta_{C,CI} = \theta_{G,CI} = \theta_{M,CI}\}\{\theta_{S,CI}\} \mid X) = \mathbb{P}(\theta_{C,CI} = \theta_{G,CI} = \theta_{M,CI} > \theta_{S,CI} \mid X) = 0.463$. The ordered partitions with the highest posterior probability are displayed in Table 4.7.

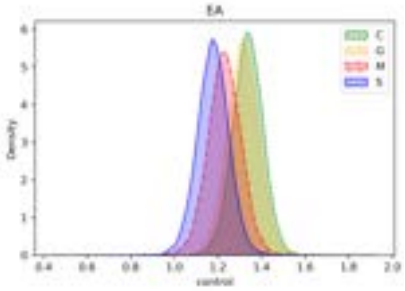
Considering the posterior probabilities summarized in Table 4.6 and in Table 4.7, we find that the cardiac index (CI) is reduced in severe preeclampsia compared to all other patients, indicating reduced myocardial contractility in the presence of the most severe disorder. The cardiac work index (CWI) is a good indicator to distinguish between cases and control, but not among cases. The left ventricular mass index (LVMI) is increased in severe preeclampsia patients compared to other pregnant women, indicating ventricular remodelling. While inter ventricular septal thickness (IVST) and left ventricular posterior wall thickness (LVPW) differ both between cases and controls and between severe preeclampsia and other disorders, indicating a progressive increase in the indexes with the severity of the disorder. The posterior probabilities associated to indexes of systolic function such as

Table 4.8: Posterior probabilities over partitions of means. Maximum a posteriori probabilities are in **bold**.

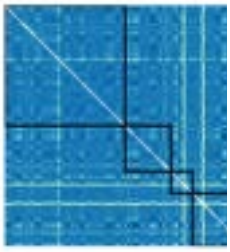
partitions	CI	CWI	LVMI	IVST	LVPW	EF	FS	EW	AW	E/A
{C,G,M,S}	0.019	0.000	0.000	0.000	0.000	**0.332**	**0.247**	0.078	0.000	0.000
{C}{G,M,S}	0.002	**0.643**	0.001	0.114	0.031	0.065	0.130	0.048	0.080	0.000
{C,G}{M,S}	0.004	0.000	0.003	0.000	0.000	0.044	0.019	0.152	0.073	0.103
{C,M,S}{G}	0.004	0.000	0.000	0.000	0.000	0.037	0.105	0.013	0.000	0.000
{C}{G}{M,S}	0.002	0.065	0.002	0.047	0.078	0.027	0.036	0.063	**0.424**	0.167
{C,G,M}{S}	**0.316**	0.000	**0.527**	0.000	0.000	0.178	0.032	**0.288**	0.002	0.000
{C,S}{G,M}	0.023	0.000	0.000	0.000	0.000	0.019	0.103	0.006	0.000	0.000
{C}{G,M}{S}	0.173	0.089	0.124	**0.472**	**0.594**	0.033	0.054	0.064	0.140	0.042
{C,M}{G,S}	0.002	0.000	0.001	0.003	0.000	0.044	0.031	0.017	0.000	0.000
{C,G,S}{M}	0.018	0.000	0.000	0.000	0.000	0.061	0.067	0.016	0.000	0.000
{C}{G,S}{M}	0.005	0.163	0.001	0.095	0.006	0.028	0.040	0.015	0.016	0.000
{C,G}{M}{S}	0.213	0.000	0.124	0.000	0.000	0.052	0.014	0.121	0.036	0.241
{C,M}{G}{S}	0.074	0.000	0.137	0.003	0.000	0.041	0.022	0.055	0.001	0.000
{C,S}{G}{M}	0.014	0.000	0.000	0.000	0.000	0.011	0.067	0.004	0.000	0.000
{C}{G}{M}{S}	0.133	0.040	0.079	0.265	0.291	0.029	0.033	0.059	0.229	**0.448**
$\sum \log_{15}\left(p_i^{-p_i}\right)$	0.687	0.407	0.509	0.501	0.371	0.828	0.886	0.823	0.582	0.505

ejection fraction (EF) and fraction shortening (FS) are relatively concentrated on the partition of complete homogeneity, letting us to conclude that no differences are present among patients. For what concerns parameters of diastolic function, the posterior distribution for the E-wave indicator identifies a modified index in severe preeclampsia patients, while the mean E/A ratio indicates a decreasing diastolic function with the severity of the disorder. The posterior for the A-wave index is actually concentrated on three distinct partitions, leaving a relatively high uncertainty regarding the modifications of the index. However, considering jointly the three partitions with the highest posterior probability, differences are detected between control and cases with a total posterior probability equal to 0.779. Figure 4.11 shows point estimates and credible intervals for disorder-specific location parameters for the first two cardiac indexes, the same plots for all cardiac indexes can be found below.

Table 4.8 shows the results obtained using the prior in (4.7), instead of (4.8). First of all note that for all ten cardiac indexes, the posterior associates negligible probabilities to partitions that are in contrast with the natural order of the diagnoses. This is particularly reassuring in that the model, even without imposing such an order a priori, is able to single it out systematically across cardiac indexes. Moreover, we observe how the partitions identified by MAP are the same of Table 4.6 for all cardiac index except AW. However, even under this alternative prior, the A-wave index is concentrated on the same three distinct partitions, that lead to conclude that it exists a difference between cases and control.

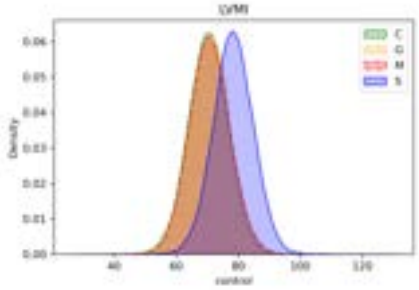


(a) density estimation

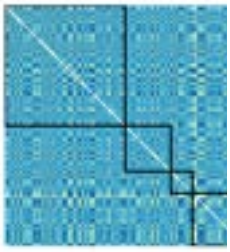


(b) co-clustering

(c) co-clustering



(d) density estimation



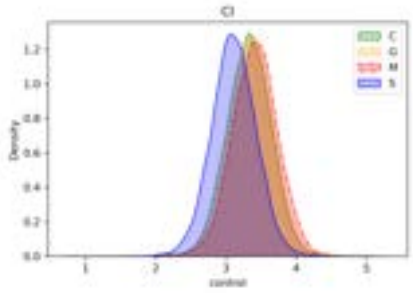
(e) co-clustering

(f) co-clustering

Figure 4.12: Panels (a) and (d): density estimates. Panels (b)–(c) and (e)–(f): heatmaps of the posterior probabilities of co-clustering; in (b) and (e) patients are ordered based on the diagnosis and the four black squares highlight the within-sample probabilities; in (c) and (f) patients are reordered based on co-clustering probabilities.

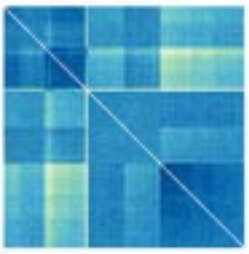
As far as prediction and second-level clustering are concerned, Figure 4.12 displays the density estimates and the heatmap of co-clustering probabilities between pairs of patients for the E/A ratio and LVMI. Figure 4.12b shows that co-clustering probabilities are similar within and across diagnoses, indicating that the effect of the diseases on the distribution of the cardiac index is mostly explained through shifts between disease-specific locations. Moreover Figure 4.12b suggests the presence of three outliers that have low probability of co-clustering with all the other subjects and that would be ignored by the model using a more traditional ANOVA structure. Contrary, Figure 4.12e shows a slightly different pattern for co-clustering probabilities in the fourth square, which suggests that the heterogeneity between severe preeclampsia patients and the others patients is not entirely explained by shifts in disease-specific locations. Finally, Figure 4.12f suggests the presence of an underlying relevant factor. The corresponding figures for all ten response variables are reported below and can be used for prediction and for a graphical analysis aimed at controlling the presence of underlying relevant factors, outliers and differences across diseases distinct from shifts between disease-specific locations.
Our results are coherent with almost all of the findings in Tatapudi & Pasumarthy (2017b) where results were obtained through a series of independent frequentist tests. However,

importantly, we are able to provide more details thanks to the simultaneous comparison approach and the latent clustering of subjects. For instance, considering the response LVMI, Tatapudi & Pasumarthy (2017b) detected a significant increase in cases compare to controls and an increase in severe preeclampsia compared to gestional hypertensive and mild preeclampsia patients. Such results do not clarify whether a modification exists between the control group and gestional hypertensive patients or between the latter and mild preeclampsia patients. Moreover, in their analysis, no information can be deducted regarding effects different than shifts in locations, presence of underlying common factors or outliers. The figures below display density estimates, heatmaps of co-clustering between patients, and locations' credible intervals for all response variables.

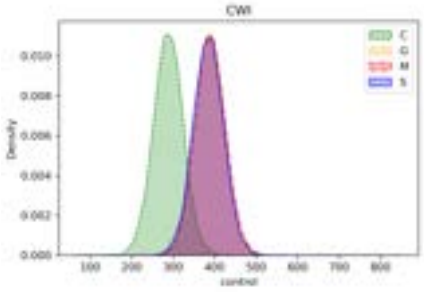


(a) density estimation

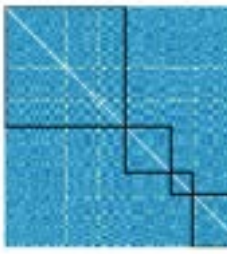
(b) co-clustering



(c) co-clustering

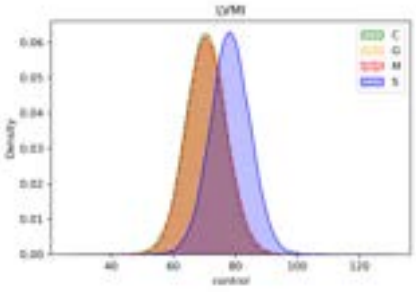


(a) density estimation

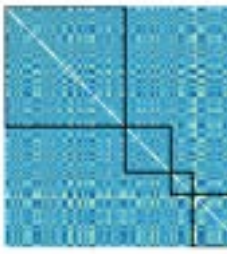


(b) co-clustering

(c) co-clustering

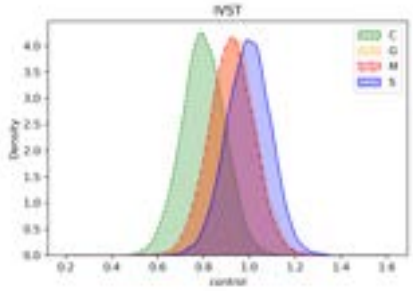


(a) density estimation

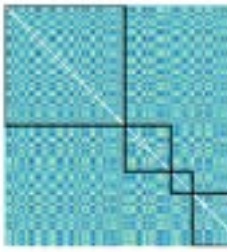


(b) co-clustering

(c) co-clustering

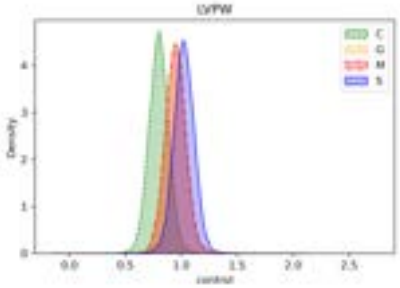


(a) density estimation

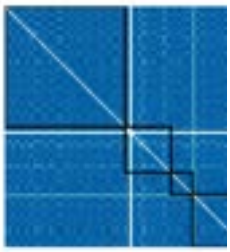


(b) co-clustering

(c) co-clustering

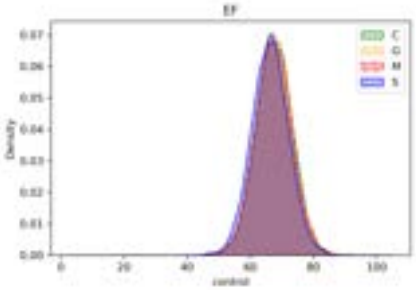


(a) density estimation

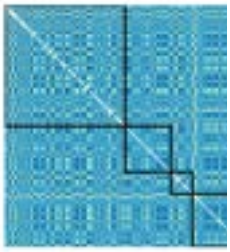


(b) co-clustering

(c) co-clustering

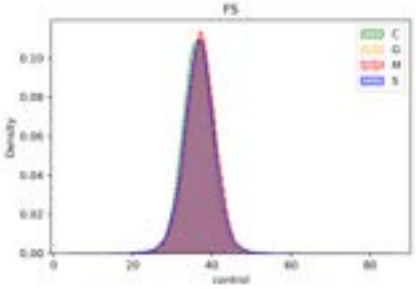


(a) density estimation

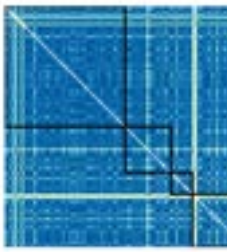


(b) co-clustering

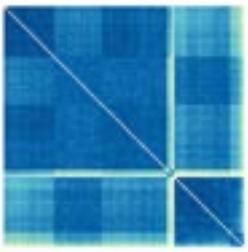
(c) co-clustering



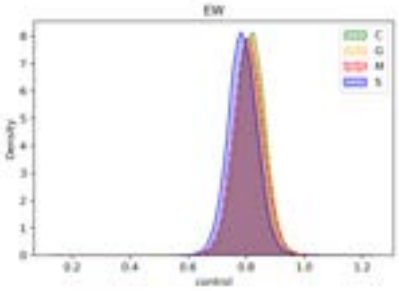
(a) density estimation



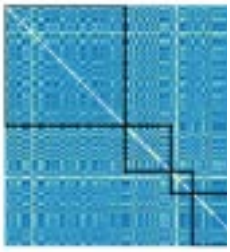
(b) co-clustering



(c) co-clustering

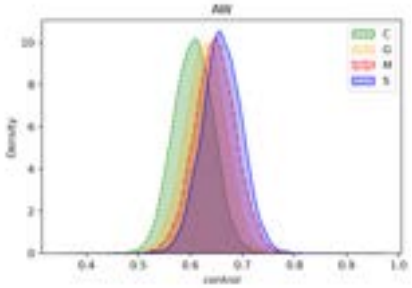


(a) density estimation

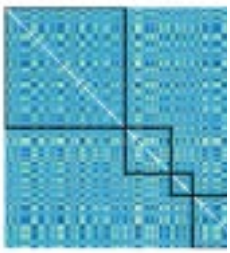


(b) co-clustering

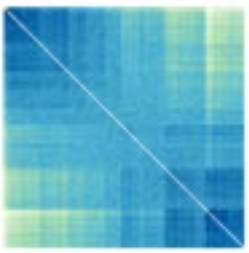
(c) co-clustering



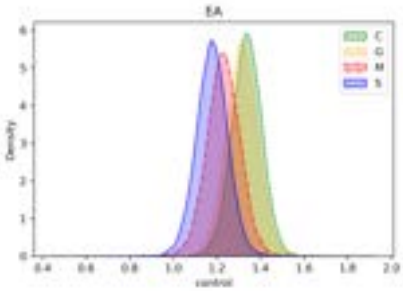
(a) density estimation



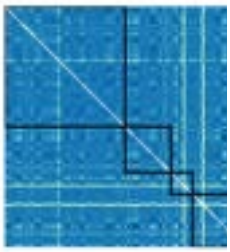
(b) co-clustering



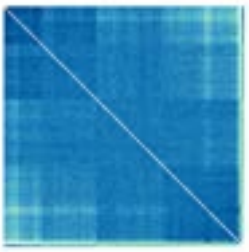
(c) co-clustering



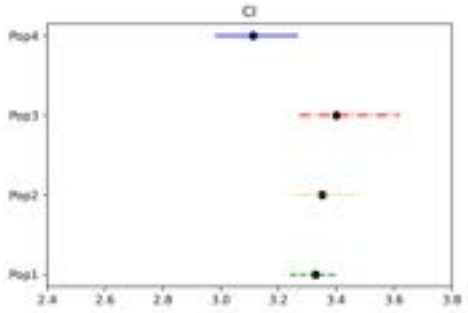
(a) density estimation

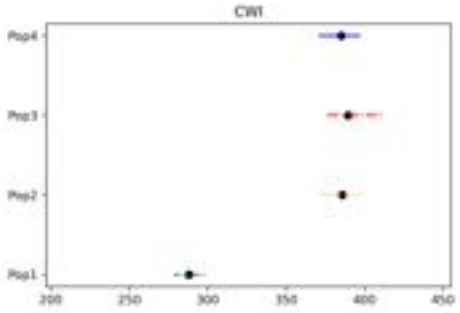


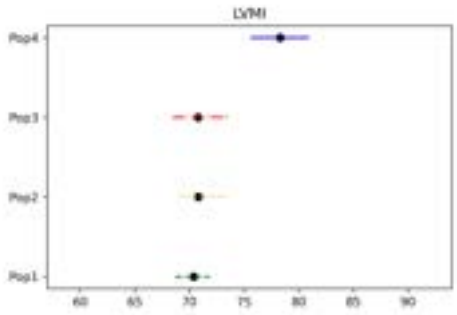
(b) co-clustering

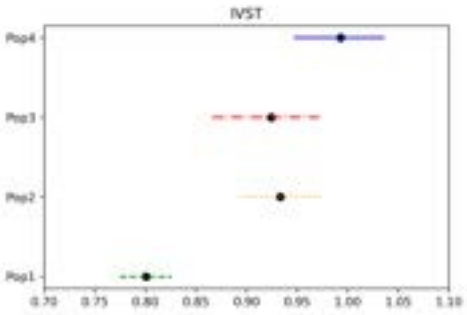


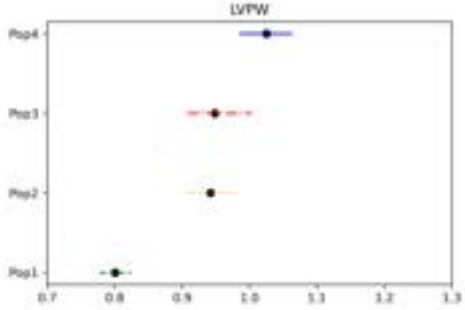
(c) co-clustering

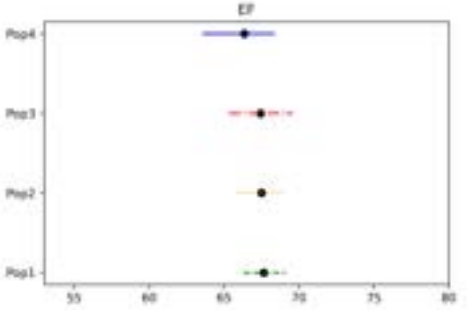


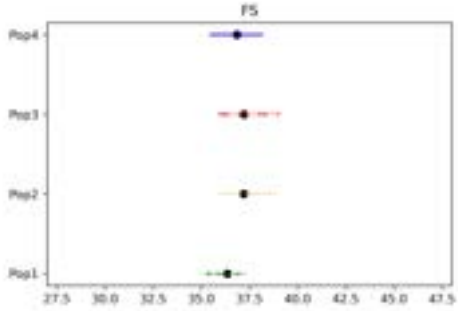


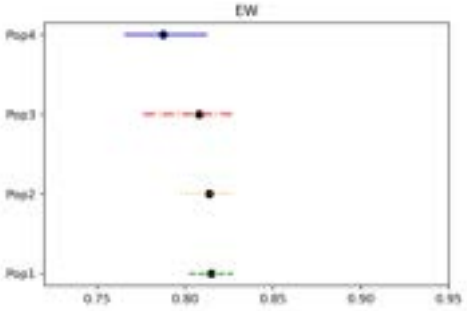

LVMI
Pop4
Pop3
Pop2
Pop1
60
65
70
75
80
85
90

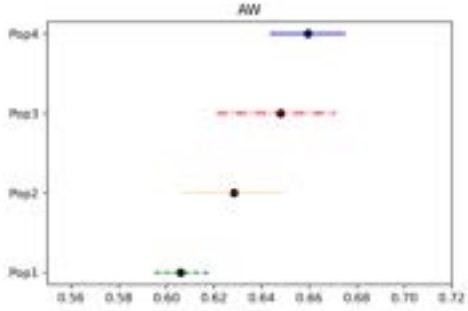

IVST
Pop4
Pop3
Pop2
Pop1
0.70
0.75
0.80
0.85
0.90
0.95
1.00
1.05
1.10

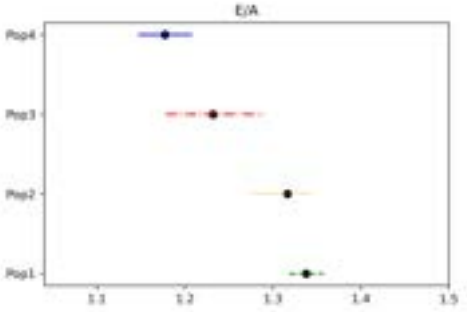

LVPW
Pop4
Pop3
Pop2
Pop1
0.7
0.8
0.9
1.0
1.1
1.2
1.3


EF
Pop4
Pop3
Pop2
Pop1
55
60
65
70
75
80


FS
Pop4
Pop3
Pop2
Pop1
27.5
30.0
32.5
35.0
37.5
40.0
42.5
45.0
47.5


EW
Pop4
Pop3
Pop2
Pop1
0.75
0.80
0.85
0.90
0.95


AW
Pop4
Pop3
Pop2
Pop1
0.56
0.58
0.60
0.62
0.64
0.66
0.68
0.70
0.72

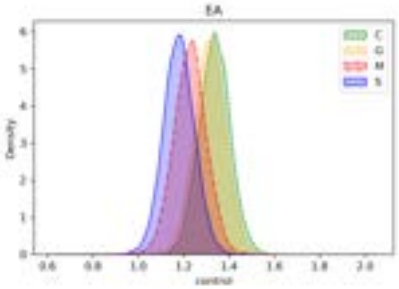

E/A
Pop4
Pop3
Pop2
Pop1
1.1
1.2
1.3
1.4
1.5

4.7.3 Prior sensitivity to hyperpriorparameters

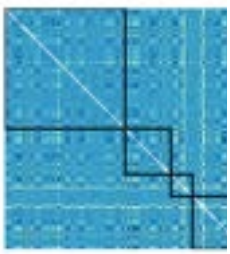
Here we verify the robustness of the model w.r.t. different specifications of the hyperparameters. We consider two alternative specifications for the hyperparameters, which differ from the one already used in the previous section, which are **Prior specification 1:** $G_m = \text{N}(0,\, 1)$; $P_{0,m} \equiv \text{NIG}(\mu = 0,\, \tau = 0.01,\, \alpha = 3,\, \beta = 3)$; all concentration parameters have prior equal to $\text{Gamma}(3,\, 3)$; **Prior specification 2:** $G_m = \text{N}(0,\, 2)$; $P_{0,m} \equiv \text{NIG}(\mu = 0,\, \tau = 1,\, \alpha = 2,\, \beta = 4)$; all concentration parameters have prior equal to $\text{Gamma}(0.1,\, 0.1)$. The model turns out to be rather robust w.r.t. the choice of the hyperparameters, leading to the same selected models for all cardiac indexes under all considered specifications. The detailed results are in the following tables and figures, which report the posterior over partitions of locations, the density estimates, and the posterior similarity matrices for the last cardiac index.

Table 4.9: Posterior probabilities over partitions of means, using prior specification 1. Maximum a posteriori probabilities are in **bold**.

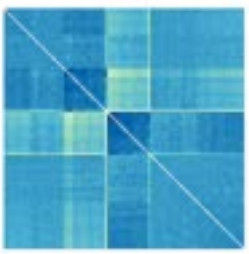
partitions	CI	CWI	LVMI	IVST	LVPW	EF	FS	EW	AW	E/A
{C,G,M,S}	0.018	0.000	0.000	0.000	0.000	**0.371**	**0.276**	0.100	0.000	0.000
{C}{G,M,S}	0.002	**0.526**	0.001	0.086	0.015	0.068	0.207	0.025	0.028	0.000
{C,G}{M,S}	0.002	0.000	0.000	0.000	0.000	0.038	0.035	0.058	0.072	0.045
{C,M,S}{G}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M,S}	0.001	0.139	0.000	0.021	0.023	0.025	0.087	0.034	0.244	0.054
{C,G,M}{S}	**0.436**	0.000	**0.612**	0.000	0.000	0.279	0.04	**0.499**	0.007	0.001
{C,S}{G,M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,M}{S}	0.157	0.100	0.180	**0.542**	**0.678**	0.073	0.172	0.103	0.265	0.026
{C,M}{G,S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G}{M}{S}	0.252	0.000	0.092	0.000	0.000	0.081	0.054	0.113	0.087	0.361
{C,M}{G}{S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,S}{G}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M}{S}	0.131	0.234	0.116	0.351	0.284	0.066	0.130	0.068	**0.295**	**0.513**



(a) density estimation



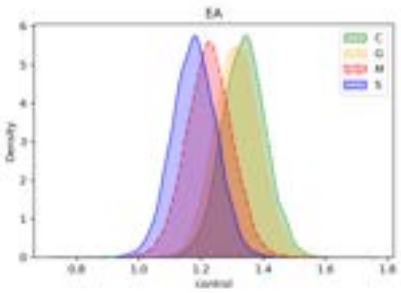
(b) co-clustering



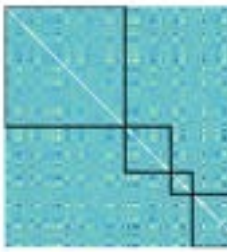
(c) co-clustering

Table 4.10: Posterior probabilities over partitions of means, using prior specification 2. Maximum a posteriori probabilities are in **bold**.

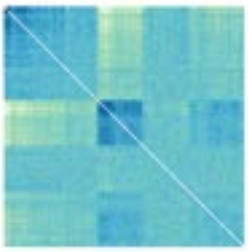
partitions	CI	CWI	LVMI	IVST	LVPW	EF	FS	EW	AW	E/A
{C,G,M,S}	0.023	0.000	0.000	0.000	0.000	**0.341**	**0.281**	0.109	0.000	0.000
{C}{G,M,S}	0.002	**0.484**	0.000	0.097	0.055	0.079	0.199	0.028	0.029	0.000
{C,G}{M,S}	0.002	0.000	0.001	0.000	0.000	0.036	0.029	0.042	0.074	0.058
{C,M,S}{G}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M,S}	0.001	0.134	0.001	0.022	0.028	0.033	0.090	0.033	0.238	0.068
{C,G,M}{S}	**0.408**	0.000	**0.585**	0.000	0.000	0.254	0.036	**0.494**	0.014	0.001
{C,S}{G,M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,M}{S}	0.145	0.111	0.184	**0.530**	**0.643**	0.077	0.172	0.105	0.254	0.019
{C,M}{G,S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G}{M}{S}	0.247	0.000	0.097	0.000	0.000	0.089	0.050	0.106	0.076	0.346
{C,M}{G}{S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,S}{G}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M}{S}	0.172	0.270	0.131	0.351	0.274	0.091	0.144	0.084	**0.315**	**0.508**



(a) density estimation



(b) co-clustering



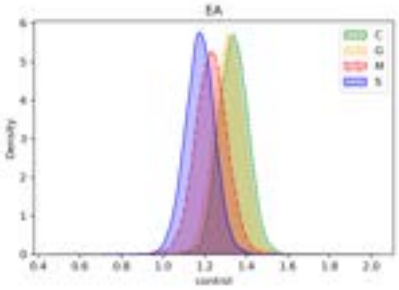
(c) co-clustering

4.7.4 s-HDP with uniform prior estimates on the Hypertensive Dataset

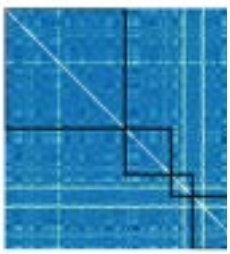
Here we report the results on the real dataset, obtained with the s-HDP with independent uniform priors on disease-specific locations, described in Section 4.6. This prior induces independence between different cardiac indexes and no borrowing of information (i.e. penalization for multiplicity) is applied. Moreover, compared to the priors used in Section 4.7.2, here the prior associates higher probability to finer partitions and, thus, does not apply a Ockham's-razor penalty, resulting in a different MAP for the EF.

Table 4.11: Posterior probabilities over partitions obtained through independent uniform priors. Maximum a posteriori probabilities are in **bold**.

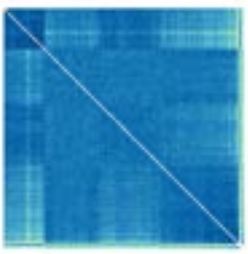
partitions	CI	CWI	LVMI	IVST	LVPW	EF	FS	EW	AW	E/A
{C,G,M,S}	0.009	0.000	0.000	0.000	0.000	0.248	**0.216**	0.047	0.000	0.000
{C}{G,M,S}	0.002	**0.568**	0.001	0.084	0.014	0.078	0.205	0.027	0.039	0.000
{C,G}{M,S}	0.003	0.000	0.002	0.000	0.000	0.082	0.079	0.160	0.102	0.055
{C,M,S}{G}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M,S}	0.001	0.143	0.001	0.024	0.029	0.027	0.087	0.041	0.262	0.064
{C,G,M}{S}	**0.376**	0.000	**0.555**	0.000	0.000	**0.324**	0.060	**0.422**	0.005	0.002
{C,S}{G,M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,M}{S}	0.157	0.115	0.188	**0.614**	**0.730**	0.078	0.189	0.096	**0.304**	0.045
{C,M}{G,S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G,S}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,G}{M}{S}	0.353	0.000	0.173	0.000	0.000	0.125	0.088	0.162	0.087	0.378
{C,M}{G}{S}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C,S}{G}{M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
{C}{G}{M}{S}	0.099	0.174	0.079	0.278	0.227	0.039	0.077	0.045	0.201	**0.457**
$\sum \log_{15}\left(p_i^{-p_i}\right)$	0.493	0.426	0.432	0.352	0.269	0.664	0.725	0.624	0.603	0.448



(a) density estimation



(b) co-clustering



(c) co-clustering

4.7.5 NDP estimates on the Hypertensive Dataset

Here we report the results obtained with ten independent NDPs used on the real dataset. As expected, the NDP tends to identify coarser partitions. Moreover, the independence between cardiac indexes of the NDP approach leads to more concentrated posterior probabilities, because no borrowing of information (i.e. penalization for multiplicity) is applied.

Table 4.12: Posterior probabilities over partitions obtained through independent NDPs. Maximum a posteriori probabilities are in **bold**.

partitions	CI	CWI	LVMI	IVST	LVPW	EF	FS	EW	AW	E/A
{C,G,M,S}	0.117	0.000	0.000	0.000	0.000	**0.613**	**0.394**	0.116	0.000	0.000
{C}{G,M,S}	0.004	**0.999**	0.001	**0.696**	**0.663**	0.047	0.099	0.049	0.313	0.000
{C,G}{M,S}	0.010	0.000	0.014	0.000	0.001	0.027	0.035	**0.206**	0.043	**0.768**
{C,M,S}{G}	0.013	0.000	0.000	0.000	0.000	0.040	0.067	0.051	0.000	0.000
{C}{G}{M,S}	0.001	0.000	0.001	0.013	0.163	0.005	0.015	0.088	**0.468**	0.013
{C,G,M}{S}	**0.552**	0.000	**0.906**	0.000	0.000	0.103	0.091	0.154	0.002	0.000
{C,S}{G,M}	0.070	0.000	0.000	0.000	0.000	0.025	0.069	0.029	0.000	0.000
{C}{G,M}{S}	0.077	0.001	0.010	0.207	0.136	0.010	0.032	0.050	0.093	0.006
{C,M}{G,S}	0.009	0.000	0.003	0.023	0.000	0.035	0.045	0.017	0.001	0.000
{C,G,S}{M}	0.047	0.000	0.000	0.000	0.000	0.068	0.081	0.073	0.000	0.000
{C}{G,S}{M}	0.003	0.001	0.000	0.052	0.027	0.007	0.022	0.012	0.030	0.000
{C,G}{M}{S}	0.065	0.000	0.047	0.000	0.000	0.011	0.016	0.071	0.007	0.208
{C,M}{G}{S}	0.025	0.000	0.017	0.004	0.000	0.007	0.017	0.033	0.001	0.000
{C,S}{G}{M}	0.006	0.000	0.000	0.000	0.000	0.006	0.014	0.023	0.000	0.000
{C}{G}{M}{S}	0.007	0.000	0.002	0.007	0.012	0.001	0.007	0.032	0.044	0.006
$\sum \log_{15}\left(p_i^{-p_i}\right)$	0.603	0.016	0.167	0.349	0.368	0.567	0.785	0.898	0.509	0.239

4.8 Concluding remarks

We designed a Bayesian nonparametric model to detect clusters of hypertensive disorders over different cardiac function indexes and found modified cardiac functions in hypertensive patients compared to healthy subjects as well as progressively increased alterations with the severity of the disorder. The proposed model has application potential also beyond the considered setup when the goal is to cluster populations according to multivariate information: it borrows strength across response variables, preserves the flexibility intrinsic to nonparametric models, and correctly detects partitions of populations even in presence of small sample sizes, when alternative distribution-based clustering models tends to underestimate the number of clusters. The key component of the model is the s-HDP, a hierarchical nonparametric structure for the error terms that offers flexibility and serves as a tool to investigate the presence of unobserved factors, outliers and effects other than changes in locations. Interesting extensions of the model include generalizations to other types of invariances in order to accommodate identifiability in generalized linear models, for instance in presence of count data and a log link function, as well as generalizations to other types of processes, beyond the Dirichlet process.

Chapter 5

Dependent Prior Processes for Panel Count Data

This chapter shows how dependent processes may be effectively used to make inference on panel count data, which are data concerning recurrent events experienced by subjects that cannot be observed in continuous time. Contrary to what has been done in all previous chapters, here dependent processes are used for data that are not partially exchangeable. However, as in all the chapters of this thesis, the dependence between the processes plays a crucial role in the definition of the appropriate Bayesian learning mechanism, as explained in Section 5.1. After presenting the model, we investigate prior and posterior distributional properties (Section 5.2 and 5.3), develop a MCMC algorithm to perform posterior inference (Section 5.4), and test the performance of our proposal in a simulation studies (Section 5.5).

5.1 Dependence in panel count data

Panel count data occur in observational studies and clinical trials that concern recurrent events, e.g., tumor, infection, or asthma attacks recurrences, where for each subject cumulative counts are recorded at discrete time points (see Thall & Lachin, 1988; Balshaw & Dean, 2002; Sun, 2013). More formally, the observed data are such that for each subject i, from a sample of n subjects, we record

- a collection of discrete time points $(t_{i,1}, \dots t_{i,m_i})$ at which the i-th subject has been observed, where $0 \leq t_{i,1} < \dots < t_{i,m_i}$ and m_i is the total number of points,
- the cumulative counts $N_{i,j}$ of the recurrent event of interest, experienced by the i-th subject up to time $t_{i,j}$, for $j = 1, \dots, m_i$,
- the follow up time C_i, which in this chapter we assume to be a fixed, non-informative time point.

For example, $t_{i,j}$ may be the time of the j-th clinical visit for patient i, $N_{i,j}$ the number of recurrent tumors between $t = 0$ and $t_{i,j}$, again for patient i, and C_i is the time at which the study ends for subject i. Both the time points and the cumulative counts are considered as realizations of point processes, namely the *observation process* and the *event process* and, typically, the main inferential goal is to estimate the mean function of the event process. We denote with $N_i = \{N_i(t) : t > 0\}$ the event process for subject i, i.e., $N_i(t)$ is the number of events experienced by subject i up to time t, and with $T_i = \{T_i(t) : t > 0\}$, the observational process for subject i, i.e., $T_i(t)$ is the number of times when the i-th subject has been observed up to time t. Contrary to what happens with standard recurrent event data, in panel count data the subject is not observed continuously and therefore the analyst has only a partial observation of the realized path of the process $N_i(t)$. Thus, $N_{i,j} = N_i(t_{i,j})$ and the inferential target is $\mathbb{E}[N_i(t)]$.

Standard models for panel count data assume that the observation process is not informative and independent from the event process (cf. Sun, 2013). Even though assuming independence between the two processes simplifies the inferential procedure, the assumption is not realistic in many applications. A typically example of dependence is when, in observational studies, those subjects that experience the event less often, tend to skip clinical visits, decreasing the rates of their observation processes. Recently there has been an increasing literature whose aim is to model *subject-level* dependence between the frequency of observation and the number of events (see, for instance, He et al., 2009; Liang et al., 2018). In simple words, subject-level dependence acts between N_i and T_i, but not between N_i and T_j. Modeling subject-level dependence is important because it allows to correct potential distortions in the mean functions estimates and can be usually done using frailties. However, a proper Bayesian model for panel count data should incorporate the information about dependence also in the learning mechanism, ultimately incorporating dependence between the two processes also at *populations-level*. Easily: if knowledge about positive association between the two processes is available a priori and, in a certain sample, we observe high frequency of observational points (higher than our prior guess), we should increase our guess for the event mean functions not only of the observed subjects but also of a new subject. However, we are not aware of any proposed statistical model for panel count data that induces such dependence at the population level and automatically incorporates the information provided by the observational points in the prediction of the event mean function of a new subject. We propose a class of Bayesian nonparametric priors over the observation and the event processes that allow for population-level dependence, incorporating prior information regarding positive association between frequency of observation and counts. The two processes are assumed to be inhomogeneous Poisson processes with random intensities. The priors are defined through mixtures with respect to CRMs, such that dependence across CRMs induces dependence between the two processes. Covariates and subject-specific frailties can be included in the model through a Cox regression (see Section 6.4).

5.2 The model

5.2.1 Cox processes with dependent mixture intensities

We exploit the common assumption that the process generating the recurrent events, namely the event process, being an inhomogeneous Poisson process (PP), with intensity function $\lambda_N(t)$ and we write

$$\{N_i(t) \,:\, t > 0\} \mid \lambda_N(t) \overset{iid}{\sim} \text{PP}(\, \lambda_N(t)\,)$$

In addition, we assume that also the observation process, from which the time points $t_{i,1}, \dots t_{i,m_i}$ are generated, is an inhomogeneous PP, with intensity function $\lambda_T(t)$

$$\{T_i(t) \,:\, t > 0\} \mid \lambda_T(t) \overset{iid}{\sim} \text{PP}(\, \lambda_T(t)\,)$$

Priors on $\lambda_T(t)$ and $\lambda_N(t)$ are then defined through mixtures with respect to CRMs

$$\begin{aligned} \lambda_T(t) &= \int_{\mathbb{Y}} k_T(t;\, y)\tilde{\mu}_T(dy) \\ \lambda_N(t) &= \int_{\mathbb{Y}} k_N(t;\, y)\tilde{\mu}_N(dy) \\ (\tilde{\mu}_T, \tilde{\mu}_N) &\sim \mathcal{Q} \end{aligned} \tag{5.1}$$

where $\tilde{\mu}_T$ and $\tilde{\mu}_N$ are CRMs defined on a measurable space $(\mathbb{Y}, \mathcal{Y})$; $k_l(\cdot;\cdot)$, for $l \in \{T,\, N\}$, are transition kernels on $\mathbb{R}^+ \times \mathbb{Y}$ and $\mathcal{Q}$ is the joint probability measure induced by $\tilde{\mu}_T$ and $\tilde{\mu}_N$. Similar CRMs mixture specifications have been used for hazard rates in survival models in Dykstra & Laud (1981) and Lo & Weng (1989) and more recently in Lijoi & Nipoti (2014) and Arbel et al. (2016) (cf. Section 1.3.2). Different choices for $\mathcal{Q}$ give rise to alternative joint prior distributions over the observation and event process and controls the dependence between the two. In particular, in this chapter we consider three different specifications for $\mathcal{Q}$: independent CRMs, GM-dependent CRMs (Lijoi et al., 2014a,b) and hierarchical CRMs (Camerlenghi et al., 2019b). Details on GM-dependent CRMs and hierarchical CRMs can been found in Section 1.4.1 and 1.4.2, while the independent case is simply obtained assuming that the joint probability measure can be factorized: $\mathcal{Q} = \mathcal{Q}_{\tilde{\mu}_T} \otimes \mathcal{Q}_{\tilde{\mu}_N}$. Clearly, when this happens, the information regarding the dependence between the observation process and the event process is ignored by the learning mechanism. We will use the independent case as benchmark in the simulation study. Notice that the model can be extended to include subject-level covariates and frailties using a Cox regression model for the intensities, such that

$$\{N_i(t) \,:\, t > 0\} \mid \lambda_{N,i}(t) \overset{ind}{\sim} \text{PP}(\, \lambda_{N,i}(t)\,)$$

$$\{T_i(t) \,:\, t > 0\} \mid \lambda_{T,i}(t) \overset{ind}{\sim} \text{PP}(\, \lambda_{T,i}(t)\,)$$

$$\lambda_{l,i}(t; \boldsymbol{x}_i) = \lambda_{0,l}(t) \exp\{\boldsymbol{x}_i' \boldsymbol{\beta}_l + \epsilon_{l,i}\} \qquad \text{for } l \in \{T, N\}$$

where the baseline intensity functions $\lambda_{0,T}(t)$ and $\lambda_{0,N}(t)$ are distributed according to (5.1) and $\boldsymbol{x}_i$ is a vector of covariates. The model is then completed choosing appropriate prior distributions over $\boldsymbol{\beta}_l$ and $\epsilon_{l,i}$. Notice, moreover, that the use of frailties $\epsilon_{l,i}$ permits to incorporate subject-level dependence. As already mentioned, such dependence is important from a modeling point of view, however is not the core of our proposal and can be simply managed using a joint prior distribution on $\epsilon_{T,i}$ and $\epsilon_{N,i}$. Therefore for sake of clarity, in this chapter, we consider the simple model without frailties and covariates, and then we extended the model in Section 6.4. We conclude this section providing expression for the first and second a priori marginal moments induced on the intensities and on the processes by equation (5.1). For sake of brevity, we report here below only results for the observation process, since the ones for the event process are analogous. We believe the two following results being a useful tool for the choice of the kernel functions and the intensities of the CRMs, during prior elicitation.

Proposition 5.1. *The prior expected value of the observation process and of the observation intensity at any time point $t \in \mathbb{R}^+$ are*

$$\mathbb{E}[\, T_i(t)\,] = \int_{\mathbb{R}^+ \times \mathbb{Y}} \left(\int_0^t k_T(t'; y) dt' \right) s\, v(ds, dy)$$

$$\mathbb{E}[\, \lambda_T(t)\,] = \int_{\mathbb{R}^+ \times \mathbb{Y}} k_T(t;\, y)\, s\, v(\mathrm{d}s, \mathrm{d}y)$$

where v is the Lévy intensity of $\tilde{\mu}_T$.

Note that both moments are finite if and only if $\int_{\mathbb{R}}^+ sv(\mathrm{d}s, \mathrm{d}y) < +\infty$.

Proposition 5.2. *The prior variance of the observation process and of the observation intensity at any time point $t \in \mathbb{R}^+$ are*

$$\mathit{Var}[\, T_i(t)\,] = \int_{\mathbb{R}^+ \times \mathbb{Y}} \left(\int_0^t k_T(t'; y) dt' \right) s^2\, v(\mathrm{d}s, \mathrm{d}y)$$

$$\mathit{Var}[\, \lambda_T(t)\,] = \int_{\mathbb{R}^+ \times \mathbb{Y}} k_T(t;\, y)\, s^2\, v(\mathrm{d}s, \mathrm{d}y)$$

where v is the Lévy intensity of $\tilde{\mu}_T$.

Note that both moments are finite if and only if $\int_{\mathbb{R}}^+ s^2 v(\mathrm{d}s, \mathrm{d}y) < +\infty$. The proofs of both propositions trivially follow from results in Appendix B.

Example 5.1. *Il $\tilde{\mu}_T$ is a gamma GM-dependent CRM and the kernel is of OU type (see Section 1.3.2), i.e. $k_T(t;y) = 2ke^{-k(t-y)}\mathbb{1}_{\{t\geq y\}}$, then*

$$\mathbb{E}[\,T_i(t)\,] = Var[\,T_i(t)\,] = c \int\limits_{\mathbb{Y}\cap\{y\,:\,y\leq t\}} 2\,(1-e^{-k(t-y)})P_0(dy)$$

$$\mathbb{E}[\,\lambda_T(t)\,] = Var[\,\lambda_T(t)\,] = c \int\limits_{\mathbb{Y}\cap\{y\,:\,y\leq t\}} 2\,k\,e^{-k(t-y)}P_0(dy)$$

where c is the concentration parameter and P_0 is the base measure of $\tilde{\mu}_T$.

5.2.2 Prior correlation between observational and event processes

The main advantage of the specification in (5.1) is the possibility to model population-level dependence between the two processes that give rise to panel count data. Such dependence is clearly ignored if the two mixing CRMs are independent. In this section we provide the a priori correlation structure between the two processes, induced by the two alternative dependent priors, respectively GM-dependent and hierarchical CRMs.

Proposition 5.3. *The pairwise a priori covariance between the two processes at two time points $t_1,\ t_2 \in \mathbb{R}^+$ is as follows.*

(i) If $(\tilde{\mu}_T, \tilde{\mu}_N) \stackrel{d}{=}$ GM-dependent CRM, then

$$Cov(T_i(t_1), N_j(t_2)) = \int\limits_{\mathbb{R}^+\times\mathbb{Y}} \left(\int\limits_0^{t_1} k_T(t;y)dt\right)\left(\int\limits_0^{t_2} k_N(t;y)dt\right) s^2\, v_0(ds,dy) \geq 0$$

where v_0 is the intensity of μ_0, the common component of $\tilde{\mu}_T$ and $\tilde{\mu}_N$ (cf. Section 1.4.1).

(ii) If $(\tilde{\mu}_T, \tilde{\mu}_N)$ are CRMs with hierarchical structure, i.e.,

$$\begin{aligned} \tilde{\mu}_l \mid \tilde{\mu}_0 &\stackrel{ind}{\sim} CRM(\tilde{v}_l) \qquad \text{with} \quad \tilde{v}_l(\mathrm{d}s,\mathrm{d}y) = \rho_l(s)\,\mathrm{d}s\,\tilde{\mu}_0(\mathrm{d}y) \quad \text{and} \quad l \in \{T,N\} \\ \tilde{\mu}_0 &\sim CRM(\tilde{v}_0) \end{aligned}$$

then

$$\begin{aligned} Cov(T_i(t_1), N_j(t_2)) = &\int\limits_{\mathbb{R}^+} s\,\rho_T(s)ds \int\limits_{\mathbb{R}^+} s\,\rho_N(s)ds \times \\ &\times \int\limits_{\mathbb{R}^+\times\mathbb{Y}} \int\limits_0^{t_1} k_T(t_1;y) \int\limits_0^{t_2} k_N(t_2;y)\, s^2\, \tilde{v}_0(ds,dy) \geq 0 \end{aligned}$$

Proof. First of all, note that

$$\begin{aligned} Cov(T_i(t_1),\, N_j(t_2)) = \mathbb{E}[\, Cov(\, T_i(t_1),\, N_j(t_2) \mid \tilde{\mu}_T, \tilde{\mu}_N)\,] \\ + Cov(\, \mathbb{E}[\, T_i(t_1) \mid \tilde{\mu}_T], \mathbb{E}[\, N_j(t_2) \mid \tilde{\mu}_N]\,) \end{aligned}$$

Where the first term equal zero, because the two processes are conditionally independent

$$Cov(\, T_i(t_1),\, N_j(t_2) \mid \tilde{\mu}_T, \tilde{\mu}_N) = 0$$

Let us compute the second term

$$\begin{aligned} &Cov(\, \mathbb{E}[\, T_i(t_1) \mid \tilde{\mu}_T], \mathbb{E}[\, N_j(t_2) \mid \tilde{\mu}_N]\,) = \\ &\quad = \mathbb{E}\Big[\mathbb{E}[\, T_i(t_1) \mid \tilde{\mu}_T] \cdot \mathbb{E}[\, N_j(t_2) \mid \tilde{\mu}_N]\Big] - \mathbb{E}\Big[\mathbb{E}[\, T_i(t_1) \mid \tilde{\mu}_T]\Big] \cdot \mathbb{E}\Big[\mathbb{E}[\, N_j(t_2) \mid \tilde{\mu}_N]\Big] = \\ &\quad = \mathbb{E}\left[\int_0^{t_1} \lambda_T(x)dx \int_0^{t_2} \lambda_N(x)dx\right] - \mathbb{E}\left[\int_0^{t_1} \lambda_T(x)dx\right] \mathbb{E}\left[\int_0^{t_2} \lambda_N(x)dx\right] \end{aligned} \tag{5.2}$$

substituting the expression of the intensities, we get

$$\begin{aligned} &Cov(\, \mathbb{E}[\, T_i(t_1) \,|\, \tilde{\mu}_T], \mathbb{E}[\, N_j(t_2) \mid \tilde{\mu}_N]\,) = \\ &\qquad = \mathbb{E}\Big[\Big(\int_0^{t_1}\int_{\mathbb{Y}} k_T(x;y)\tilde{\mu}_T(dy)dx\Big)\Big(\int_0^{t_2}\int_{\mathbb{Y}} k_N(x;y)\tilde{\mu}_N(dy)dx\Big)\Big] + \\ &\qquad - \mathbb{E}\Big[\Big(\int_0^{t_1}\int_{\mathbb{Y}} k_T(x;y)\tilde{\mu}_T(dy)dx\Big)\Big] \cdot \mathbb{E}\Big[\Big(\int_0^{t_2}\int_{\mathbb{Y}} k_N(x;y)\tilde{\mu}_N(dy)dx\Big)\Big] \end{aligned} \tag{5.3}$$

Moreover if $(\tilde{\mu}_T, \tilde{\mu}_N)$ are GM-dependent CRM, (5.3) simplifies to

$$\begin{aligned} &\mathbb{E}\Big[\Big(\int_{\mathbb{Y}}\int_0^{t_1} k_T(x;y)dx\,\mu_0(dy)\Big)\Big(\int_{\mathbb{Y}}\int_0^{t_2} k_N(x;y)dx\,\mu_0(dy)\Big)\Big] + \\ &-\mathbb{E}\Big[\Big(\int_{\mathbb{Y}}\int_0^{t_1} k_T(x;y)dx\,\mu_0(dy)\Big)\Big]\mathbb{E}\Big[\Big(\int_{\mathbb{Y}}\int_0^{t_2} k_N(x;y)dx\,\mu_0(dy)\Big)\Big] \end{aligned}$$

which depends on the common component μ_0 only and, using the results in Appendix B, we have

$$\int_{\mathbb{R}^+\times\mathbb{Y}} \Big(\int_0^{t_1} k_T(x;y)dx\Big)\Big(\int_0^{t_2} k_N(x;y)dx\Big)\, s^2\, v_0(ds,dy)$$

proving point (i) of the theorem.

While if $(\tilde{\mu}_T, \tilde{\mu}_N)$ are CRMs with hierarchical structure, according to results in Appendix B, for $l \in \{T, N\}$, we have

$$\begin{aligned}\mathbb{E}\left[\int_0^t \lambda_l(x)dx\right] &= \mathbb{E}\left[\mathbb{E}\left[\int_{\mathbb{Y}}\int_0^t k_l(x;y)dx\,\tilde{\mu}_l(dy) \mid \tilde{\mu}_0\right]\right] = \mathbb{E}\left[\int_{\mathbb{R}^+\times\mathbb{Y}}\int_0^t k_l(x;y)dx\,s\,\tilde{v}_l(ds,dy)\right] = \\ &= \int_{\mathbb{R}^+} s\,\rho_l(s)ds \int_{\mathbb{R}^+\times\mathbb{Y}}\int_0^t k_l(x;y)dx\,s\,\tilde{v}_0(ds,dy)\end{aligned} \tag{5.4}$$

Moreover

$$\begin{aligned}&\mathbb{E}\left[\int_0^{t_1}\lambda_T(x)dx\int_0^{t_2}\lambda_N(x)dx\right] = \mathbb{E}\left[\mathbb{E}\left[\int_{\mathbb{Y}}\int_0^{t_1}k_T(x;y)dx\,\tilde{\mu}_T(dy)\int_{\mathbb{Y}}\int_0^{t_2}k_N(x;y)dx\,\tilde{\mu}_N(dy) \mid \tilde{\mu}_0\right]\right] = \\ &= \int_{\mathbb{R}^+} s\,\rho_T(s)ds \int_{\mathbb{R}^+} s\,\rho_N(s)ds \left(\int_{\mathbb{R}^+\times\mathbb{Y}}\int_0^{t_1}k_T(t_1;y)\int_0^{t_2}k_N(t_2;y)\,s^2\,\tilde{v}_0(ds,dy) + \right. \\ &\left. + \int_{\mathbb{R}^+\times\mathbb{Y}}\int_0^{t_1}k_T(t_1;y)\,s\,\tilde{v}_0(ds,dy)\int_{\mathbb{R}^+\times\mathbb{Y}}\int_0^{t_2}k_N(t_2;y)\,s\,\tilde{v}_0(ds,dy)\right)\end{aligned} \tag{5.5}$$

Plugging (5.4) and (5.5) into (5.2) proves point (ii) of the theorem. □

Example 5.2. *Il $\tilde{\mu}_N$ and $\tilde{\mu}_T$ are gamma GM-dependent CRMs and the two kernels are of OU type, i.e., $k_T(t;y) = k_N(t;y) = 2ke^{-k(t-y)}\mathbb{1}_{\{t\geq y\}}$, then*

$$Cov(T_i(t_1), N_j(t_2)) = \int_{\mathbb{Y}\cap\{y\,:\,y\leq\min\{t_1,t_2\}\}} c\,4\,(1-e^{-k(t_1-y)})\,(1-e^{-k(t_2-y)})P_0(dy)$$

5.3 Posterior characterization

In this section we provide a posterior characterisation of the vector of CRMs and of the intensity functions. The results are conditional to some appropriate auxiliary random variables both for the GM-dependent and the hierarchical priors and they can be used to make inference through the MCMC algorithm in the next section. For sake of exposition in the remaining of this chapter we assume equal marginals for the two intensities and, in particular, $k_T(\cdot;\cdot) = k_N(\cdot;\cdot) = k(\cdot;\cdot)$. It is important to stress that such assumption may often be unrealistic in real applications, however extensions beyond this assumption can easily be constructed, as done for instance later in Section 6.4 of the next chapter using multiplicative terms for the two intensities. When this strategy will be applied the results derived here are extended straightforwardly.

5.3.1 Likelihood

We denote the collection of the observation points as $\boldsymbol{t} = \{t_{i,j} \,:\, j = 1, \dots m_i,\, i = 1, \dots n\}$. The marginal likelihood function for the observation times is obtained as product across i of the likelihood of the observed event times in the time interval $[0, C_i)$ for the inhomogeneous PPs $T_i(t)$:

$$\mathcal{L}(\tilde{\mu}_T\,; \boldsymbol{t}) = \prod_{i=1}^{n} \left\{ e^{-\int_0^{C_i} \int_{\mathbb{Y}} k(t;y)\tilde{\mu}_T(dy)dt} \prod_{j=1}^{m_i} \int_{\mathbb{Y}} k(t_{i,j};y)\tilde{\mu}_T(dy) \right\}$$

We define $x_{i,j} = N_{i,j} - N_{i,j-1}$, $\forall j$ and $\forall i$, and denote the collection of incremental counts as $\boldsymbol{x} = \{x_{i,j} \,:\, j = 1, \dots m_i,\, i = 1, \dots n\}$. The incremental counts are conditionally independent realizations of Poisson random variables:

$$x_{i,j} \mid \boldsymbol{t}, \tilde{\mu}_N \overset{ind}{\sim} \text{Poisson}\left(\int_{t_{i,j-1}}^{t_{i,j}} \int_{\mathbb{Y}} k(t;\, y)\tilde{\mu}_N(dy)dt \right)$$

So that the likelihood function for the counts, conditional on the observation points, is

$$\mathcal{L}(\tilde{\mu}_N\,; \boldsymbol{x} \mid \boldsymbol{t}) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left[e^{-\int_{t_{i,j-1}}^{t_{i,j}} \int_{\mathbb{Y}} k(t;y)\tilde{\mu}_N(dy)dt} \times \frac{1}{x_{i,j}!} \left(\int_{t_{i,j-1}}^{t_{i,j}} \int_{\mathbb{Y}} k(t;\, y)\tilde{\mu}_N(dy)dt \right)^{x_{i,j}} \right]$$

Therefore, the joint likelihood can be rewritten as

$$\begin{aligned} \mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}) = e^{-\int_{\mathbb{Y}} K_T(y)\tilde{\mu}_T(dy)} e^{-\int_{\mathbb{Y}} K_N(y)\tilde{\mu}_N(dy)} \times \\ \times \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left[\int_{\mathbb{Y}} k(t_{i,j};y)\tilde{\mu}_T(dy) \frac{1}{x_{i,j}!} \left(\int_{\mathbb{Y}} H_{i,j}(y)\tilde{\mu}_N(dy) \right)^{x_{i,j}} \right] \end{aligned} \tag{5.6}$$

where $K_T(y) = \sum_{i=1}^{n} \int_0^{C_i} k(t;\, y)dt$, $K_N(y) = \sum_{i=1}^{n} \int_0^{t_{i,m_i}} k(t;\, y)dt$ and $H_{i,j}(y) = \int_{t_{i,j-1}}^{t_{i,j}} k(t;\, y)dt$.

We introduce some latent random variables $\boldsymbol{Y}_l = \{Y_{i,j,l} \,:\, j = 1, \dots, m_i\,,\, i = 1, \dots, n\}$ for $l \in \{T, N\}$ to simplify the expression in (5.6) removing the integrals, such that the joint law of $(\boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_T, \boldsymbol{Y}_N)$, conditionally on $\tilde{\mu}_T, \tilde{\mu}_N$ is given by

$$\begin{aligned} \mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) = e^{-\int_{\mathbb{Y}} K_T(y)\tilde{\mu}_T(dy)} e^{-\int_{\mathbb{Y}} K_N(y)\tilde{\mu}_N(dy)} \times \\ \times \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left[k(t_{i,j};y_{i,j,T})\tilde{\mu}_T(dy_{i,j,T}) \frac{1}{x_{i,j}!} \left(H_{i,j}(y_{i,j,N})\tilde{\mu}_N(dy_{i,j,N}) \right)^{x_{i,j}} \right] \end{aligned} \tag{5.7}$$

5.3.2 GM-dependent CRMs posterior law

By discreteness of $\tilde{\mu}_T$ and $\tilde{\mu}_N$, there will be ties between the auxiliary variables $Y_{i,j,l}$'s. We denote the distinct values in $\boldsymbol{y}_T$ and $\boldsymbol{y}_N$ respectively as

$$\{y^*_{1,T}, \dots, y^*_{k_T,T}, y^*_1, \dots, y^*_k\} \qquad \{y^*_{1,N}, \dots, y^*_{k_N,N}, y^*_1, \dots, y^*_k\}$$

with $\{y^*_{1,T}, \dots, y^*_{k_T,T}\} \cap \{y^*_{1,N}, \dots, y^*_{k_N,N}\} = \emptyset$.
We define the frequencies of the first collection $\boldsymbol{y}_T$ as

$$n_h = \sum_i \sum_j \mathbb{1}(y_{i,j,T} = y^*_{h,T}) \qquad \text{and} \qquad q_m = \sum_i \sum_j \mathbb{1}(y_{i,j,T} = y^*_m)$$

and the counts corresponding to the second collection $\boldsymbol{y}_N$ as

$$x'_r = \sum_{\substack{(i,j):\\ y_{i,j,N} = y^*_{r,N}}} x_{i,j} \qquad \text{and} \qquad x''_m = \sum_{\substack{(i,j):\\ y_{i,j,N} = y^*_m}} x_{i,j}$$

Moreover, let us introduce two additional independent sequences of i.i.d auxiliary random variables $\boldsymbol{V}_T = (V_{h,T})_{h=1}^{k_T}$ and $\boldsymbol{V}_N = (V_{r,N})_{r=1}^{k_N}$ such that

$$\mathbb{P}[V_{i,l} = 0] = 1 - \mathbb{P}[V_{i,l} = 1] = z \qquad \text{for } \forall i = 1, \dots, k_l \quad \text{and} \quad l \in T, N$$

Theorem 5.1. *Let $Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} k(t_{i,j}; y_{i,j,T}) \frac{1}{x_{i,j}!} H_{i,j}(y_{i,j,N})^{x_{i,j}}$, the probability distribution of $(\boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_N}, \boldsymbol{Y_Z})$ conditionally on $\boldsymbol{V_T}$ and $\boldsymbol{V_N}$ equals*

$$\begin{aligned}
\pi(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N} | \boldsymbol{V_T}, \boldsymbol{V_N}) = Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})\ & e^{-c\,\psi_z(K_T, K_N)} (1-z)^k\, c^{k_T + k_N + k} \times \\
& \times \prod_{h=1}^{k_T} P_0(dy^*_{h,T}) \int_{\mathbb{R}^+} s^{n_h}\, e^{-s\,(K_T(y^*_{h,T}) + K_N(y^*_{h,T}) V_{h,T})} \rho(s) ds \times \\
& \times \prod_{r=1}^{k_N} P_0(dy^*_{r,N}) \int_{\mathbb{R}^+} s^{x'_r}\, e^{-s\,(K_T(y^*_{r,N}) V_{r,N} + K_N(y^*_{r,N}))} \rho(s) ds \times \\
& \times \prod_{m=1}^{k} P_0(dy^*_m) \int_{\mathbb{R}^+} s^{q_m + x''_m}\, e^{-s\,(K_T(y^*_m) + K_N(y^*_m))} \rho(s) ds
\end{aligned}$$

Proof. The joint distribution is obtained taking the expected value of the likelihood in (5.7) with respect to the distribution of the vector of CRMs.

$$\pi(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) = \mathbb{E}[\mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})].$$

Define: $\mathbb{Y}^* = \mathbb{Y} \setminus \{dy^*_{1,T}, \dots, dy^*_{k_T,T}, dy^*_{1,N}, \dots, dy^*_{k_N,N}, dy^*_1, \dots, dy^*_k\}$ where $dy = [y, y + \epsilon)$

with $\epsilon > 0$, the likelihood in (5.7) can be rewritten as

$$\begin{aligned}\mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N \,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) \;&= Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})\; e^{-\int_{\mathbb{Y}^*} K_T(y)\tilde{\mu}_T(dy) - \int_{\mathbb{Y}^*} K_N(y)\tilde{\mu}_N(dy)} \times \\ &\times \prod_{h=1}^{k_T} e^{-K_T(y^*_{h,T})\tilde{\mu}_T(dy^*_{h,T}) - K_N(y^*_{h,T})\tilde{\mu}_N(dy^*_{h,T})} \tilde{\mu}_T(dy^*_{h,T})^{n_h} \times \\ &\times \prod_{r=1}^{k_N} e^{-K_T(y^*_{r,N})\tilde{\mu}_T(dy^*_{r,N}) - K_N(y^*_{r,N})\tilde{\mu}_N(dy^*_{r,N})} \tilde{\mu}_N(dy^*_{r,N})^{x'_r} \times \\ &\times \prod_{m=1}^{k} e^{-K_T(y^*_m)\tilde{\mu}_T(dy^*_m) - K_N(y^*_m)\tilde{\mu}_N(dy^*_m)} \tilde{\mu}_T(dy^*_m)^{q_m} \tilde{\mu}_N(dy^*_m)^{x''_m}\end{aligned}$$

For ϵ arbitrarily small, the intervals dy^* are disjoint, thus the expected value can be rewritten as

$$\begin{aligned}\mathbb{E}\left[\mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N \,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})\right] \;&= Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})\; \mathbb{E}\left[e^{-\int_{\mathbb{Y}^*} K_T(y)\tilde{\mu}_T(dy) - \int_{\mathbb{Y}^*} K_N(y)\tilde{\mu}_N(dy)}\right] \times \\ &\times \prod_{h=1}^{k_T} \mathbb{E}\left[e^{-K_T(y^*_{h,T})\tilde{\mu}_T(dy^*_{h,T}) - K_N(y^*_{h,T})\tilde{\mu}_N(dy^*_{h,T})} \tilde{\mu}_T(dy^*_{h,T})^{n_h}\right] \times \\ &\times \prod_{r=1}^{k_N} \mathbb{E}\left[e^{-K_T(y^*_{r,N})\tilde{\mu}_T(dy^*_{r,N}) - K_N(y^*_{r,N})\tilde{\mu}_N(dy^*_{r,N})} \tilde{\mu}_N(dy^*_{r,N})^{x'_r}\right] \times \\ &\times \prod_{m=1}^{k} \mathbb{E}\left[e^{-K_T(y^*_m)\tilde{\mu}_T(dy^*_m) - K_N(y^*_m)\tilde{\mu}_N(dy^*_m)} \tilde{\mu}_T(dy^*_m)^{q_m} \tilde{\mu}_N(dy^*_m)^{x''_m}\right]\end{aligned} \tag{5.8}$$

Notice that the expectations in (5.8) are given by the following four equations

$$\mathbb{E}\left[e^{-\int_{\mathbb{Y}^*} K_T(y)\tilde{\mu}_T(dy) - \int_{\mathbb{Y}^*} K_N(y)\tilde{\mu}_N(dy)}\right] = e^{-c\psi_z(K_T(y)\mathbb{1}_{\mathbb{Y}^*}, K_N(y)\mathbb{1}_{\mathbb{Y}^*})}$$

$$\begin{aligned}&\mathbb{E}\left[e^{-K_T(y^*_{h,T})\tilde{\mu}_T(dy^*_{h,T}) - K_N(y^*_{h,T})\tilde{\mu}_N(dy^*_{h,T})} \tilde{\mu}_T(dy^*_{h,T})^{n_{h,T}}\right] = \\ &= (-1)^{n_h} \frac{\partial^{n_h}}{\partial \gamma^{n_h}} \mathbb{E}\left[e^{-\gamma\tilde{\mu}_T(dy^*_{h,T}) - K_N(y^*_{h,T})\tilde{\mu}_N(dy^*_{h,T})}\right]\Big|_{\gamma = K_T(y^*_{h,T})} = \\ &= (-1)^{n_h} \frac{\partial^{n_h}}{\partial \gamma^{n_h}} e^{-c\psi_z(\gamma\mathbb{1}_{dy^*_{h,T}}, K_N(y)\mathbb{1}_{dy^*_{h,T}})}\Big|_{\gamma = K_T(y^*_{h,T})}\end{aligned} \tag{5.9}$$

$$\begin{aligned}&\mathbb{E}\left[e^{-K_T(y^*_{r,N})\tilde{\mu}_T(dy^*_{r,N})-K_N(y^*_{r,N})\tilde{\mu}_N(dy^*_{r,N})}\tilde{\mu}_N(dy^*_{r,N})^{x'_r}\right]=\\&=(-1)^{x'_r}\frac{\partial^{x'_r}}{\partial\gamma^{x'_r}}e^{-c\psi_z(K_T(y)\mathbb{1}_{dy^*_{r,N}},\gamma\mathbb{1}_{dy^*_{r,N}})}\Big|_{\gamma=K_N(y^*_{r,N})}\end{aligned}\tag{5.10}$$

$$\begin{aligned}&\mathbb{E}\left[e^{-K_T(y^*_m)\tilde{\mu}_T(dy^*_m)-K_N(y^*_m)\tilde{\mu}_N(dy^*_m)}\tilde{\mu}_T(dy^*_m)^{q_m}\tilde{\mu}_N(dy^*_m)^{x''_m}\right]=\\&=(-1)^{q_m+x''_m}\frac{\partial^{q_m+x''_m}}{\partial\gamma_1^{q_m}\partial\gamma_2^{x''_m}}e^{-c\psi_z(\gamma_1\mathbb{1}_{dy^*_m},\gamma_2\mathbb{1}_{dy^*_m})}\Big|_{\substack{\gamma_1=K_T(y^*_m)\\\gamma_2=K_N(y^*_m)}}\end{aligned}\tag{5.11}$$

Applying Faà di Bruno's formula, as shown in Appendix C, equations (5.9)-(5.11) can be rewritten respectively as

$$\begin{aligned}(-1)^{n_h}\frac{\partial^{n_h}}{\partial\gamma^{n_h}}e^{-c\psi_z(\gamma\mathbb{1}_{dy^*_{h,T}},K_N(y)\mathbb{1}_{dy^*_{h,T}})}\Big|_{\gamma=K_T(y^*_{h,T})}&=e^{-c\psi_z(K_T(y^*_{h,T}),K_N(y^*_{h,T}))}\times\\&\times c\,P_0(dy^*_{h,T})\Bigg\{z\int_{\mathbb{R}^+}s^{n_h}\,e^{-s\,K_T(y^*_{h,T})}\rho(s)ds+\\&+(1-z)\int_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(y^*_{h,T})+K_N(y^*_{h,T}))}\rho(s)ds\Bigg\}+o(P_0(dy^*_{h,T}))\end{aligned}\tag{5.12}$$

$$\begin{aligned}(-1)^{x'_r}\frac{\partial^{x'_r}}{\partial\gamma^{x'_r}}e^{-c\psi_z(K_T(y)\mathbb{1}_{dy^*_{r,N}},\gamma\mathbb{1}_{dy^*_{r,N}})}\Big|_{\gamma=K_N(y^*_{r,N})}&=e^{-c\psi_z(K_T(y^*_{r,N}),K_N(y^*_{r,N}))}\times\\&\times c\,P_0(dy^*_{r,N})\Bigg\{z\int_{\mathbb{R}^+}s^{x'_r}\,e^{-s\,K_N(y^*_{r,N})}\rho(s)ds+\\&+(1-z)\int_{\mathbb{R}^+}s^{x'_r}\,e^{-s\,(K_T(y^*_{r,N})+K_N(y^*_{r,N}))}\rho(s)ds\Bigg\}+o(P_0(dy^*_{r,N}))\end{aligned}\tag{5.13}$$

equations (5.9)-(5.11) come from iteratively applying: $e^{-cx}\cdot x=(-1)\frac{\partial}{\partial\gamma}e^{-\gamma x}\Big|_{\gamma=c}$

$$(-1)^{q_m+x''_m}\frac{\partial^{q_m+x''_m}}{\partial\gamma_1^{q_m}\partial\gamma_2^{x''_m}}e^{-c\psi_z(\gamma_1\mathbb{1}_{dy^*_m},\gamma_2\mathbb{1}_{dy^*_m})}\Big|_{\substack{\gamma_1=K_T(y^*_m)\\ \gamma_2=K_N(y^*_m)}} =$$
$$= e^{-c\psi_z(K_T(y^*_m),K_N(y^*_m))}\times$$
$$\times\, c\,P_0(dy^*_m)\left\{(1-z)\int_{\mathbb{R}^+} s^{q_m+x''_m}\,e^{-s\,(K_T(y^*_m)+K_N(y^*_m))}\rho(s)ds\right\}$$
$$+\,o(P_0(dy^*_m))$$

Lastly, we use the auxiliary variables $\boldsymbol{V}_T$ and $\boldsymbol{V}_N$, so that equations (5.12) and (5.13) conditional on them simplify to

$$(-1)^{n_h}\frac{\partial^{n_h}}{\partial\gamma^{n_h}}e^{-c\psi_z(\gamma\mathbb{1}_{dy^*_{h,T}},K_N(y)\mathbb{1}_{dy^*_{h,T}})}\Big|_{\gamma=K_T(y^*_{h,T})} = e^{-c\psi_z(K_T(y^*_{h,T}),K_N(y^*_{h,T}))}\times$$
$$\times\, c\,P_0(dy^*_{h,T})\int_{\mathbb{R}^+} s^{n_h}\,e^{-s\,(K_T(y^*_{h,T})+K_N(y^*_{h,T})V_{h,T})}\rho(s)ds+o(P_0(dy^*_{h,T}))$$

$$(-1)^{x'_r}\frac{\partial^{x'_r}}{\partial\gamma^{x'_r}}e^{-c\psi_z(K_T(y)\mathbb{1}_{dy^*_{r,N}},\gamma\mathbb{1}_{dy^*_{r,N}})}\Big|_{\gamma=K_N(y^*_{r,N})} = e^{-c\psi_z(K_T(y^*_{r,N}),K_N(y^*_{r,N}))}\times$$
$$\times\, c\,P_0(dy^*_{r,N})\int_{\mathbb{R}^+} s^{x'_r}\,e^{-s\,(K_T(y^*_{r,N})V_{r,N}+K_N(y^*_{r,N}))}\rho(s)ds+o(P_0(dy^*_{r,N}))$$

Putting everything together and letting ϵ go to zero, we get

$$\mathbb{E}\Big[\mathcal{L}(\tilde\mu_T,\tilde\mu_N\,;\boldsymbol{x},\boldsymbol{t},\boldsymbol{y_T},\boldsymbol{y_N}\,|\boldsymbol{V}_T,\boldsymbol{V}_N)\Big] = Q(\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\;e^{-c\,\psi_z(K_T,K_N)}(1-z)^k\,c^{k_T+k_N+k}\;\times$$
$$\times\prod_{h=1}^{k_T}P_0(dy^*_{h,T})\int_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(y^*_{h,T})+K_N(y^*_{h,T})V_{h,T})}\rho(s)ds\times$$
$$\times\prod_{r=1}^{k_N}P_0(dy^*_{r,N})\int_{\mathbb{R}^+}s^{x'_r}\,e^{-s\,(K_T(y^*_{r,N})V_{r,N}+K_N(y^*_{r,N}))}\rho(s)ds\times$$
$$\times\prod_{m=1}^{k}P_0(dy^*_m)\int_{\mathbb{R}^+}s^{q_m+x''_m}\,e^{-s\,(K_T(y^*_m)+K_N(y^*_m))}\rho(s)ds$$

□

Theorem 5.2. *The posterior distribution of $\tilde{\mu}_T$ and $\tilde{\mu}_N$, conditional on $\boldsymbol{T}$, $\boldsymbol{X}$, $\boldsymbol{Y_T}$, $\boldsymbol{Y_N}$, $\boldsymbol{V_T}$ and $\boldsymbol{V_N}$, equals the distribution of the vector of CRMs*

$$
\begin{aligned}
(\tilde{\mu}_T^*, \tilde{\mu}_N^*) + &\left(\sum_{h=1}^{k_T} J_{h,T}\, \delta_{Y^*_{h,T}}, \sum_{h=1}^{k_T} J_{h,T}\, V_{h,T}\, \delta_{Y^*_{h,T}} \right) \\
&+ \left(\sum_{r=1}^{k_N} J_{h,N}\, V_{h,N}\, \delta_{Y^*_{h,N}}, \sum_{r=1}^{k_N} J_{h,N}\, \delta_{Y^*_{h,N}} \right) \\
&+ \left(\sum_{m=k}^{k} J_m\, \delta_{Y^*_m}, \sum_{m=1}^{k_0} J_m\, \delta_{Y^*_m} \right)
\end{aligned}
$$

where $\tilde{\mu}_T^*$ and $\tilde{\mu}_N^*$ are CRMs such that:

$$
\begin{aligned}
\tilde{\mu}_T^* &\stackrel{d}{=} \mu_0^* + \mu_T^* \\
\tilde{\mu}_N^* &\stackrel{d}{=} \mu_0^* + \mu_N^*
\end{aligned}
$$

where μ_0^*, μ_T^* and μ_N^* are independent CRMs with Lévy intensities respectively equal to:

$$
\begin{aligned}
\nu_0^*(ds, dy) =& c\,(1-z)\, e^{-s\,(K_T(y)+K_N(y))} P_0(dy)\, \rho(s) ds \\
\nu_T^*(ds, dy) =& c\, z\, e^{-s\, K_T(y)} P_0(dy)\, \rho(s) ds \\
\nu_N^*(ds, dy) =& c\, z\, e^{-s\, K_N(y)} P_0(dy)\, \rho(s) ds
\end{aligned}
$$

The jumps $J_{1,T}, \ldots, J_{k_T,T}$, $J_{1,N}, \ldots, J_{k_N,N}$ and $J_1, \ldots, J_k$ are mutually independent and independent from $\tilde{\mu}_T^*$ and $\tilde{\mu}_N^*$ and have densities:

$$
\begin{aligned}
f_{J_{h,T}}(s) &\propto s^{n_h}\, e^{-s\,(K_T(Y^*_{h,T}) + K_N(Y^*_{h,T}) V_{h,T})} \rho(s) ds \\
f_{J_{r,N}}(s) &\propto s^{x'_r}\, e^{-s\,(K_T(Y^*_{r,N}) V_{r,N} + K_N(Y^*_{h,T}))} \rho(s) ds \\
f_{J_k}(s) &\propto s^{q_m + x''_m}\, e^{-s\,(K_T(Y^*_m) + K_N(Y^*_m)} \rho(s) ds
\end{aligned}
$$

Proof. The posterior distribution of $\tilde{\mu}_T$ and $\tilde{\mu}_N$ is uniquely determined by the posterior joint Laplace functional tansform:

$$
\begin{aligned}
&\mathbb{E}[e^{-\tilde{\mu}_T(f_T) - \tilde{\mu}_N(f_N)} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}, \boldsymbol{V_T}, \boldsymbol{V_N})] = \\
=&\frac{\mathbb{E}[e^{-\tilde{\mu}_T(f_t) - \tilde{\mu}_N(f_N)} \boldsymbol{\mathcal{L}}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N} \mid \boldsymbol{V_T}, \boldsymbol{V_N})]}{\mathbb{E}[\boldsymbol{\mathcal{L}}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N} \mid \boldsymbol{V_T}, \boldsymbol{V_N})]}
\end{aligned}
$$

where $f_l : \mathbb{Y} \to \mathbb{R}^+$ and $\tilde{\mu}(f) = \int_{\mathbb{Y}} f(y) \tilde{\mu}(dy)$. The expected value at the denominator is the probability distribution provided by Theorem 5.1 and the numerator can be rewritten analogously as:

$$\mathbb{E}\left[e^{-\tilde{\mu}_T(f_T)-\tilde{\mu}_N(f_N)}\mathcal{L}(\tilde{\mu}_T,\tilde{\mu}_N\,;\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N}\mid \boldsymbol{V}_T,\boldsymbol{V}_N)\right] =$$
$$= Q(\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\; e^{-c\,\psi_z(K_T+f_T,K_N+f_N)}(1-z)^k\,c^{k_T+k_N+k}\;\times$$
$$\times\prod_{h=1}^{k_T}P_0(dy^*_{h,T})\int\limits_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(y^*_{h,T})+f_T(y^*_{h,T})+[K_N(y^*_{h,T})+f_N(y^*_{h,T})]V_{h,T})}\rho(s)ds\times$$
$$\times\prod_{r=1}^{k_N}P_0(dy^*_{r,N})\int\limits_{\mathbb{R}^+}s^{x'_r}\,e^{-s\,([K_T(y^*_{r,N})+f_T(y^*_{r,N})]V_{r,N}+K_N(y^*_{r,N})+f_N(y^*_{r,N}))}\rho(s)ds\times$$
$$\times\prod_{m=1}^{k}P_0(dy^*_m)\int\limits_{\mathbb{R}^+}s^{q_m+x''_m}\,e^{-s\,(K_T(y^*_m)+f_T(y^*_m)+K_N(y^*_m)+f_N(y^*_m))}\rho(s)ds$$

So that, when ϵ goes to 0 the posterior joint Laplace functional transform is equal to

$$\mathbb{E}[e^{-\tilde{\mu}_T(f_T)-\tilde{\mu}_N(f_N)}\mid\boldsymbol{T},\boldsymbol{X},\boldsymbol{Y}_T,\boldsymbol{Y}_N,\boldsymbol{V}_T,\boldsymbol{V}_N] =$$
$$= e^{-c\,\psi_z(K_T+f_T,K_N+f_N)+c\psi_z(K_T,K_N)}\;\times$$
$$\times\prod_{h=1}^{k_T}\frac{\int\limits_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(Y^*_{h,T})+f_T(Y^*_{h,T})+[K_N(Y^*_{h,T})+f_N(Y^*_{h,T})]V_{h,T})}\rho(s)ds}{\int\limits_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T})}\rho(s)ds}\times$$
$$\times\prod_{r=1}^{k_N}\frac{\int\limits_{\mathbb{R}^+}s^{x'_r}\,e^{-s\,([K_T(Y^*_{r,N})+f_T(Y^*_{r,N})]V_{r,N}+K_N(Y^*_{r,N})+f_N(Y^*_{r,N}))}\rho(s)ds}{\int\limits_{\mathbb{R}^+}s^{x'_r}\,e^{-s\,(K_T(Y^*_{r,N})V_{r,N}+K_N(Y^*_{r,N}))}\rho(s)ds}\times$$
$$\times\prod_{m=1}^{k}\frac{\int\limits_{\mathbb{R}^+}s^{q_m+x''_m}\,e^{-s\,(K_T(Y^*_m)+f_T(Y^*_m)+K_N(Y^*_m)+f_N(Y^*_m))}\rho(s)ds}{\int\limits_{\mathbb{R}^+}s^{q_m+x''_m}\,e^{-s\,(K_T(Y^*_m)+K_N(Y^*_m))}\rho(s)ds}$$

where the first factor is the joint Laplace functional transform of the two CRMs $\tilde{\mu}^*_T$ and $\tilde{\mu}^*_N$. Moreover, for the second factor we have

$$\frac{\int\limits_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(Y^*_{h,T})+f_T(Y^*_{h,T})+[K_N(Y^*_{h,T})+f_N(Y^*_{h,T})]V_{h,T})}\rho(s)ds}{\int\limits_{\mathbb{R}^+}s^{n_h}\,e^{-s\,(K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T})}\rho(s)ds} = \mathbb{E}[e^{-s\,(f_T(Y^*_{h,T})+f_N(Y^*_{h,T})V_{h,T})}\mid V_{h,T}]$$

which is the conditional Laplace transform of the vector $(s, V_{h,T}\cdot s)$, where s has density proportional to

$$s^{n_h}\,e^{-s\,(K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T})}\rho(s)ds$$

Similar interpretation can be applied to the last two factors, concluding the proof. □

Corollary 5.1. *For any $t > 0$, the posterior estimate of the intensity function $\lambda_N(t)$ conditionally given the observations $\boldsymbol{T}$ and $\boldsymbol{X}$ and the auxiliary variables $(\boldsymbol{Y_l})_{l\in\{T,N\}}$ and $(\boldsymbol{V_l})_{l\in\{T,N\}}$ under a square loss function is*

$$\begin{aligned}
&\int_{\mathbb{R}^+\times\mathbb{Y}} s\, k_N(t;\, y) e^{-s\, K_N(y)} \left[(1-z)e^{-s\, K_T(y)} + z\right] \rho(s) ds\, c\, P_0(dy) + \\
&+ \sum_{h=1}^{k_T} V_{h,T}\, k_N(t; Y^*_{h,T}) \int_{\mathbb{R}^+} u f_{J_{h,T}}(u) du + \\
&+ \sum_{r=1}^{k_N} k_N(t; Y^*_{r,N}) \int_{\mathbb{R}^+} u f_{J_{r,N}}(u) du + \\
&+ \sum_{m=1}^{k_0} k_N(t; Y^*_m) \int_{\mathbb{R}^+} u f_{J_k}(u) du
\end{aligned} \tag{5.14}$$

Proof.

$$\begin{aligned}
&\mathbb{E}[\lambda_N(t) \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_T, \boldsymbol{Y}_N, \boldsymbol{V}_T, \boldsymbol{V}_N] = \\
&\qquad = \mathbb{E}\left[\int_{\mathbb{Y}} k_N(t;\, y)\tilde{\mu}_N(dy) \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_T, \boldsymbol{Y}_N, \boldsymbol{V}_T, \boldsymbol{V}_N\right] = \\
&\qquad = \mathbb{E}\left[\int_{\mathbb{Y}} k_N(t;\, y)\tilde{\mu}^*_N(dy)\right] + \sum_{h=1}^{k_T} \mathbb{E}[J_{h,T}]\, V_{h,T}\, k_N(t; Y^*_{h,T}) + \\
&\qquad + \sum_{r=1}^{k_N} \mathbb{E}[J_{r,N}]\, k_N(t; Y^*_{r,N}) + \sum_{m=1}^{k_0} \mathbb{E}[J_m]\, k_N(t; Y^*_m)
\end{aligned}$$

where

$$\mathbb{E}\left[\int_{\mathbb{Y}} k_N(t;\, y)\tilde{\mu}^*_N(dy)\right] = \mathbb{E}\left[\int_{\mathbb{Y}} k_N(t;\, y)\mu^*_0(dy)\right] + \mathbb{E}\left[\int_{\mathbb{Y}} k_N(t;\, y)\mu^*_N(dy)\right]$$

which, according to the results in Appendix B, becomes

$$\int_{\mathbb{R}^+\times\mathbb{Y}} s\, k_N(t;\, y)\, v^*_0(ds, dy) + \int_{\mathbb{R}^+\times\mathbb{Y}} s\, k_N(t;\, y)\, v^*_N(ds, dy)$$

substituiting the Lévy intensities provided by Theorem 5.2

$$\int_{\mathbb{R}^+\times\mathbb{Y}} s\, k_N(t;\, y) e^{-s\, K_N(y)} \left[(1-z)e^{-s\, K_T(y)} + z\right] \rho(s) ds\, c\, P_0(dy)$$

while $\mathbb{E}[J_{h,T}]$, $\mathbb{E}[J_{r,N}]$ and $\mathbb{E}[J_m]$ can be computed using their respective densities provided by Theorem 5.2 and the thesis follows. □

5.3.3 Hierarchical CRMs posterior law

Also in the case of hierarchical CRMs, by discretness of $\tilde{\mu}_T$ and $\tilde{\mu}_N$, there will be ties between the auxiliary random variables $Y_{i,j,l}$'s. We denote the joint collection of the distinct values in $\boldsymbol{y}_T$ and $\boldsymbol{y}_N$ as $\{y_1^*, \dots, y_k^*\}$. Moreover, we define the frequencies of $\boldsymbol{y}_T$ and the counts of $\boldsymbol{y}_N$ respectively as

$$n_{h,T} = \sum_i \sum_j \mathbb{1}(y_{i,j,T} = y_h^*) \qquad \text{and} \qquad n_{h,N} = \sum_{\substack{(i,j)\,: \\ y_{i,j,N} = y_h^*}} x_{i,j}$$

Notice that differently from the previous section both frequencies and counts can be equal to zero.

Moreover, let us introduce two addictional vectors of latent random variables

$$\boldsymbol{C}_l = \{C_{i,j,l} \,:\, j = 1, \dots, m_i \,,\, i = 1, \dots, n\} \qquad \text{for } \; l \in \{T, N\}$$

such that in the Chinese franchise metaphor they will correspond to the labels of the tables.

Theorem 5.3. *Let* $Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} k(t_{i,j}; y_{i,j,T}) \frac{1}{x_{i,j}!} H_{i,j}(y_{i,j,N})^{x_{i,j}}$*, the probability distribution of* $(T,\, X,\, Y_N,\, Y_Z)$ *conditionally on* C_T *and* C_N *equals*

$$\begin{aligned}
\pi(T,\, X,\, Y_N,\, Y_Z | C_T, C_N) = \; & Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})\, c_0^k \prod_{h=1}^{k} P_0(dy_h^*) \times \\
& e^{-c_0 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(T)}(K_T(y)) + \psi^{(N)}(K_N(y)))P_0(dy)} \times \\
& \times \prod_{h=1}^{k} \tau^{(0)}_{r_{T,h}+r_{N,h}} \left(\psi^{(T)}(K_T(y_h^*)) + \psi^{(N)}(K_N(y_h^*))\right) \times \\
& \times \prod_{l \in T,N} \prod_{h=1}^{k} \binom{n_{h,l}}{q_{l,h,1}, \cdots, q_{l,h,r_{l,h}}} \frac{1}{r_{l,h}!} \prod_{c=1}^{r_{l,h}} \tau^{(l)}_{q_{l,h,c}}(K_l(y_h^*))
\end{aligned}$$

where, according to the Chinese franchise metaphor, $r_{l,h}$ *denote the number of tables eating dish* y_h^* *in restaurant* l*, while* $q_{l,h,c}$ *denotes the number of customers in restaurant* l *seated at the* c*-th table that serves dish* y_h^**.*

Proof. The density $\pi(T,\, X,\, Y_N,\, Y_Z)$ can be computed as the expected value of the quantity

in (5.7) as follows

$$\begin{aligned}\mathbb{E}\Bigg[\mathbb{E}\Big[\mathcal{L}(\tilde{\mu}_T,\tilde{\mu}_N;\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\mid\tilde{\mu}_0\Big]\Bigg] &= Q(\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\times\\ &\times\mathbb{E}\Bigg[\mathbb{E}\Big[e^{-\int_{\mathbb{Y}}K_T(y)\tilde{\mu}_T(dy)}\prod_{h=1}^{k}\mu_T(dy_h^*)^{n_{h,T}}\mid\tilde{\mu}_0\Big]\times\\ &\times\mathbb{E}\Big[e^{-\int_{\mathbb{Y}}K_N(y)\tilde{\mu}_N(dy)}\prod_{h=1}^{k}\mu_N(dy_h^*)^{n_{h,N}}\mid\tilde{\mu}_0\Big]\Bigg]\end{aligned}\tag{5.15}$$

Set, as before $\mathbb{Y}^*=\mathbb{Y}\setminus\{dy_1^*,\ldots,dy_k^*\}$ where $dy=[y,y+\epsilon)$ for $\epsilon>0$ and arbitrarily small so that the intervals dy^* are disjoint. The last two rows in (5.15) equals

$$\begin{aligned}&\mathbb{E}\Bigg[\mathbb{E}\Big[e^{-\int_{\mathbb{Y}^*}K_T(y)\tilde{\mu}_T(dy)}\mid\tilde{\mu}_0\Big]\mathbb{E}\Big[\prod_{h=1}^{k}e^{-K_T(y_h^*)\tilde{\mu}_T(dy_h^*)}\mu_T(dy_h^*)^{n_{h,T}}\mid\tilde{\mu}_0\Big]\times\\ &\times\mathbb{E}\Big[e^{-\int_{\mathbb{Y}^*}K_N(y)\tilde{\mu}_N(dy)}\mid\tilde{\mu}_0\Big]\mathbb{E}\Big[\prod_{h=1}^{k}e^{-K_N(y_h^*)\tilde{\mu}_N(dy_h^*)}\mu_N(dy_h^*)^{n_{h,N}}\mid\tilde{\mu}_0\Big]\Bigg]=\end{aligned}$$

$$\begin{aligned}&=\mathbb{E}\Bigg[e^{-\int_{\mathbb{Y}^*}\psi^{(T)}(K_T(y))\tilde{\mu}_0(dy)}\prod_{h=1}^{k}(-1)^{n_{h,T}}\frac{\partial^{n_{h,T}}}{\partial\gamma^{n_{h,T}}}\mathbb{E}\left[e^{-\gamma\tilde{\mu}_T(dy_h^*)}\mid\tilde{\mu}_0\right]\Bigg|_{\gamma=K_T(y_h^*)}\times\\ &\times e^{-\int_{\mathbb{Y}^*}\psi^{(N)}(K_N(y))\tilde{\mu}_0(dy)}\prod_{h=1}^{k}(-1)^{n_{h,N}}\frac{\partial^{n_{h,N}}}{\partial\gamma^{n_{h,N}}}\mathbb{E}\left[e^{-\gamma\tilde{\mu}_N(dy_h^*)}\mid\tilde{\mu}_0\right]\Bigg|_{\gamma=K_N(y_h^*)}\Bigg]=\end{aligned}\tag{5.16}$$

$$\begin{aligned}&=\mathbb{E}\Bigg[e^{-\int_{\mathbb{Y}^*}\psi^{(T)}(K_T(y))\tilde{\mu}_0(dy)}\prod_{h=1}^{k}(-1)^{n_{h,T}}\frac{\partial^{n_{h,T}}}{\partial\gamma^{n_{h,T}}}\,e^{-\psi^{(T)}(\gamma)\tilde{\mu}_0(dy_h^*)}\Bigg|_{\gamma=K_T(y_h^*)}\times\\ &\times e^{-\int_{\mathbb{Y}^*}\psi^{(N)}(K_N(y))\tilde{\mu}_0(dy)}\prod_{h=1}^{k}(-1)^{n_{h,N}}\frac{\partial^{n_{h,N}}}{\partial\gamma^{n_{h,N}}}\,e^{-\psi^{(N)}(\gamma)\tilde{\mu}_0(dy_h^*)}\Bigg|_{\gamma=K_N(y_h^*)}\Bigg]\end{aligned}$$

Applying Faà di Bruno's formula, as shown in Appendix C, equation (5.16) can be rewritten as

$$\begin{aligned}&\mathbb{E}\left[e^{-\int_{\mathbb{Y}^*}\psi^{(T)}(K_T(y))\tilde{\mu}_0(dy)-\int_{\mathbb{Y}^*}\psi^{(N)}(K_N(y))\tilde{\mu}_0(dy)}\right]\times\\ &\times\mathbb{E}\left[\prod_{h=1}^{k}\prod_{l\in\{T,N\}}\sum_{r_{l,h}=1}^{n_{h,l}}\xi_{n_{h,l},l,r_{l,h}}(K_l(y_h^*))\,e^{-\psi^{(l)}(K_l(y_h^*))\tilde{\mu}_0(dy_h^*)}\tilde{\mu}_0(dy_h^*)^{r_{l,h}}\right]\end{aligned}\tag{5.17}$$

where $\xi_{n_{h,l},l,r}(K_l(y_h^*)) = \sum_{(*)} \binom{n_{h,l}}{q_1,\cdots,q_r} \frac{1}{r!} \tau_{q_1}^{(l)}(K_l(y_h^*)) \cdots \tau_{q_r}^{(l)}(K_l(y_h^*))$, where $\tau_q^{(l)}(u) = \int_{\mathbb{R}^+} s^q \, e^{-su} \rho_l(s) ds$ and the sum $(*)$ runs over all vectors $(q_1, \ldots, q_r)$ of positive integers such that $\sum_{j=1}^r q_j = n_{h,l}$, for $l \in \{T, N\}$. Computing the product in h and l and denoting with $\boldsymbol{r}$ the set of all vectors corresponding to a term in the resulting summation: $\boldsymbol{r} = \{(r_{T,1}, \ldots, r_{T,k}, r_{N,1}, \ldots, r_{N,k}) \, : \, r_{l,h} \in \{1, \ldots, n_{h,l}\}\}$, we get

$$\begin{aligned} \sum_{\boldsymbol{r}} \left(\prod_{l \in T,N} \prod_{h=1}^{k} \xi_{n_{h,l},l,r_{l,h}}(K_l(y_h^*)) \right) \mathbb{E}\left[e^{-\int_{\mathbb{Y}^*} \psi^{(T)}(K_T(y)) \tilde{\mu}_0(dy) - \int_{\mathbb{Y}^*} \psi^{(N)}(K_N(y)) \tilde{\mu}_0(dy)} \right] \times \\ \times \prod_{h=1}^{k} \mathbb{E}\left[e^{-[\psi^{(T)}(K_T(y_h^*)) - \psi^{(N)}(K_N(y_h^*))] \tilde{\mu}_0(dy_h^*)} \tilde{\mu}_0(dy_h^*)^{r_{T,h}+r_{N,h}} \right] \end{aligned} \tag{5.18}$$

Applying again Faà di Bruno's formula, equation (5.15) equals

$$\begin{aligned} \mathbb{E}\left[\mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N \, ; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) \right] = Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) \, c_0^k \prod_{h=1}^{k} P_0(dy_h^*) \times \\ e^{-c_0 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(T)}(K_T(y)) + \psi^{(N)}(K_N(y))) P_0(dy)} \times \\ \times \sum_{\boldsymbol{r}} \left[\prod_{l \in T,N} \prod_{h=1}^{k} \xi_{n_{h,l},l,r_{l,h}}(K_l(y_h^*)) \, \tau_{r_{T,h}+r_{N,h}}\left(\psi^{(T)}(K_T(y_h^*)) + \psi^{(N)}(K_N(y_h^*))\right) \right] \\ + o(\prod_{h=1}^{k} P_0(dy_h^*)) \end{aligned} \tag{5.19}$$

susbstituting the expression of $\xi_{n_{h,l},l,r}(K_l(y_h^*))$ we get

$$\begin{aligned} \mathbb{E}\left[\mathcal{L}(\tilde{\mu}_T, \tilde{\mu}_N \, ; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) \right] = Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N}) \, c_0^k \prod_{h=1}^{k} P_0(dy_h^*) \times \\ e^{-c_0 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(T)}(K_T(y)) + \psi^{(N)}(K_N(y))) P_0(dy)} \times \\ \times \sum_{\boldsymbol{r}} \left[\prod_{l \in T,N} \prod_{h=1}^{k} \sum_{(*)} \binom{n_{h,l}}{q_1, \cdots, q_{r_{l,h}}} \frac{1}{r_{l,h}!} \tau_{q_1}^{(l)}(K_l(y_h^*)) \cdots \tau_{q_{r_{l,h}}}^{(l)}(K_l(y_h^*)) \right. \\ \left. \tau_{r_{T,h}+r_{N,h}}^{(0)}\left(\psi^{(T)}(K_T(y_h^*)) + \psi^{(N)}(K_N(y_h^*))\right) \right] + o(\prod_{h=1}^{k} P_0(dy_h^*)) \end{aligned}$$

computing the products in l and h

$$\mathbb{E}\left[\mathcal{L}(\tilde{\mu}_T,\tilde{\mu}_N\,;\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\right] = Q(\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\,c_0^k\prod_{h=1}^{k}P_0(dy_h^*)\times$$
$$e^{-c_0\int_{\mathbb{Y}}\psi^{(0)}(\psi^{(T)}(K_T(y))+\psi^{(N)}(K_N(y)))P_0(dy)}\times$$
$$\times\sum_{\boldsymbol{r}}\sum_{\boldsymbol{q}}\left[\prod_{l\in T,N}\prod_{h=1}^{k}\binom{n_{h,l}}{q_{l,h,1},\cdots,q_{l,h,r_{l,h}}}\frac{1}{r_{l,h}!}\,\tau^{(l)}_{q_{l,h,1}}(K_l(y_h^*))\cdots\tau^{(l)}_{q_{l,h,r_{l,h}}}(K_l(y_h^*))\right.$$
$$\left.\tau^{(0)}_{r_{T,h}+r_{N,h}}\big(\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))\big)\right]+o(\prod_{h=1}^{k}P_0(dy_h^*))$$

Finally, we introduce the latent random variables $\boldsymbol{C}_T$ and $\boldsymbol{C}_N$, to get rid of the sum over $\boldsymbol{r}$ and $\boldsymbol{q}$

$$\pi(T,X,Y_T,Y_N\mid C_T,C_N)=\mathbb{E}\left[\mathcal{L}(\tilde{\mu}_T,\tilde{\mu}_N\,;\boldsymbol{t},\boldsymbol{n},\boldsymbol{y_T},\boldsymbol{y_N}\mid\boldsymbol{C_T},\boldsymbol{C_N})\right]=$$
$$=Q(\boldsymbol{t},\boldsymbol{x},\boldsymbol{y_T},\boldsymbol{y_N})\,c_0^k\prod_{h=1}^{k}P_0(dy_h^*)e^{-c_0\int_{\mathbb{Y}}\psi^{(0)}(\psi^{(T)}(K_T(y))+\psi^{(N)}(K_N(y)))P_0(dy)}\times$$
$$\times\prod_{h=1}^{k}\tau^{(0)}_{r_{T,h}+r_{N,h}}\big(\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))\big)\times$$
$$\times\prod_{l\in T,N}\prod_{h=1}^{k}\binom{n_{h,l}}{q_{l,h,1},\cdots,q_{l,h,r_{l,h}}}\frac{1}{r_{l,h}!}\,\tau^{(l)}_{q_{l,h,1}}(K_l(y_h^*))\cdots\tau^{(l)}_{q_{l,h,r_{l,h}}}(K_l(y_h^*))+$$
$$+o(\prod_{h=1}^{k}P_0(dy_h^*))$$

letting ϵ go to zero completes the proof. □

Theorem 5.4. *The posterior distribution of $\tilde{\mu}_T$ and $\tilde{\mu}_N$, conditional on T, X, Y_T, Y_N, C_T and C_N equals the distribution of the vector of CRMs*

$$(\tilde{\mu}_T^*,\tilde{\mu}_N^*)+\left(\sum_{h=1}^{k}\sum_{c=1}^{r_{T,h}}J_{T,h,c}\,\delta_{Y_h^*},\sum_{h=1}^{k}\sum_{c=1}^{r_{N,h}}J_{N,h,c}\,\delta_{Y_h^*}\right)$$

where $\tilde{\mu}_T^*$ and $\tilde{\mu}_N^*$ are CRMs such that:

$$\tilde{\mu}_l^*\mid\tilde{\mu}_0^*\overset{ind}{\sim}CRM(\tilde{\nu}_l^*)$$

for $l \in \{T, N\}$, with $\tilde{\nu}_l^*(ds, dy) = e^{-sK_l(y)}\rho_l(s)ds\tilde{\mu}_0^*(dy)$ and $\tilde{\mu}_0^*$ is a CRMs such that:

$$\tilde{\mu}_0^* \stackrel{d}{=} \eta_0^* + \sum_{h=1}^{k} I_h \delta_{Y_h^*}$$

where η_0^* is a CRM with no fixed jumps and Lévy intensity given by

$$\nu_0^*(ds, dy) = e^{-s[\psi^{(T)}(K_T(y))+\psi^{(N)}(K_N(y))]}\rho_0(s)ds\, c_0\, P_0(dy)$$

and the jumps $I_1, \ldots, I_k$ are mutually independent and independent from η_0^* and have density:

$$f_{I_h}(s) \propto s^{r_{T,h}+r_{N,h}} e^{-s[\psi^{(T)}(K_T(Y_h^*))+\psi^{(N)}(K_N(Y_h^*))]}\rho_0(s)ds$$

Lastly, the sequences of independent jumps $(J_{T,h,c})_{h,c}$ and $(J_{N,h,c})_{h,c}$ are independent from one another, independent from $(\tilde{\mu}_T^*, \tilde{\mu}_N^*)$ and have densities:

$$f_{J_{l,h,c}}(s) \propto s^{q_{l,h,c}} e^{-s\, K_l(Y_h^*)}\rho_l(s)ds$$

Proof. The posterior of $\tilde{\mu}_T$ and $\tilde{\mu}_N$ is uniquely determined by

$$\begin{aligned} &\mathbb{E}[e^{-\tilde{\mu}_T(f_T)-\tilde{\mu}_N(f_N)} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_T, \boldsymbol{Y}_N, \boldsymbol{C}_T, \boldsymbol{C}_N)] = \\ &= \frac{\mathbb{E}[e^{-\tilde{\mu}_T(f_t)-\tilde{\mu}_N(f_N)}\boldsymbol{\mathcal{L}}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N} \mid \boldsymbol{C_T}, \boldsymbol{C_N})]}{\mathbb{E}[\boldsymbol{\mathcal{L}}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N} \mid \boldsymbol{C_T}, \boldsymbol{C_N})]} \end{aligned} \tag{5.20}$$

where $f_l : \mathbb{Y} \to \mathbb{R}^+$ and $\tilde{\mu}(f) = \int_{\mathbb{Y}} f(y)\tilde{\mu}(dy)$.

The denominator is given by Theorem 5.3, while using the same techniques the numerator can be rewritten as

$$\begin{aligned} \mathbb{E}\left[e^{-\tilde{\mu}_T(f_t)-\tilde{\mu}_N(f_N)}\boldsymbol{\mathcal{L}}(\tilde{\mu}_T, \tilde{\mu}_N\,; \boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N} \mid C_T, C_N) \right] &= Q(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y_T}, \boldsymbol{y_N})\times \\ &\times c_0^k \prod_{h=1}^{k} P_0(dy_h^*)\, e^{-c_0 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(T)}(K_T(y)+f_T(y))+\psi^{(N)}(K_N(y)+f_N(y))P_0(dy)} \times \\ &\times \prod_{h=1}^{k} \tau^{(0)}_{r_{T,h}+r_{N,h}}\big(\psi^{(T)}(K_T(y_h^*)+f_T(y_h^*)) + \psi^{(N)}(K_N(y_h^*)+f_N(y_h^*))\big)\times \\ &\times \prod_{l\in\{T,N\}} \prod_{h=1}^{k} \binom{n_{h,l}}{q_{l,h,1}, \cdots, q_{l,h,r_{l,h}}} \frac{1}{r_{l,h}!} \prod_{c=1}^{r_{l,h}} \tau^{(l)}_{q_{l,h,c}}(K_l(y_h^*)+f_l(y_h^*))+ \\ &+ o(\prod_{h=1}^{k} P_0(dy_h^*)) \end{aligned}$$

So equation (5.20) equals

$$\begin{aligned}
&\mathbb{E}[e^{-\tilde{\mu}_T(f_T)-\tilde{\mu}_N(f_N)} \mid \boldsymbol{X}, \boldsymbol{T}, \boldsymbol{Y}_T, \boldsymbol{Y}_N, \boldsymbol{C}_T, \boldsymbol{C}_N)] = \\
&\exp\Big\{ - c_0 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(T)}(K_T(y) + f_T(y)) + \psi^{(N)}(K_N(y) + f_N(y))P_0(dy) + \\
&c_0 \int_{\mathbb{Y}} \psi^{(0)}(\psi^{(T)}(K_T(y)) + \psi^{(N)}(K_N(y)))P_0(dy)\Big\} \times \\
&\times \prod_{h=1}^{k} \frac{\tau^{(0)}_{r_{T,h}+r_{N,h}}\big(\psi^{(T)}(K_T(y_h^*) + f_T(y_h^*)) + \psi^{(N)}(K_N(y_h^*) + f_N(y_h^*))\big)}{\tau^{(0)}_{r_{T,h}+r_{N,h}}\big(\psi^{(T)}(K_T(y_h^*)) + \psi^{(N)}(K_N(y_h^*))\big)} \times \\
&\times \prod_{l\in\{T,N\}} \prod_{h=1}^{k} \prod_{c=1}^{r_{l,h}} \frac{\tau^{(l)}_{q_{l,h,c}}(K_l(y_h^*) + f_l(y_h^*))}{\tau^{(l)}_{q_{l,h,c}}(K_l(y_h^*))}
\end{aligned}$$

where the first two factors are the joint Laplace functional of the two CRMs $\tilde{\mu}_T^*$ and $\tilde{\mu}_N^*$. Moreover, for the last factor we have

$$\frac{\tau^{(l)}_{q_{l,h,c}}(K_l(y_h^*) + f_l(y_h^*))}{\tau^{(l)}_{q_{l,h,c}}(K_l(y_h^*))} = [e^{-s\, f_l(y_h^*)}]$$

is the Laplace transform of $J_{l,h,c}$. □

5.4 Posterior Inference

In this section, we derive a MCMC marginal sampler for the GM-dependent prior: firstly, in presence of general kernel, intensity and base distribution and, secondly, using OU-kernel, gamma CRMs and uniform base distribution.

5.4.1 GM-dependent CRMs marginal sampler

Full conditional distributions for the latent variables

We provide in this section the full conditional distribution for the vector of latent variables $(Y_{i,j,l}, V_{i,j,l})$ for $j = 1, \ldots, m_i$, $i = 1, \ldots, n$ and $l \in \{T, N\}$. Firstly, we have that

$$\begin{aligned}
&\mathbb{P}[Y_{i,j,T} \in dy \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y}_{\boldsymbol{N}}, \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V}_{\boldsymbol{N}}] \\
&= w_0\, G_0(dy) + \sum_{h=1}^{k_T^{-(i,j,T)}} w_{h,T}\, \delta_{Y^*_{h,T}}(dy) + \sum_{r=1}^{k_N} w_{r,N}\, \delta_{Y^*_{r,N}}(dy) + \sum_{m=1}^{k^{-(i,j,T)}} w_m\, \delta_{Y^*_m}(dy)
\end{aligned} \tag{5.21}$$

where $-(i,j)$ and $-(i,j,T)$ are used to denote the fact that the element arising from the j-th observation for the observational process of the i-th unit has been removed. Using the

result from Theorem 5.2 and denoting with $(Y_{i,j,T} = \text{new})$ the event that $Y_{i,j,T}$ does not coincide with any element in $\boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}$ and $\boldsymbol{Y_N}$, we have

$$\begin{aligned} G_0(dy) =& \mathbb{P}[Y_{i,j,T} \in dy \mid Y_{i,j,T} = \text{new}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N}, \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}] \\ \propto& \mathbb{P}[Y_{i,j,T} \in dy, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N} \mid Y_{i,j,T} = \text{new}, \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}] \\ \propto& z \int_0^{+\infty} s e^{-s\, K_T(y)} \rho(s) ds\, k(t_{i,j}; y) P_0(dy) \\ &+ (1-z) \int_0^{+\infty} s e^{-s\,(K_T(y)+K_N(y))} \rho(s) ds\, k(t_{i,j}; y) P_0(dy) \end{aligned} \tag{5.22}$$

$$\begin{aligned} w_0 =& \mathbb{P}[Y_{i,j,T} = \text{new} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N}, \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}] \\ =& \int_{\mathbb{Y}} \mathbb{P}[Y_{i,j,T} \in dy, Y_{i,j,T} = \text{new} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N}, \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}] dy \\ =& \int_{\mathbb{Y}} \frac{\mathbb{P}[Y_{i,j,T} \in dy, Y_{i,j,T} = \text{new}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N} \mid \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}]}{\mathbb{P}[\boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N} \mid \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}]} dy \\ \propto& c\, z \int_{\mathbb{Y}} \int_0^{+\infty} s e^{-s\, K_T(y)} \rho(s) ds\, k(t_{i,j}; y) P_0(dy) \\ &+ c\,(1-z) \int_{\mathbb{Y}} \int_0^{+\infty} s e^{-s\,(K_T(y)+K_N(y))} \rho(s) ds\, k(t_{i,j}; y) P_0(dy) \end{aligned} \tag{5.23}$$

$$\begin{aligned} w_{h,T} =& \mathbb{P}[Y_{i,j,T} = Y^*_{h,T} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N}, \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}] \\ =& \frac{\mathbb{P}[Y_{i,j,T} = Y^*_{h,T}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N} \mid \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}]}{\mathbb{P}[\boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{Y_N} \mid \boldsymbol{V}_{\boldsymbol{T}}^{-(i,j)}, \boldsymbol{V_N}]} \\ \propto& \frac{\int_0^{+\infty} s^{n_h^{-(i,j)}+1} e^{-s(K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T})} \rho(s) ds}{\int_0^{+\infty} s^{n_h^{-(i,j)}} e^{-s(K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T})} \rho(s) ds} k(t_{i,j}; Y^*_{h,T}) \end{aligned} \tag{5.24}$$

$$w_{r,N} \propto (1-z) \frac{\int_0^{+\infty} s^{x'_r+1} e^{-s(K_T(Y^*_{r,N})V_{r,N}+K_N(Y^*_{r,N}))} \rho(s) ds\, \mathbb{1}_{\{V_{r,N}=1\}}}{\int_0^{+\infty} s^{x'_r} e^{-s(K_T(Y^*_{r,N})V_{r,N}+K_N(Y^*_{r,N}))} \rho(s) ds} k(t_{i,j}; Y^*_{r,N}) \tag{5.25}$$

$$w_m \propto \frac{\int_0^{+\infty} s^{q_m^{-(i,j)}+x''_m+1} e^{-s(K_T(Y^*_m)+K_N(Y^*_m))}\rho(s)ds}{\int_0^{+\infty} s^{q_m^{-(i,j)}+x''_m} e^{-s(K_T(Y^*_m)+K_N(Y^*_m))}\rho(s)ds}\, k(t_{i,j}; Y^*_m) \tag{5.26}$$

Secondly, consider the conditional distribution $\mathbb{P}[\boldsymbol{V_T} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}, \boldsymbol{V_T}^{-(i,j)}, \boldsymbol{V_N}]$, if $Y_{i,j,T}$ equals some element in $\boldsymbol{Y_T}^{-(i,j)}$ and/or $\boldsymbol{Y_N}$, then the distribution is degenerate because $\boldsymbol{V_T} = \boldsymbol{V_T}^{-(i,j)}$. Contrary, if $Y_{i,j,T}$ does not coincide with any element in $\boldsymbol{Y_T}^{-(i,j)}$ and $\boldsymbol{Y_N}$, i.e. a new value $y_{h,T}$ has been sampled according to (5.22), the full conditional of $V_{h,T}$ for h such that $Y_{i,j,T} = y_{h,T}$ is

$$\begin{aligned} &\mathbb{P}[V_{h,T} = v \mid Y_{i,j,T} = y_{h,T}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}^{-(i,j)}, \boldsymbol{Y_N}, \boldsymbol{V_T}^{-(i,j)}, \boldsymbol{V_N}] \\ &\propto \mathbb{P}[V_{h,T} = v]\, \mathbb{P}[\boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N} \mid \boldsymbol{V_T}, \boldsymbol{V_N}] \\ &\propto z^{(1-v)}\,(1-z)^{v} \int_0^{+\infty} s\, e^{-s(K_T(y_{h,T})+K_N(y_{h,T})\,v)}\rho(s)ds \end{aligned} \tag{5.27}$$

Analogously, we find the full conditional distribution for $(Y_{i,j,N}, V_{i,j,N})$ given by:

$$\begin{aligned} &\mathbb{P}[Y_{i,j,N} \in dy \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}^{-(i,j)}, \boldsymbol{V_T}, \boldsymbol{V_N}^{-(i,j)}] \\ &= p_0\, H_0(dy) + \sum_{h=1}^{k_T} p_{h,T}\, \delta_{Y^*_{h,T}}(dy) + \sum_{r=1}^{k_N^{-(i,j,N)}} p_{r,N}\, \delta_{Y^*_{r,N}}(dy) + \sum_{m=1}^{k^{-(i,j,N)}} p_m\, \delta_{Y^*_m}(dy) \end{aligned} \tag{5.28}$$

where

$$\begin{aligned} H_0(dy) =& \mathbb{P}[Y_{i,j,N} \in dy \mid Y_{i,j,N} = \text{new}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}^{-(i,j)}, \boldsymbol{V_T}, \boldsymbol{V_N}^{-(i,j)}] \\ \propto z & \int_0^{+\infty} s^{x_{i,j}} e^{-s\,K_N(y)}\rho(s)ds\, \frac{1}{x_{i,j}!} H_{i,j}(y)^{x_{i,j}}\, P_0(dy) \\ &+ (1-z) \int_0^{+\infty} s^{x_{i,j}} e^{-s\,(K_T(y)+K_N(y))}\rho(s)ds\, \frac{1}{x_{i,j}!} H_{i,j}(y)^{x_{i,j}}\, P_0(dy) \end{aligned} \tag{5.29}$$

$$
\begin{aligned}
p_0 =& \mathbb{P}[Y_{i,j,N} = \text{new} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}^{-(i,j)}, \boldsymbol{V_T}, \boldsymbol{V_N}^{-(i,j)}] \\
&\propto c\, z \int_{\mathbb{Y}} \int_0^{+\infty} s^{x_{i,j}} e^{-s\, K_N(y)} \rho(s) ds \, \frac{1}{x_{i,j}!} H_{i,j}(y)^{x_{i,j}} P_0(dy) \\
&\quad + c\,(1-z) \int_{\mathbb{Y}} \int_0^{+\infty} s^{x_{i,j}} e^{-s\,(K_T(y)+K_N(y))} \rho(s) ds \, \frac{1}{x_{i,j}!} H_{i,j}(y)^{x_{i,j}} P_0(dy)
\end{aligned}
\tag{5.30}
$$

$$
\begin{aligned}
p_{h,T} =& \ \mathbb{P}[Y_{i,j,N} = Y^*_{h,T} \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}^{-(i,j)}, \boldsymbol{V_T}, \boldsymbol{V_N}^{-(i,j)}] \\
\propto& (1-z) \frac{\int_0^{+\infty} s^{n_h + x_{i,j}} e^{-s(K_T(Y^*_{h,T}) + K_N(Y^*_{h,T}) V_{h,T})} \rho(s) ds \, \mathbb{1}_{\{V_{h,T}=1\}}}{\int_0^{+\infty} s^{n_h} e^{-s(K_T(Y^*_{h,T}) + K_N(Y^*_{h,T}) V_{h,T})} \rho(s) ds} \, \frac{1}{x_{i,j}!} H_{i,j}(Y^*_{h,T})^{x_{i,j}}
\end{aligned}
\tag{5.31}
$$

$$
p_{r,N} \propto \frac{\int_0^{+\infty} s^{{x'}^{-(i,j)}_r + x_{i,j}} e^{-s(K_T(Y^*_{r,N}) V_{r,N} + K_N(Y^*_{r,N}))} \rho(s) ds}{\int_0^{+\infty} s^{{x'}^{-(i,j)}_r} e^{-s(K_T(Y^*_{r,N}) V_{r,N} + K_N(Y^*_{r,N}))} \rho(s) ds} \, \frac{1}{x_{i,j}!} H_{i,j}(Y^*_{r,N})^{x_{i,j}} \tag{5.32}
$$

$$
p_m \propto \frac{\int_0^{+\infty} s^{q_m + {x''}^{-(i,j)}_m + x_{i,j}} e^{-s(K_T(Y^*_m) + K_N(Y^*_m))} \rho(s) ds}{\int_0^{+\infty} s^{q_m + {x''}^{-(i,j)}_m} e^{-s(K_T(Y^*_m) + K_N(Y^*_m))} \rho(s) ds} \, \frac{1}{x_{i,j}!} H_{i,j}(Y^*_m)^{x_{i,j}} \tag{5.33}
$$

Lastly, if $Y_{i,j,N}$ is different than any element in $\boldsymbol{Y_N}^{-(i,j)}$ and $\boldsymbol{Y_T}$, the full conditional of $V_{r,N}$ for r such that $Y_{i,j,N} = y_{r,N}$ is not degenerate and is given by

$$
\begin{aligned}
&\mathbb{P}[V_{r,N} = v \mid Y_{i,j,N} = y_{r,N}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}^{-(i,j)}, \boldsymbol{V_T}, \boldsymbol{V_N}^{-(i,j)}] \\
&\propto z^{(1-v)} (1-z)^v \int_0^{+\infty} s\, e^{-s(K_T(y_{r,N}) v + K_N(y_{r,N}))} \rho(s) ds
\end{aligned}
\tag{5.34}
$$

Full conditional distributions for the hyperparameters

Until now, we assumed the hyperparameters c and z to be fixed, but one may desire to use some hyperprior distribution for those. We derive here the full conditional distributions for the concentration parameter c and for the dependence parameter z, in the usual case in

which the two parameters are a priori independent

$$\mathcal{L}(z \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}, \boldsymbol{V_T}, \boldsymbol{V_N}, c) \propto \mathcal{L}(z)\,(1-z)^k\, e^{-c\,\psi_z(K_T, K_N)} \tag{5.35}$$

where $\mathcal{L}(z)$ denotes the prior distribution for z. Analogously, we have

$$\mathcal{L}(c \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}, \boldsymbol{V_T}, \boldsymbol{V_N}, z) \propto \mathcal{L}(c)\, c^{k_1+k_2+k}\, e^{-c\,\psi_z(K_T, K_N)} \tag{5.36}$$

Algorithm

Algorithm 2: Algorithm for the estimate of λ_N under general kernel and CRMs

Result: Posterior mean of λ_N
Set initial values for $(, \boldsymbol{Y_T}^{(0)}, \boldsymbol{Y_N}^{(0)}, \boldsymbol{V_T}^{(0)}, \boldsymbol{V_N}^{(0)}, z^{(0)}, c^{(0)})$;
for *iter=1* **to** *burnin+tot_iter* **do**
  Sample $(\boldsymbol{Y_T}^{(iter)}, \boldsymbol{V_T}^{(iter)})$ according to (5.21)-(5.27);
  Sample $(\boldsymbol{Y_N}^{(iter)}, \boldsymbol{V_N}^{(iter)})$ according to (5.28)-(5.34);
  Sample $(c^{(iter)}, z^{(iter)})$ according to (5.35)-(5.36);
  if *iter>burnin* **then**
    Compute $\lambda_N^{(iter)}$ according to (5.14);
  end
end
Compute $\widehat{\lambda_N} = \frac{1}{tot\_iter} \sum \lambda_N^{(iter)}$

5.4.2 GM-dependent gamma CRMs with Ornstein-Uhlenbeck kernel and uniform base-measure

Now, we set $\rho(s)ds = e^{-s}s^{-1}$, $k(t;y) = 2ke^{-k(t-y)}\mathbb{1}_{\{t\geq y\}}$ and $P_0 \sim \mathrm{U}(0,T)$.

Posterior estimate of the event intensity

First of all, we derive the posterior estimate $\widehat{\lambda_N(t)}$ of the intensity function $\lambda_N(t)$, conditionally given the observations $\boldsymbol{T}$ and $\boldsymbol{X}$ and the auxiliary variables $(\boldsymbol{Y_l})_{l\in\{T,N\}}$ and $(\boldsymbol{V_l})_{l\in\{T,N\}}$ under a square loss function.

From Corollary 5.1, we have that

$$\begin{aligned}\widehat{\lambda_N(t)} = &\int_{\mathbb{R}^+\times\mathbb{Y}} s\, k_N(t;\, y) e^{-s\, K_N(y)} \left[(1-z)e^{-s\, K_T(y)} + z\right] \rho(s) ds\, c\, P_0(dy) + \\ &+ \sum_{h=1}^{k_T} V_{h,T}\, k_N(t; Y^*_{h,T}) \int_{\mathbb{R}^+} u f_{J_{h,T}}(u) du + \\ &+ \sum_{r=1}^{k_N} k_N(t; Y^*_{r,N}) \int_{\mathbb{R}^+} u f_{J_{r,N}}(u) du + \\ &+ \sum_{m=1}^{k_0} k_N(t; Y^*_m) \int_{\mathbb{R}^+} u f_{J_k}(u) du\end{aligned} \tag{5.37}$$

We notice that, when $\tilde{\mu}_N$ and $\tilde{\mu}_T$ are gamma CRMs and P_0 is a uniform on $[0, T]$, the first term simplifies as follows.

$$\begin{aligned}&(1-z) \int_{\mathbb{R}^+\times\mathbb{Y}} s\, k(t;\, y) e^{-s\,(K_N(y)+K_T(y))}\, \rho(s) ds\, c\, P_0(dy) \\ &+\, z \int_{\mathbb{R}^+\times\mathbb{Y}} s\, k(t;\, y) e^{-s\, K_N(y)}\, \rho(s) ds\, c\, P_0(dy) \\ =\, & c\, \frac{1-z}{T} \int_{\mathbb{Y}\setminus(T,+\infty)} \frac{k(t;\, y)}{K_N(y) + K_T(y) + 1} dy + c\, \frac{z}{T} \int_{\mathbb{Y}\setminus(T,+\infty)} \frac{k(t;\, y)}{K_N(y) + 1} dy\end{aligned}$$

Substituting the expression of the Ornstein-Uhlenbeck kernel, we get

$$c\, \frac{1-z}{T} \int_0^{\min\{t,T\}} \frac{2k e^{-k(t-y)}}{K_N(y) + K_T(y) + 1} dy + c\, \frac{z}{T} \int_0^{\min\{t,T\}} \frac{2k e^{-k(t-y)}}{K_N(y) + 1} dy$$

Moreover, we have that

$$\begin{aligned}K_N(y) &= \sum_{i=1}^{n} \int_0^{t_{i,m_i}} k(t; y) dt = \sum_{i=1}^{n} \int_0^{t_{i,m_i}} 2k e^{-k(t-y)} \mathbb{1}_{\{t \geq y\}} dt \\ &= \sum_{i=1}^{n} 2k \int_{\min\{\max\{0,y\}, t_{i,m_i}\}}^{t_{i,m_i}} e^{-k(t-y)} dt \\ &= \sum_{i=1}^{n} 2(1 - e^{-k(t_{i,m_i} - y)}) \mathbb{1}_{\{0 \leq y \leq t_{i,m_i}\}} + \sum_{i=1}^{n} 2(e^{ky} - e^{-k(t_{i,m_i} - y)}) \mathbb{1}_{\{y \leq 0\}}\end{aligned}$$

Notice that the terms multiplied by $\mathbb{1}_{\{y\leq 0\}}$ are associated with an event with null probability. Reorder the observations $\{t_{1,m_1},\ldots,t_{n,m_n}\}$ from the smallest to the highest and denote the ordered collection as $\{t^{(1)} < t^{(2)} < \cdots t^{(n)}\}$, e.g. $t^{(1)} = \min\{t_{1,m_1},\ldots,t_{n,m_n}\}$, and set $t^{(0)} = 0$

$$\begin{aligned}\int_0^{\min\{t,T\}} \frac{2k\, e^{-k(t-y)}}{K_N(y)+1}\, dy &= \sum_{i=0}^{n-1} \int_{\min\{t^{(i)},t,T\}}^{\min\{t^{(i+1)},t,T\}} \frac{2k\, e^{-k(t-y)}}{2[(n-i)-\sum_{j=i+1}^{n} e^{-k(t^{(j)}-y)}]+1} dy \\ &= \sum_{i=0}^{n-1}\left(\frac{1}{\sum_{j=i+1}^{n} e^{-k(t^{(j)}-t)}} \log\left(\frac{2(n-i)-2\sum_{j=i+1}^{n} e^{-k(t^{(j)}-\min\{t^{(i)},t,T\})}+1}{2(n-i)-2\sum_{j=i+1}^{n} e^{-k(t^{(j)}-\min\{t^{(i+1)},t,T\})}+1}\right)\right)\end{aligned}$$

In the same spirit, reorder the observations $\{t_{1,m_1},\ldots,t_{n,m_n}\}$ and the values $\{C_1,\ldots,C_n\}$ from the smallest to the highest and denote the ordered collection as $\{a^{(1)} < a^{(2)} < \cdots a^{(2n)}\}$, e.g. $a^{(1)} = \min\{t_{1,m_1},\ldots,t_{n,m_n},C_1,\ldots,C_n\}$, and set $a^{(0)} = 0$.

$$\begin{aligned}&\int_0^{\min\{t,T\}} \frac{2ke^{-k(t-y)}}{K_N(y)+K_T(y)+1} dy = \\ &= \sum_{i=0}^{2n-1}\left(\frac{1}{\sum_{j=i+1}^{2n} e^{-k(a^{(j)}-t)}} \log\left(\frac{2(2n-i)-2\sum_{j=i+1}^{2n} e^{-k(a^{(j)}-\min\{a^{(i)},t,T\})}+1}{2(2n-i)-2\sum_{j=i+1}^{2n} e^{-k(a^{(j)}-\min\{a^{(i+1)},t,T\})}+1}\right)\right)\end{aligned}$$

Using Theorem 5.2 to compute the remaining three terms in (5.37) we have that

$$f_{J_{h,T}}(u) = C^{-1} u^{n_h} e^{-u(K_T(y*_{h,T})+K_N(y*_{h,T})V_{h,T})}\rho(u)$$

where the normalizing constant, with gamma CRMs, equals

$$\begin{aligned}C &= \int_0^{+\infty} u^{n_h} e^{-u(K_T(y^*_{h,T})+K_N(y^*_{h,T})V_{h,T})}\rho(u)du \\ &= \int_0^{+\infty} u^{n_h-1} e^{-u(K_T(y^*_{h,T})+K_N(y^*_{h,T})V_{h,T}+1)}du \\ &= \frac{(n_h-1)!}{(K_T(y^*_{h,T})+K_N(y^*_{h,T})V_{h,T}+1)^{n_h}}\end{aligned}$$

So that

$$\sum_{h=1}^{k_T} V_{h,T}\, k_N(t; Y^*_{h,T}) \int_{\mathbb{R}^+} u f_{J_{h,T}}(u)du = \sum_{h=1}^{k_T} \frac{n_h\, V_{h,T}\, 2ke^{-k(t-Y^*_{h,T})}\mathbb{1}_{\{t\geq Y^*_{h,T}\}}}{K_T(Y^*_{h,T}) + K_N(Y^*_{h,T})V_{h,T} + 1}$$

$$\sum_{r=1}^{k_N} k_N(t; Y^*_{r,N}) \int_{\mathbb{R}^+} u f_{J_{r,N}}(u)du = \sum_{r=1}^{k_N} \frac{x'_r\, 2ke^{-k(t-Y^*_{r,N})}\mathbb{1}_{\{t\geq Y^*_{r,N}\}}}{K_T(Y^*_{r,N})V_{r,N} + K_N(Y^*_{r,N}) + 1}$$

$$\sum_{m=1}^{k_0} k_N(t; Y^*_{m}) \int_{\mathbb{R}^+} u f_{J_{k}}(u)du = \sum_{m=1}^{k_0} \frac{(q_m + x''_m)\, 2ke^{-k(t-Y^*_{m})}\mathbb{1}_{\{t\geq Y^*_{m}\}}}{K_T(Y^*_{m}) + K_N(Y^*_{m}) + 1}$$

Putting everything together we get that the posterior estimate $\widehat{\lambda_N(t)}$ for λ_N is

$$\begin{aligned}
&\frac{c\,(1-z)}{T}\sum_{i=0}^{2n-1}\left(\frac{1}{\sum_{j=i+1}^{2n} e^{-k(a^{(j)}-t)}}\log\left(\frac{2(2n-i)-2\sum_{j=i+1}^{2n} e^{-k(a^{(j)}-\min\{a^{(i)},t,T\})}+1}{2(2n-i)-2\sum_{j=i+1}^{2n} e^{-k(a^{(j)}-\min\{a^{(i+1)},t,T\})}+1}\right)\right)\\
&+\frac{c\,z}{T}\sum_{i=0}^{n-1}\left(\frac{1}{\sum_{j=i+1}^{n} e^{-k(t^{(j)}-t)}}\log\left(\frac{2(n-i)-2\sum_{j=i+1}^{n} e^{-k(t^{(j)}-\min\{t^{(i)},t,T\})}+1}{2(n-i)-2\sum_{j=i+1}^{n} e^{-k(t^{(j)}-\min\{t^{(i+1)},t,T\})}+1}\right)\right)\\
&+\sum_{h=1}^{k_T}\frac{n_h\, V_{h,T}\, 2ke^{-k(t-Y^*_{h,T})}\mathbb{1}_{\{t\geq Y^*_{h,T}\}}}{K_T(Y^*_{h,T}) + K_N(Y^*_{h,T})V_{h,T} + 1}+\sum_{r=1}^{k_N}\frac{x'_r\, 2ke^{-k(t-Y^*_{r,N})}\mathbb{1}_{\{t\geq Y^*_{r,N}\}}}{K_T(Y^*_{r,N})V_{r,N} + K_N(Y^*_{r,N}) + 1}\\
&+\sum_{m=1}^{k_0}\frac{(q_m + x''_m)\, 2ke^{-k(t-Y^*_{m})}\mathbb{1}_{\{t\geq Y^*_{m}\}}}{K_T(Y^*_{m}) + K_N(Y^*_{m}) + 1}
\end{aligned}$$

Full conditional distributions

The full conditional distributions for the vector of latent variables $(Y_{i,j,T}, V_{i,j,T})$ are given by

$$G_0(dy) \propto \mathbb{1}_{\{t_{i,j}\geq y\}}\frac{2ke^{-k(t_{i,j}-y)}}{T}\left(\frac{z}{K_T(y)+1}+\frac{1-z}{K_T(y)+K_N(y)+1}\right)$$

$$w_0 \propto \frac{c\,z}{T}\int_0^{\min\{t_{i,j},T\}} \frac{2ke^{-k(t_{i,j}-y)}}{K_T(y)+1}dy + \frac{c\,(1-z)}{T}\int_0^{\min\{t_{i,j},T\}} \frac{2ke^{-k(t_{i,j}-y)}}{K_T(y)+K_N(y)+1}dy$$

$$=\frac{c\,z}{T}\sum_{\iota=0}^{n-1}\left(\frac{1}{\sum_{\gamma=\iota+1}^{n} e^{-k(t^{(\gamma)}-t_{i,j})}}\log\left(\frac{2(n-\iota)-2\sum_{\gamma=\iota+1}^{n} e^{-k(t^{(\gamma)}-\min\{t^{(\iota)},t_{i,j},T\})}+1}{2(n-\iota)-2\sum_{\gamma=\iota+1}^{n} e^{-k(t^{(\gamma)}-\min\{t^{(\iota+1)},t_{i,j},T\})}+1}\right)\right)$$

$$+\frac{c\,(1-z)}{T}\sum_{\iota=0}^{2n-1}\left(\frac{1}{\sum_{\gamma=\iota+1}^{2n} e^{-k(a^{(\gamma)}-t_{i,j})}}\log\left(\frac{2(2n-\iota)-2\sum_{\gamma=\iota+1}^{2n} e^{-k(a^{(\gamma)}-\min\{a^{(\iota)},t_{i,j},T\})}+1}{2(2n-\iota)-2\sum_{\gamma=\iota+1}^{2n} e^{-k(a^{(\gamma)}-\min\{a^{(\iota+1)},t_{i,j},T\})}+1}\right)\right)$$

$$w_{h,T} \propto \frac{n_h^{-(i,j)}2ke^{-k(t_{i,j}-Y^*_{h,T})}\mathbb{1}_{\{t_{i,j}\geq Y^*_{h,T}\}}}{K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T}+1}$$

$$w_{r,N} \propto (1-z)\frac{x'_r\,2ke^{-k(t_{i,j}-Y^*_{r,N})}\mathbb{1}_{\{t_{i,j}\geq Y^*_{r,N}\}}\mathbb{1}_{\{V_{r,N}=1\}}}{K_T(Y^*_{r,N})V_{r,N}+K_N(Y^*_{r,N})+1}$$

$$w_m \propto \frac{(q_m^{-(i,j)}+x''_m)2ke^{-k(t_{i,j}-Y^*_m)}\mathbb{1}_{\{t_{i,j}\geq Y^*_m\}}}{K_T(Y^*_m)+K_N(Y^*_m)+1}$$

$$\begin{aligned}\mathbb{P}[V_{h,T}=v\mid Y_{i,j,T}=y_{h,T},\boldsymbol{T},\boldsymbol{X},\boldsymbol{Y_T}^{-(i,j)},\boldsymbol{Y_N},\boldsymbol{V_T}^{-(i,j)},\boldsymbol{V_N}]\\ \propto\ z^{(1-v)}\,(1-z)^v\,\frac{1}{K_T(y_{h,T})+K_N(y_{h,T})\,v+1}\end{aligned}$$

The full conditional distributions for the vector of latent variables $(Y_{i,j,N}, V_{i,j,N})$ are

$$H_0(dy) \propto \left[\frac{z}{T\,x_{i,j}}\left(\frac{H_{i,j}(y)}{K_N(y)+1}\right)^{x_{i,j}} + \frac{1-z}{T\,x_{i,j}}\left(\frac{H_{i,j}(y)}{K_N(y)+K_T(y)+1}\right)^{x_{i,j}}\right]\mathbb{1}_{\{y<T\}}$$

$$p_0 \propto \frac{c\,z}{T\,x_{i,j}}\int_0^T\left(\frac{H_{i,j}(y)}{K_N(y)+1}\right)^{x_{i,j}}dy + \frac{c\,(1-z)}{T\,x_{i,j}}\int_0^T\left(\frac{H_{i,j}(y)}{K_N(y)+K_T(y)+1}\right)^{x_{i,j}}dy$$

Now notice that $H_{i,j}(y) = 2\big(\mathbb{1}_{\{y<t_{i,j-1}\}}(e^{-k(t_{i,j-1}-y)}-e^{-k(t_{i,j}-y)})+\mathbb{1}_{\{t_{i,j-1}<y<t_{i,j}\}}(1-e^{-k(t_{i,j}-y)})\big)$,

therefore

$$\int_0^T \left(\frac{H_{i,j}(y)}{K_N(y)+1}\right)^{x_{i,j}} dy =$$

$$= \sum_{\iota=0}^{n-1} \int_{\min\{t^{(\iota)},T\}}^{\min\{t^{(\iota+1)},T\}} \left(\frac{H_{i,j}(y)}{2[(n-\iota)-\sum_{\gamma=\iota+1}^{n} e^{-k(t^{(\gamma)}-y)}]+1}\right)^{x_{i,j}} dy$$

$$= \sum_{\iota=0}^{n-1} \int_{\min\{t^{(\iota)},t_{i,j-1},T\}}^{\min\{t^{(\iota+1)},t_{i,j-1},T\}} \left(\frac{2(e^{-k(t_{i,j-1}-y)}-e^{-k(t_{i,j}-y)})}{2[(n-\iota)-\sum_{\gamma=\iota+1}^{n} e^{-k(t^{(\gamma)}-y)}]+1}\right)^{x_{i,j}} dy$$

$$+ \sum_{\iota=0}^{n-1} \int_{\min\{\max\{\min\{t^{(\iota)},T\},t_{i,j-1}\},t_{i,j}\}}^{\max\{\min\{t^{(\iota+1)},T,t_{i,j}\},t_{i,j-1}\}} \left(\frac{2(1-e^{-k(t_{i,j}-y)})}{2[(n-\iota)-\sum_{\gamma=\iota+1}^{n} e^{-k(t^{(\gamma)}-y)}]+1}\right)^{x_{i,j}} dy$$

whose closed form can be found using the binomial expansion and the result

$$\int \left(\frac{e^{cx}}{a-be^{cx}}\right)^d dx = \frac{(e^{cx}/a)^d \, {}_2F_1(d,d;d+1;\frac{be^{xc}}{a})}{cd} + C$$

$$p_{h,T} \propto (1-z)1, \frac{n_h^{(x_{i,j})}[H_{i,j}(Y^*_{h,T}]^{x_{i,j}}}{x_{i,j}!\,[K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T}+1]^{x_{i,j}}} \mathbb{1}_{\{V_{h,T}=1\}}$$

where $n^{(x)}$ denotes the rising factorial, $n^{(x)} = \frac{(n+x-1)!}{(n-1)!}$

$$p_{r,N} \propto \frac{[x'^{-(i,j)}_r]^{(x_{i,j})}[H_{i,j}(Y^*_{r,N})]^{x_{i,j}}}{x_{i,j}!\,[K_T(Y^*_{r,N})V_{r,N}+K_N(Y^*_{r,N})+1]^{x_{i,j}}}$$

$$p_m \propto \frac{[q_m + x''^{-(i,j)}_m]^{(x_{i,j})}[H_{i,j}(Y^*_m)]^{x_{i,j}}}{x_{i,j}!\,[K_T(Y^*_m)+K_N(Y^*_m)+1]^{x_{i,j}}}$$

Lastly, if $Y_{i,j,N}$ is different than any element in $\boldsymbol{Y}_{\boldsymbol{N}}^{-(i,j)}$ and $\boldsymbol{Y_T}$, the full conditional of $V_{r,N}$ for r such that $Y_{i,j,N} = y_{r,N}$ is not degenerate and is given by

$$\mathbb{P}[V_{r,N} = v \mid Y_{i,j,N} = y_{r,N}, \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y}_{\boldsymbol{N}}^{-(i,j)}, \boldsymbol{V_T}, \boldsymbol{V}_{\boldsymbol{N}}^{-(i,j)}]$$
$$\propto z^{(1-v)}(1-z)^v \frac{1}{K_T(y_{r,N})v + K_N(y_{r,N})}$$

5.5 Simulation study

We provide here three simple simulation studies where we compare inference made through three different approaches: the GM-dependent approach treated in the previous section, the independent approach which preserves the same gamma marginals and OU-kernel for $\tilde{\mu}_T$ and $\tilde{\mu}_N$ and a naive frequentist estimator provided by

$$\lambda_N(t) = \frac{1}{\sum_1^n \mathbb{1}_{\{t \leq t_{i,m_i}\}}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{x_{i,j}}{t_{i,j} - t_{i,j-1}} \mathbb{1}_{\{t_{i,j-1} < t \leq t_{i,j}\}}$$

In the first two simulation studies, we generate data for $n = 10$ and $n = 20$ subjects setting both the observation process and the event process to two Poisson processes with constant intensity equal to 1. Figure 5.1 e Figure 5.2 show the results. In the last simulation study, in Figure 5.3, we generate data for $n = 20$ subjects and both intensities equal to $\exp\{-0.2\, t\}$.

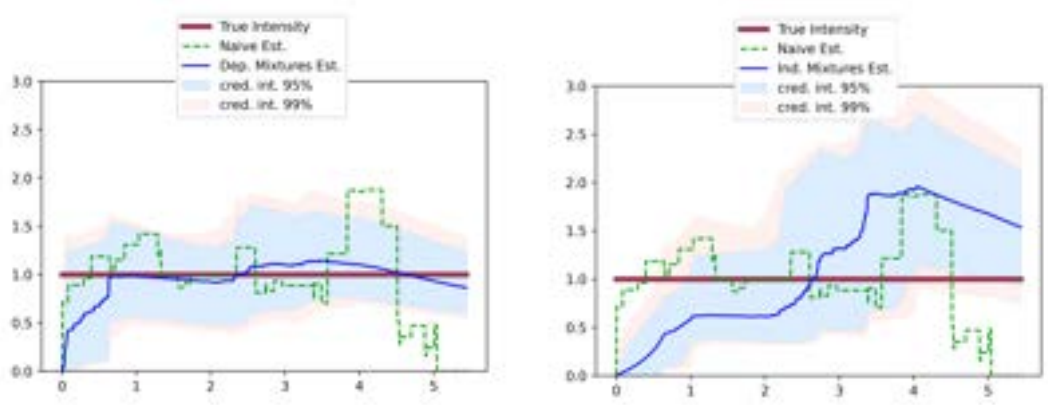


Figure 5.1: Simulation study n.1. Left: GM-dependent model estimate for λ_N. Right: Independent model estimate for λ_N.

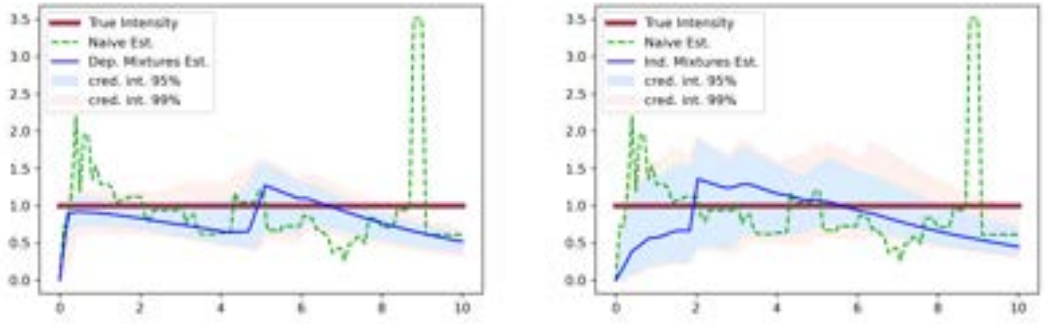


Figure 5.2: Simulation study n.2. Left: GM-dependent model estimate for λ_N. Right: Independent model estimate for λ_N.

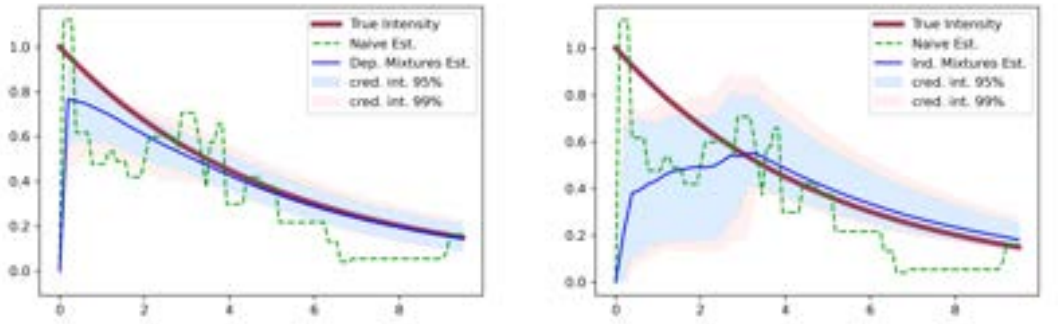


Figure 5.3: Simulation study n.3. Left: GM-dependent model estimate for λ_N. Right: Independent model estimate for λ_N.

The GM-dependent model outperforms the independent approach both in terms of point estimate as well as presenting a smaller uncertainty.

5.6 Concluding remarks

The proposed model allows to account for positive correlation between the two Poisson processes' intensities involved in the generation of panel count data and it performs promisingly in the simulation studies conducted in the previous section. From the point of view of the application, the positive correlation across the CRMs permits to catch for instance the behaviour of those patients that feeling more or less pain will go to doctor's appointments respectively more and less often. However, the results presented in this chapter should be intended only as a first step to extend the use of dependent processes beyond partial exchangeability and to different and more structured data. We plan to extend the model presented here before applying it to real data. As already mentioned, two important developments to consider are a generalizations to different kernels and marginals for the two intensities and the inclusion of subject specific covariates and frialties, more details on both are provided in Chapter 6. Moreover, in some applications, it is easy to imagine the presence of an opposite effect, different from positive correlation, for instance when very sick patients skip an increasing number of clinical visit due to inability to attend or when the event process $N_i(t)$ refers to the number of smoked cigarettes and the counts are self-reported (Moreno et al., 2020). The specifications considered in this chapter ignore the possible presence of negative correlation between the intensities. Therefore an interesting extension will be a generalization of the model that allows for both effects. One way to do so, it is to define two negatively correlated random measures, which does not appear an easy task. In the next chapter we provide an idea to construct such measures which we also plan to explore further in the future.

Chapter 6

Further Extensions

In this last concluding chapter we describe extensions of the works presented in previous chapters. In particular, Section 6.1 briefly describes possible generalization of the class of mSSMs introduced in Chapter 2. Section 6.2 defines FuRBI priors, which are a natural generalization of n-FuRBIs of Chapter 3, obtained relaxing the assumption of independence of increments for the CRV in the product space. Section 6.3 reformulates the model proposed in Chapter 4 in terms of dependent linear regression models with discrete covariates and shows how to modified the invariance conditions to deal with log link functions in a regression setting. Finally, Sections 6.4 and 6.5 provide detailed tools to adapt the model proposed in Chapter 5 in order to respectively consider subjects-level covariates/dependence and incorporate negative-association.

6.1 Extensions of mSSM

The results provided in Chapter 2 may be extended in different ways. Let us recall the generative definition of mSSM in Definition 2.7, which is $\boldsymbol{X} = (X_i^{(j_i)})_{i\geq 1}$ is distributed according to a mSSM if to sample $X_1^{(j_1)}, \ldots, X_n^{(j_n)}$, we can

1. sample a partially exchangeable random partition;
2. associate to each cluster identified at step 1 a value, sampling independent and identically distributed atoms from a non-atomic base measure, independently from the partition sampled at step 1.

Starting from here, we may relax some of the conditions required by the sampling procedure and obtain different laws for the sequence of observations, some examples are provided here below.

Extension 1. If we relax the requirement that the random partition at step 1 being partially exchangeable, many of the results derived in Chapter 2 still holds. Choosing appropriately the law of the partition, models for row-column exchangeable data may be represented and constructed through this procedure.

Extension 2: If instead we relax the hypothesis of independence of the atoms across each other and/or independence of the partition, we may retrieve both n-FURBIs, invariant dependent processes, repulsive mixtures (Petralia et al., 2012; Quinlan et al., 2017; Xie & Xu, 2020) and density regression models.

6.2 From n-FuRBI to FuRBI priors

Recall that if $(\tilde{p}_1, \tilde{p}_2)$ are n-FuRBI on $\mathbb{X}$, then there exist two random probability measures p_1 and p_2 on $\mathbb{X} \times \mathbb{X}$ such that

$$\tilde{p}_1(\cdot) = p_1(\cdot \times \mathbb{X}) \qquad \tilde{p}_2(\cdot) = p_2(\mathbb{X} \times \cdot)$$

and p_1 and p_2 share almost surely the entire sequence of atoms (cf. Definition 3.1). Moreover, notice that tractability of n-FuRBIs is a consequence of the tractability of the objects constructed in the product spaces. Thus, the idea behind n-FURBIs can be generalized also to those cases in which (p_1, p_2) are not normalizations of the coordinates of a CRV and, also more generally, not NRMIs. In particular, we may define FuRBI priors as projections of any mSSP: tractability will be guaranteed as long as the pEPPF is tractable.

6.3 Invariant dependent processes for log link functions

Recall the model introduced in Chapter 4 and consider just one response variable, if we denote with $X_{i,j}$ the i-th observation from the j-th population, the model can be written as a linear regression model as

$$\mathbb{E}[X_{i,j} \mid (\theta_1, \theta_2, \theta_3, \theta_4), \tilde{\Pi}_k^{(N)}, (\xi_c, \sigma_c^2)_{c \geq 1}] = \sum_{k=1}^{4} \theta_k \mathbb{1}_{\{j=k\}} + \xi_{c_{i,j}}$$

where $\tilde{\Pi}_k^{(N)}$ is the partition in (4.14) induced by the s-HDP and $c_{i,j}$ is the label associated to the cluster to which $X_{i,j}$ belongs according to $\tilde{\Pi}_k^{(N)}$. Therefore, $\xi_{c_{i,j}}$ can be seen as a latent random effect and the error term

$$e_{i,j} = X_{i,j} - \mathbb{E}[X_{i,j} \mid (\theta_1, \theta_2, \theta_3, \theta_4), \tilde{\Pi}_k^{(N)}, (\xi_c, \sigma_c^2)_{c \geq 1}]$$

has a covariate-dependent distribution, which induces heteroscedasticity.

Similar proposals for latent random effects can be found in Kleinman & Ibrahim (1998); Guglielmi et al. (2014); Bush & MacEachern (1996); Berger & Tutz (2018).

The model can be extended beyond linearity using a function $f : \mathbb{X} \to \mathbb{X}$ to define the

expected value as

$$\mathbb{E}[X_{i,j} \mid (\theta_1, \theta_2, \theta_3, \theta_4), \tilde{\Pi}_k^{(N)}, (\xi_c)_{c\geq 1}] = f\left(\sum_{k=1}^{4} \theta_k \mathbb{1}_{\{j=k\}} + \xi_{c_{i,j}}\right)$$

where, as in Chapter 4, $\theta_k \mid \tilde{p} \overset{iid}{\sim} \tilde{p}$, $\tilde{p} \sim DP(\alpha, P_0)$ and $\xi_{c_{i,j}} \mid \tilde{q}_j \overset{ind}{\sim} \tilde{q}_j$. However, in order to perform model selection, i.e. in order to estimate the partition of $(\theta_1, \dots, \theta_4)$ without losing identifiability, $(\tilde{q}_1, \dots, \tilde{q}_4)$ needs to satisfy the following invariance condition

$$\int f\left(\sum_{k=1}^{4} \theta_k \mathbb{1}_{\{j=k\}} + x\right) \tilde{q}_j(\mathrm{d}x) \overset{a.s.}{=} f\left(\sum_{k=1}^{4} \theta_k \mathbb{1}_{\{j=k\}}\right) \quad \text{for } j = 1, \dots, 4 \tag{6.1}$$

Trivially, if $f(x) = x$ and $\tilde{q}_j$ is almost surely symmetric around 0 for $j = 1, 2, 3, 4$, (6.1) is satisfied and, in this case, one obtains the model studied in Chapter 4.
When the mean function is $f(x) = \exp\{x\}$, the generalized linear model is said to have a log link function and it is typically used with count data. If $f(x) = \exp\{x\}$ and $\tilde{q}_j$ is almost surely invariant with respect to the identity and the transformation $g(\xi) = \log(2 - \exp\{\xi\})$, for $j = 1, 2, 3, 4$, (6.1) is satisfied and, in this case, one obtains a generalization of the model studied in Chapter 4.

6.4 Dependent priors for panel count data with covariates and frailties

The model presented in Chapter 5 can be generalized in order to model real data adding frailties and considering the possible availability of covariates. Here, we denote with x_i a vector of covariates associated to subject i and we define the model as

$$\{N_i(t) \,:\, t > 0\} \mid \lambda_{N,i}(t) \overset{ind}{\sim} \mathrm{PP}(\,\lambda_{N,i}(t)\,) \qquad \text{for } i = 1, \dots, n$$

$$\{T_i(t) \,:\, t > 0\} \mid \lambda_{T,i}(t) \overset{ind}{\sim} \mathrm{PP}(\,\lambda_{T,i}(t)\,) \qquad \text{for } i = 1, \dots, n$$

$$\lambda_{T,i}(t) = \lambda_{T,0}(t) \exp\{x_i'\gamma + \xi_i\} \qquad \text{with } \lambda_{T,0}(t) = \int_0^{+\infty} \sigma_T \, \kappa \, e^{-\kappa(t-y)} \, \mathbb{1}_{(0,t]}(y) \, \tilde{\mu}_T(dy)$$

$$\lambda_{N,i}(t) = \lambda_{N,0}(t) \exp\{x_i'\beta + \epsilon_i\} \qquad \text{with } \lambda_{N,0}(t) = \int_0^{+\infty} \sigma_N \, \kappa \, e^{-\kappa(t-y)} \, \mathbb{1}_{(0,t]}(y) \, \tilde{\mu}_N(dy)$$

$$p(\gamma) \propto 1 \quad p(\beta) \propto 1 \quad p(\sigma_T) \propto 1 \quad p(\sigma_N) \propto 1$$

$$(\xi_i, \epsilon_i) \mid \Sigma \overset{iid}{\sim} \mathrm{N}_2\Big(0, \Sigma\Big) \qquad \Sigma \sim \mathcal{W}^{-1}\Big(I_d, \nu\Big)$$

$$(\tilde{\mu}_T, \tilde{\mu}_N) \mid c,\, z \overset{d}{=} \text{GM-dependent CRMs}(c,\, z,\, \rho(s) = e^{-s}s^{-1},\, P_0)$$

$$c \sim \text{Gamma}(\alpha_c, \beta_c)$$

$$z \sim \text{Beta}(\alpha_z, \beta_z)$$

Setting $\theta = (\tilde{\mu}_N,\, \tilde{\mu}_T,\, \beta, \gamma,\, \sigma_T,\, \sigma_N,\, \xi_1,\, \ldots,\, \xi_n,\, \epsilon_1,\, \ldots,\, \epsilon_n,\ c,\, z)$, the likelihood function corresponding to the model is

$$\begin{aligned}\mathcal{L}(\theta\,; \boldsymbol{t}, \boldsymbol{x}) = & e^{-\int_0^{+\infty} K_T(y)\tilde{\mu}_T(dy)}\, e^{-\int_0^{+\infty} K_N(y)\tilde{\mu}_N(dy)} \times \\ & \exp\left\{\sum_{i=1}^{n}[m_i\,(x_i'\gamma + \xi_i) + N_{i,m_i}\,(x_i'\beta + \epsilon_i)]\right\} \times \\ & \prod_{i=1}^{n}\prod_{j=1}^{m_i}\left[\int_0^{t_{i,j}} \sigma_T\,\kappa\, e^{-\kappa(t_{i,j}-y)}\,\tilde{\mu}_T(dy)\,\frac{1}{x_{i,j}!}\left(\int_{\mathbb{Y}} H_{i,j}(y)\tilde{\mu}_N(dy)\right)^{x_{i,j}}\right]\end{aligned}$$

where

$$K_T(y) = \sigma_T \sum_{i=1}^{n}\left[\exp\{x_i'\gamma + \xi_i\}\,(1 - e^{-\kappa(C_i - y)})\mathbb{1}_{(0,C_i]}(y)\right]$$

$$K_N(y) = \sigma_N \sum_{i=1}^{n}\left[\exp\{x_i'\beta + \epsilon_i\}\,(1 - e^{-\kappa(t_{i,m_i} - y)})\mathbb{1}_{(0,t_{i,m_i}]}(y)\right]$$

and

$$H_{i,j}(y) = \begin{cases} \sigma_N(e^{-\kappa(t_{i,j-1}-y)} - e^{-\kappa(t_{i,j}-y)}), & \text{if } y < t_{i,j-1} \\ \sigma_N(1 - e^{-\kappa(t_{i,j}-y)}), & \text{if } t_{i,j-1} < y < t_{i,j} \\ 0, & \text{otherwise} \end{cases}$$

Setting $\theta^* = (\beta, \gamma,\, \sigma_T,\, \sigma_N,\, \xi_1,\, \ldots,\, \xi_n,\, \epsilon_1,\, \ldots,\, \epsilon_n,\, c,\, z)$ and applying the same arguments of Theorem 5.2 we get the posterior distribution provided in next theorem.

Theorem 6.1. *The posterior distribution of $\tilde{\mu}_T$ and $\tilde{\mu}_N$, conditional on $\boldsymbol{T}$, $\boldsymbol{X}$, $\boldsymbol{Y_T}$, $\boldsymbol{Y_N}$, $\boldsymbol{V_T}$, $\boldsymbol{V_N}$ and θ^*, equals the distribution of the vector of CRMs*

$$\begin{aligned}(\tilde{\mu}_T^*, \tilde{\mu}_N^*) & + \left(\sum_{h=1}^{k_T} J_{h,T}\,\delta_{Y^*_{h,T}}, \sum_{h=1}^{k_T} J_{h,T}\,V_{h,T}\,\delta_{Y^*_{h,T}}\right) \\ & + \left(\sum_{r=1}^{k_N} J_{h,N}\,V_{h,N}\,\delta_{Y^*_{h,N}}, \sum_{r=1}^{k_N} J_{h,N}\,\delta_{Y^*_{h,N}}\right) \\ & + \left(\sum_{m=k}^{k} J_m\,\delta_{Y^*_m}, \sum_{m=1}^{k} J_m\,\delta_{Y^*_m}\right)\end{aligned}$$

where $\tilde{\mu}^*_T$ and $\tilde{\mu}^*_N$ are CRMs such that

$$\tilde{\mu}^*_T \stackrel{d}{=} \mu^*_0 + \mu^*_T$$
$$\tilde{\mu}^*_N \stackrel{d}{=} \mu^*_0 + \mu^*_N$$

where μ^*_0, μ^*_T and μ^*_N are independent CRMs with Lévy intensities respectively equal to

$$\begin{aligned}
\nu^*_0(ds, dy) =& c\,(1-z)\,e^{-s\,(K_T(y)+K_N(y)+1)} P_0(dy)\,s^{-1}ds \\
\nu^*_T(ds, dy) =& c\,z\,e^{-s\,(K_T(y)+1)} P_0(dy)\,s^{-1}ds \\
\nu^*_N(ds, dy) =& c\,z\,e^{-s\,(K_N(y)+1)} P_0(dy)\,s^{-1}ds
\end{aligned}$$

The jumps $J_{1,T}, \ldots, J_{k_T,T}$, $J_{1,N}, \ldots, J_{k_N,N}$ and $J_1, \ldots, J_k$ are mutually independent and independent from $\tilde{\mu}^*_T$ and $\tilde{\mu}^*_N$ and have densities

$$\begin{aligned}
f_{J_{h,T}}(s) &\propto s^{n_h-1}\,e^{-s\,(K_T(Y^*_{h,T})+K_N(Y^*_{h,T})V_{h,T}+1)}ds \equiv \textit{Gamma}(n_h, K_T(Y^*_{h,T}) + K_N(Y^*_{h,T})V_{h,T} + 1) \\
f_{J_{r,N}}(s) &\propto s^{x'_r-1}\,e^{-s\,(K_T(Y^*_{r,N})V_{r,N}+K_N(Y^*_{h,T})+1)}ds \equiv \textit{Gamma}(x'_r, K_T(Y^*_{r,N})V_{r,N} + K_N(Y^*_{h,T}) + 1) \\
f_{J_k}(s) &\propto s^{q_m+x''_m-1}\,e^{-s\,(K_T(Y^*_m)+K_N(Y^*_m)+1)}ds \equiv \textit{Gamma}(q_m + x''_m, K_T(Y^*_m) + K_N(Y^*_m) + 1)
\end{aligned}$$

where Gamma(a, b) denotes a Gamma distribution with expected value a/b.

From Theorem 6.1 we have that

$$\begin{aligned}
\widehat{\lambda}_{N,0}(t) &= \mathbb{E}\left[\int_0^{+\infty} \sigma_N\,\kappa\,e^{-\kappa(t-y)}\,\mathbb{1}_{(0,t]}(y)\,\tilde{\mu}_N(dy) \mid \boldsymbol{T}, \boldsymbol{X}, \boldsymbol{Y_T}, \boldsymbol{Y_N}, \boldsymbol{V_T}, \boldsymbol{V_N}, \theta^*\right] = \\
&= c\,(1-z)\int_{\mathbb{R}^+}\int_0^t s\,\sigma_N\,\kappa\,e^{-\kappa(t-y)}\,e^{-s\,(K_T(y)+K_N(y)+1)}P_0(dy)\,s^{-1}ds + \\
&+ c\,z\int_{\mathbb{R}^+}\int_0^t s\,\sigma_N\,\kappa\,e^{-\kappa(t-y)}\,e^{-s\,(K_N(y)+1)}P_0(dy)\,s^{-1}ds + \\
&+ \sum_{h=1}^{k_T} V_{h,T}\,\sigma_N\,\kappa\,e^{-\kappa(t-Y^*_{h,T})}\,\mathbb{1}_{(0,t]}(Y^*_{h,T})\,\mathbb{E}\left[J_{h,T}\right] + \\
&+ \sum_{r=1}^{k_N} \sigma_N\,\kappa\,e^{-\kappa(t-Y^*_{r,N})}\,\mathbb{1}_{(0,t]}(Y^*_{r,N})\mathbb{E}\left[J_{r,N}\right] + \\
&+ \sum_{m=1}^{k} \sigma_N\,\kappa\,e^{-\kappa(t-Y^*_m)}\,\mathbb{1}_{(0,t]}(Y^*_m)\mathbb{E}\left[J_m\right]
\end{aligned}$$

where, the first term is

$$c\,(1-z)\int\limits_{\mathbb{R}^+}\int\limits_{0}^{t} s\,\sigma_N\,\kappa\,e^{-\kappa(t-y)}\,e^{-s\,(K_T(y)+K_N(y)+1)}P_0(dy)\,s^{-1}ds =$$

$$= \frac{c\,(1-z)\,\sigma_N\,\kappa}{T}\int\limits_{0}^{t}\int\limits_{\mathbb{R}^+} e^{-[\kappa(t-y)+s\,(K_T(y)+K_N(y)+1)]}ds\,dy =$$

$$= \frac{c\,(1-z)\,\sigma_N\,\kappa}{T}\int\limits_{0}^{t}\frac{e^{-\kappa(t-y)}}{K_T(y)+K_N(y)+1}dy$$

Consider now: $K_T(y)+K_N(y)+1$

$$\begin{aligned}K_T(y)+K_N(y)+1 =& \sigma_T\sum_{i=1}^{n}\exp\{x_i'\gamma+\xi_i\}\,(1-e^{-\kappa(C_i-y)})\mathbb{1}_{(0,C_i]}(y)\\ &+\sigma_N\sum_{i=1}^{n}\exp\{x_i'\beta+\epsilon_i\}\,(1-e^{-\kappa(t_{i,m_i}-y)})\mathbb{1}_{(0,t_{i,m_i}]}(y)+1\end{aligned}$$

Reorder the observations $\{t_{1,m_1},\ldots,t_{n,m_n}\}$ and the values $\{C_1,\ldots,C_n\}$ from the highest to the smallest and denote the ordered collection as $\{a^{(1)}>a^{(2)}>\cdots>a^{(2n)}\}$, e.g. $a^{(1)}=\max\{t_{1,m_1},\ldots,t_{n,m_n},C_1,\ldots,C_n\}$.
Moreover, create the vector w defined according to $w_i=1$ if $a^{(i)}\in\{C_1,\ldots,C_n\}$ and $w_i=0$ otherwise and a duplicate and reorder the covariates and error terms in such a way that they correspond to the new order induces by the $a^{(i)}$,we denote this new ordered terms $x^{(i)}$, $\xi^{(i)}$ and $\epsilon^{(i)}$. We have that

$$K_T(y)+K_N(y)+1 \;=\; 1 \quad \text{when } a^{(1)}<y$$

$$\begin{aligned}K_T(y)+K_N(y)+1 \;=\;& \sigma_T\,w_1\,\exp\{x^{(1)}\gamma+\xi^{(1)}\}\,(1-e^{-\kappa(a^{(1)}-y)})+\\ &+\sigma_N\,(1-w_1)\,\exp\{x^{(1)}\beta+\epsilon^{(i)}\}\,(1-e^{-\kappa(a^{(1)}-y)})+1 \quad \text{when } a^{(2)}<y<a^{(1)}\end{aligned}$$

$$\begin{aligned}K_T(y)+K_N(y)+1 \;=\;& \sigma_T\sum_{i=1}^{2}w_i\,\exp\{x^{(i)}\gamma+\xi^{(i)}\}(1-e^{-\kappa(a^{(i)}-y)})+\\ &+\sigma_N\sum_{i=1}^{2}(1-w_i)\,\exp\{x^{(i)}\beta+\epsilon^{(i)}\}\,(1-e^{-\kappa(a^{(i)}-y)})+1 \quad \text{when } a^{(3)}<y<a^{(2)}\\ &\ldots\end{aligned}$$

$$K_T(y) + K_N(y) + 1 \;=\; \sigma_T \sum_{i=1}^{j} w_i \, \exp\{x^{(i)}\gamma + \xi^{(i)}\} \, (1 - e^{-\kappa(a^{(i)} - y)}) +$$

$$+ \, \sigma_N \sum_{i=1}^{j} (1 - w_i) \, \exp\{x^{(i)}\beta + \epsilon^{(i)}\} \, (1 - e^{-\kappa(a^{(i)} - y)}) + 1 \quad \text{when } \; a^{(j+1)} < y < a^{(j)}$$

$$\dots$$

$$K_T(y) + K_N(y) + 1 \;=\; \sigma_T \sum_{i=1}^{2n} w_i \, \exp\{x^{(i)}\gamma + \xi^{(i)}\} \, (1 - e^{-\kappa(a^{(i)} - y)}) +$$

$$+ \, \sigma_N \sum_{i=1}^{2n} (1 - w_i) \, \exp\{x^{(i)}\beta + \epsilon^{(i)}\} \, (1 - e^{-\kappa(a^{(i)} - y)}) + 1 \quad \text{when } \; 0 < y < a^{(2n)}$$

Therefore, setting $a^{(0)} = +\infty$, $a^{(2n+1)} = 0$ and $\sum_{i=1}^{0} = 0$

$$\int_0^t \frac{e^{-\kappa(t-y)}}{K_T(y) + K_N(y) + 1} dy =$$

$$\sum_{j=0}^{2n} \int_{\min\{a^{(j+1)}, t\}}^{\min\{a^{(j)}, t\}} \left[e^{-\kappa(t-y)} \left(\sigma_T \sum_{i=1}^{j} w_i \exp\{x^{(i)}\gamma + \xi^{(i)}\}(1 - e^{-\kappa(a^{(i)} - y)}) + \right. \right.$$

$$\left. \left. \sigma_N \sum_{i=1}^{j} (1 - w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\}(1 - e^{-\kappa(a^{(i)} - y)}) + 1 \right)^{-1} \right] dy$$

$$= \sum_{j=0}^{2n} \frac{1}{\kappa \sum_{i=1}^{j} \left(W_i \, e^{-\kappa \, (a^{(i)} - t)} \right)} \, log \left(\frac{1 + \sum_{i=1}^{j} W_i (1 - e^{-\kappa \, (a^{(i)} - \min\{a^{(j+1)}, t\})})}{1 + \sum_{i=1}^{j} W_i (1 - e^{-\kappa \, (a^{(i)} - \min\{a^{(j)}, t\})})} \right)$$

where $W_i = \sigma_T \, w_i \exp\{x^{(i)}\gamma + \xi^{(i)}\} + \sigma_N \, (1 - w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\}$.
Indeed, notice that

$$\int_{x_1}^{x_2} \frac{e^{a\,x}}{b - c\, e^{a\,x}} dx = \left[- \frac{log(b - c\, e^{a\,x})}{a\, c} \right]_{x_1}^{x_2} = \frac{log\left(\frac{b - c\, e^{a\,x_1}}{b - c\, e^{a\,x_2}} \right)}{a\, c}$$

where one still has to multiply by $e^{-\kappa\, t}$ and set

$$a = \kappa$$

$$b = \sigma_T \sum_{i=1}^{j} w_i \exp\{x^{(i)}\gamma + \xi^{(i)}\} + \sigma_N \sum_{i=1}^{j} (1 - w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\} + 1 = \sum_{i=1}^{j} W_i + 1$$

$$c = \sum_{i=1}^{j} \left[(\sigma_T w_i \exp\{x^{(i)}\gamma + \xi^{(i)}\} + \sigma_N (1-w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\})\, e^{-\kappa\, a^{(i)}} \right] = \sum_{i=1}^{j} \left(W_i e^{-\kappa\, a^{(i)}} \right)$$

So that finally the first term is equal to

$$\frac{c\,(1-z)\,\sigma_N}{T} \sum_{j=0}^{2n} \frac{1}{\sum_{i=1}^{j} \left(W_i\, e^{-\kappa\,(a^{(i)}-t)} \right)} log \left(\frac{1 + \sum_{i=1}^{j} W_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j+1)}, t\})})}{1 + \sum_{i=1}^{j} W_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j)}, t\})})} \right)$$

Analogously one can compute the second term as follows

$$\begin{aligned}
& c\, z \int_{\mathbb{R}^+} \int_{0}^{t} s\, \sigma_N\, \kappa\, e^{-\kappa(t-y)}\, e^{-s\,(K_N(y)+1)} P_0(dy)\, s^{-1} ds = \\
& = \frac{c\, z\, \sigma_N\, \kappa}{T} \int_{0}^{t} \frac{e^{-\kappa(t-y)}}{K_N(y)+1} dy = \\
& = \frac{c\, z\, \sigma_N}{T} \sum_{j=0}^{2n} \frac{1}{\sum_{i=1}^{j} \left(S_i\, e^{-\kappa\,(a^{(i)}-t)} \right)} log \left(\frac{1 + \sum_{i=1}^{j} S_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j+1)}, t\})})}{1 + \sum_{i=1}^{j} S_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j)}, t\})})} \right)
\end{aligned}$$

where $S_i = \sigma_N\,(1-w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\}$. While the remaining three terms are

$$\begin{aligned}
& \sum_{h=1}^{k_T} V_{h,T}\, \sigma_N\, \kappa\, e^{-\kappa(t - Y^*_{h,T})}\, \mathbb{1}_{(0,t]}(Y^*_{h,T})\, \mathbb{E}\,[J_{h,T}] = \\
& = \sum_{h=1}^{k_T} V_{h,T}\, \frac{n_h\, \sigma_N\, \kappa}{K_T(Y^*_{h,T}) + K_N(Y^*_{h,T}) V_{h,T} + 1}\, e^{-\kappa(t - Y^*_{h,T})}\, \mathbb{1}_{(0,t]}(Y^*_{h,T}) \\
& \sum_{r=1}^{k_N} \sigma_N\, \kappa\, e^{-\kappa(t - Y^*_{r,N})}\, \mathbb{1}_{(0,t]}(Y^*_{r,N}) \mathbb{E}\,[J_{r,N}] = \\
& = \sum_{r=1}^{k_N} \frac{x'_r\, \sigma_N\, \kappa}{K_T(Y^*_{r,N}) V_{r,N} + K_N(Y^*_{r,N}) + 1}\, e^{-\kappa(t - Y^*_{r,N})}\, \mathbb{1}_{(0,t]}(Y^*_{r,N}) \\
& \sum_{m=1}^{k} \sigma_N\, \kappa\, e^{-\kappa(t - Y^*_{m})}\, \mathbb{1}_{(0,t]}(Y^*_{m}) \mathbb{E}\,[J_{m}] = \\
& = \sum_{m=1}^{k_0} \frac{(q_m + x''_m)\, \sigma_N\, \kappa}{K_T(Y^*_m) + K_N(Y^*_m) + 1}\, e^{-\kappa(t - Y^*_m)}\, \mathbb{1}_{(0,t]}(Y^*_m)
\end{aligned}$$

Finally, the conditional expected value of $\lambda_{N,0}$ is

$$\begin{aligned}
\widehat{\lambda}_{N,0}(t) = & \frac{c\,(1-z)\,\sigma_N}{T} \sum_{j=0}^{2n} \frac{1}{\sum_{i=1}^{j} \left(W_i\, e^{-\kappa\,(a^{(i)}-t)} \right)} \, log \left(\frac{1 + \sum_{i=1}^{j} W_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j+1)}, t\})})}{1 + \sum_{i=1}^{j} W_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j)}, t\})})} \right) \\
& + \frac{c\,z\,\sigma_N}{T} \sum_{j=0}^{2n} \frac{1}{\sum_{i=1}^{j} \left(S_i\, e^{-\kappa\,(a^{(i)}-t)} \right)} \, log \left(\frac{1 + \sum_{i=1}^{j} S_i (1 - e^{-\kappa\,(a^{(i)} - \min\{a^{(j+1)}, t\})})}{1 + \sum_{i=1}^{j} S_i (1 - e^{-\kappa\,(a^{(i)} - \min\{t^{(j)}, t\})})} \right) \\
& + \sum_{h=1}^{k_T} V_{h,T} \frac{n_h\,\sigma_N\,\kappa}{K_T(Y^*_{h,T}) + K_N(Y^*_{h,T}) V_{h,T} + 1} \, e^{-\kappa(t - Y^*_{h,T})} \, \mathbb{1}_{(0,t]}(Y^*_{h,T}) \\
& + \sum_{r=1}^{k_N} \frac{x'_r\,\sigma_N\,\kappa}{K_T(Y^*_{r,N}) V_{r,N} + K_N(Y^*_{r,N}) + 1} \, e^{-\kappa(t - Y^*_{r,N})} \, \mathbb{1}_{(0,t]}(Y^*_{r,N}) \\
& + \sum_{m=1}^{k_0} \frac{(q_m + x''_m)\,\sigma_N\,\kappa}{K_T(Y^*_m) + K_N(Y^*_m) + 1} \, e^{-\kappa(t - Y^*_m)} \, \mathbb{1}_{(0,t]}(Y^*_m)
\end{aligned}$$

where $W_i = \sigma_T\, w_i \exp\{x^{(i)}\gamma + \xi^{(i)}\} + \sigma_N\,(1 - w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\}$ and $S_i = \sigma_N\,(1 - w_i) \exp\{x^{(i)}\beta + \epsilon^{(i)}\}$.

6.5 Random measures with signed correlation

As anticipated at the end of Chapter 5, a nice tool to model dependence in panel count data would be a pair of random measures that may display negative correlation. More precisely the goal of this section is to define two random measures $\tilde{\mu}_1$ and $\tilde{\mu}_2$ on $(\mathbb{X}, \mathcal{X})$ that may display negative correlation when evaluated over the same Borel set, i.e.

$$\text{Cov}[\tilde{\mu}_1(A), \tilde{\mu}_2(A)] < 0$$

The main idea we employ is to use partially exchangeable sequences as atoms of the random measures, instead of i.i.d. sequences as it usually happens CRMs.

Let $M_{\mathbb{X}}$ be the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$ equipped with corresponding Borel σ-algebra $\mathcal{M}_{\mathbb{X}}$. Consider random elements $\tilde{\mu}$ taking values in $(M_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ such that $\tilde{\mu}$ is almost surely discrete and has no fixed point of discountinuity: $\tilde{\mu} = \sum_{k \geq 1} J_k \delta_{\theta_k}$.

Definition 6.1. *Two random measures $\tilde{\mu}_1$ and $\tilde{\mu}_2$ on $(\mathbb{X}, \mathcal{X})$ are called random measures with signed correlation if and only if*

$$\tilde{\mu}_1 \overset{a.s.}{=} \sum_{k \geq 1} J_k \delta_{(\theta_{1,k})} \qquad\qquad \tilde{\mu}_2 \overset{a.s.}{=} \sum_{k \geq 1} W_k \delta_{(\theta_{2,k})}$$

with

$$\theta_{j,k} \mid \theta_{j,0} \overset{ind}{\sim} \tilde{H}(\cdot\,;\theta_{j,0}) \qquad \textit{for } j = 1,2$$
$$(\theta_{1,0}, \theta_{2,0}) \sim H_0(\cdot\,,\cdot)$$

where $\tilde{H}$ is a random probability measure and H_0 is a diffuse probability measures.

Notice that, by de Finetti theorem of partial exchangeability, Definition 6.1 implies that the two sequences of atoms are partially exchangeable. For now we are also going to assume that $(J_k, W_k)_{k\geq 1} \perp (\theta_{1,k}, \theta_{2,k})_{k\geq 1}$ and we say that the random measures are homogenous. Notice that when H_0 is degenerate on a single point where $\theta_{1,0} = \theta_{2,0}$, i.e. $H_0 = \delta_{\{x,x\}}$, many popular constructions of dependent completely random measures can be recovered, as GM-dependent CRMs, hierarchical CRMs, compound random measures, etc. However, exept when H_0 is degenerate on some point, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are not CRMs, but mixtures of CRMs as shown by their marginal Laplace functional

Proposition 6.1. *Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be random measures with signed correlation, then for every measurable function f from $\mathbb{X}$ to $\mathbb{R}^+$, the marginal laplace functional transform of $\tilde{\mu}_j$ is given by*

$$\mathbb{E}[e^{-\tilde{\mu}_j(f)}] = \int_{\Theta} \exp\left\{ -\theta \int_{\mathbb{X}} \int_{\mathbb{R}^+} (1 - e^{-s\, f(x)})\, \rho_j(s) ds\, \tilde{H}(dx\,;\theta_{j,0}) \right\} H_j(d\theta_{j,0})$$

where we use $\tilde{\mu}(f)$ to denote $\int_{\mathbb{X}} f(x)\tilde{\mu}(dx)$ and H_j to denote the marginal probability distribution induce by H_0 on the j coordinate.

Proof. Proof is straightforward using tower rule conditioning on $\theta_{j,0}$. □

Theorem 6.2. *Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be homogeneous random measures with signed correlation, then*

$$Cov(\tilde{\mu}_1(A), \tilde{\mu}_2(B)) = \beta\, Cov(\tilde{\mu}_1(\mathbb{X}),\, \tilde{\mu}_2(\mathbb{X})) + \gamma\, Cov_{H_0}(\tilde{H}(A;\theta_{1,0}),\, \tilde{H}(B;\theta_{2,0}))$$

where

$$\beta = \mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0}), \tilde{H}(B;\theta_{2,0})] \textit{ and } \gamma = \mathbb{E}[\tilde{\mu}_1(\mathbb{X})]\mathbb{E}[\tilde{\mu}_2(\mathbb{X})]$$

or equivalently

$$\beta = \mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})]\mathbb{E}_{H_0}[\tilde{H}(B;\theta_{2,0})] \textit{ and } \gamma = \mathbb{E}[\tilde{\mu}_1(\mathbb{X})\, \tilde{\mu}_2(\mathbb{X})]$$

Proof.

$$\mathrm{Cov}(\tilde{\mu}_1(A), \tilde{\mu}_2(B)) = \mathbb{E}[\mathbb{E}[\tilde{\mu}_1(A)\, \tilde{\mu}_2(B) \mid \theta_{1,0}, \theta_{2,0}]\,] - \mathbb{E}[\mathbb{E}[\tilde{\mu}_1(A) \mid \theta_{1,0}]\,]\, \mathbb{E}[\mathbb{E}[\tilde{\mu}_2(B) \mid \theta_{2,0}]\,]$$

where

$$\begin{aligned}\mathbb{E}[\mathbb{E}[\tilde{\mu}_1(A)\,\tilde{\mu}_2(B)\mid\theta_{1,0},\theta_{2,0}]\,] =&\mathbb{E}[\mathbb{E}[\sum_{k\geq1}J_k\delta_{(\theta_{1,k})}(A)\sum_{k\geq1}W_k\delta_{(\theta_{2,k})}(B)\mid\theta_{1,0},\theta_{2,0}]\,]\\ =&\mathbb{E}[\sum_{k\geq1}J_k\tilde{H}(A;\theta_{1,0})\sum_{k\geq1}W_k\tilde{H}(B;\theta_{2,0})]\\ =&\mathbb{E}[\sum_{k\geq1}J_k\sum_{k\geq1}W_k]\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})\tilde{H}(B;\theta_{2,0})]\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mathbb{E}[\tilde{\mu}_1(A)\mid\theta_{1,0}]\,] =&\mathbb{E}[\mathbb{E}[\sum_{k\geq1}J_k\delta_{(\theta_{1,k})}(A)]\,]\\ =&\mathbb{E}[\sum_{k\geq1}J_k\tilde{H}(A;\theta_{1,0})]\\ =&\mathbb{E}[\sum_{k\geq1}J_k]\,\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})]\end{aligned}$$

therefore

$$\begin{aligned}\mathrm{Cov}(\tilde{\mu}_1(A),\tilde{\mu}_2(B)) =&\mathbb{E}[\sum_{k\geq1}J_k\sum_{k\geq1}W_k]\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})\tilde{H}(B;\theta_{2,0})]\\ &-\mathbb{E}[\sum_{k\geq1}J_k]\,\mathbb{E}[\sum_{k\geq1}W_k]\,\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})]\,\mathbb{E}_{H_0}[\tilde{H}(B;\theta_{2,0})]\\ &-\mathbb{E}[\sum_{k\geq1}J_k]\,\mathbb{E}[\sum_{k\geq1}W_k]\,\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})\tilde{H}(B;\theta_{2,0})]\\ &+\mathbb{E}[\sum_{k\geq1}J_k]\,\mathbb{E}[\sum_{k\geq1}W_k]\,\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})\tilde{H}(B;\theta_{2,0})]\\ =&\,\mathbb{E}_{H_0}[\tilde{H}(A;\theta_{1,0})\tilde{H}(B;\theta_{2,0})]\,Cov(\sum_{k\geq1}J_k,\sum_{k\geq1}W_k)\\ &+\mathbb{E}[\sum_{k\geq1}J_k]\,\mathbb{E}[\sum_{k\geq1}W_k]\,Cov_{H_0}(\tilde{H}(A;\theta_{1,0}),\,\tilde{H}(B;\theta_{2,0}))\end{aligned}$$

□

Theorem 6.2 shows that the covariance between the two CRMs with signed correlation can be decomposed into two terms, whose signs are given respectively by the covariance between the total masses (i.e. by the weights) and the covariance induced on the atoms.

Corollary 6.1. *Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be random measures with signed correlation such that $(J_k)_{k\geq1}$ and $(W_k)_{k\geq1}$ are independent, then*

$$Cov(\tilde{\mu}_1(A),\tilde{\mu}_2(B)) = \theta^2\,Cov_{H_0}(\tilde{H}(A;\theta_{1,0}),\,\tilde{H}(B;\theta_{2,0}))$$

Proof. The proof is immediate applying Theorem 6.2. □

Theorem 6.3. *Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be homogeneous random measures with signed correlation such that $(J_k)_{k\geq 1}$ and $(W_k)_{k\geq 1}$ are independent, then for every pair of measurable function f_1 and f_2 from $\mathbb{X}$ to $\mathbb{R}^+$, the joint Laplace functional transform of $\tilde{\mu}_1$ and $\tilde{\mu}_2$ is given by*

$$\mathbb{E}[e^{-\tilde{\mu}_1(f_1)-\tilde{\mu}_2(f_2)}] = \int_{\Theta^2} \exp\Big\{ -\theta^2 \Big(\int_{\mathbb{X}} \int_{\mathbb{R}^+} (1-e^{-s\, f_1(x)})\, \rho_1(s) ds\, \tilde{H}(dx\,;\theta_{1,0}) \\ + \int_{\mathbb{X}} \int_{\mathbb{R}^+} (1-e^{-s\, f_2(x)})\, \rho_2(s) ds\, \tilde{H}(dx\,;\theta_{2,0}) \Big) \Big\} H_0(d\theta_{1,0}, d\theta_{2,0})$$

Proof. Proof is straightforward using tower rule conditioning on $\theta_{1,0}$ and $\theta_{2,0}$. □

This Page Intentionally Left Blank

Appendix A

Finite Dirichlet distribution

A.1 Some properties of the Dirichlet distribution

The Dirichlet distribution defined in Definition 1.2 admits also a famous representation in terms of Gamma random variables, according to the following proposition.

Proposition (A.1). *Consider k independent random variables $Y_1, \dots, Y_k$ such that $Y_i \overset{ind}{\sim}$ Gamma(α_i) and define*

$$p_i = \frac{Y_i}{\sum_{i=1}^{k} Y_i} \qquad \text{for } i = 1, \dots, k$$

then $(p_1, \dots, p_k) \sim D_{k-1}(\alpha_1, \dots, \alpha_k)$.

where $X \sim \text{Gamma}(a, b)$ denotes a Gamma distributed random variable X parameterized such that $\mathbb{E}[X] = a/b$. Moreover the marginals of a Dirichlet distribution are Beta. Denote with α_0 the sum of the parameters, i.e. $\alpha_0 := \sum_{i=1}^{k} \alpha_i$.

Proposition (A.2). *Consider $(p_1, \dots, p_k) \sim D_{k-1}(\alpha_1, \dots, \alpha_k)$ then*

$$p_i \sim \text{Beta}(\alpha_i,\ \alpha_0 - \alpha_i)$$

From proposition A.2 is immediate to obtain the first and second marginal moments of the Dirichlet distribution given by

$$\mathbb{E}[p_i] = \frac{\alpha_i}{\alpha_0} \qquad \text{for } i = 1, ..., k$$

$$\text{Var}[p_i] = \frac{(\alpha_0 - \alpha_i)\alpha_i}{\alpha_0^2(\alpha_0 + 1)} \qquad \text{for } i = 1, ..., k$$

Moreover, the covariance is

$$\text{Cov}[p_i, p_j] = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)} \qquad \text{for } i \neq j$$

Notice that the variance of each coordinate of $(p_1, \ldots, p_k)$ can be rearranged in the following way

$$\text{Var}[p_i] = \mathbb{E}[p_i]\,(1 - \mathbb{E}[p_i])\left(\frac{1}{\alpha_0 + 1}\right)$$

where it is evident that, for fixed expected values of the coordinates, the higher the sum α_0, the smaller the variance of the prior. More details about the Dirichlet distribution can be found in many Bayesian statistics textbooks, as for instance, Bernardo & Smith (2009, p. 134-136) and Ghosal & Van der Vaart (2017, p. 562-569).

A.2 Multinomial-Dirichlet Model

Given a population with k mutually exclusive and collectively exhaustive categories. One can arbitrarily identify each category with a natural number, in such a way that when a random sample $X_1, ..., X_n$ is drawn, each X_i takes value in $\{1, \ldots, k\}$ and $X_i = j$ indicates the i-th subject belongs to the j-th category. In this contest, assuming exchangeability, one has that $X_i \mid p \overset{iid}{\sim} f$, where $p = (p_1, p_2, \ldots, p_k)$ with $p_j \geq 0$ and $\sum_{j=1}^{k} p_j = 1$ and where f is a probability mass function with support on $\{1, 2, \ldots, k\}$ and such that $f(j) = p_j$, for $j = 1, \ldots, k$. The model can be equivalently rewritten in terms of the random counts $N_j = \sum_{i=1}^{n} \mathbb{1}_{\{j\}}(X_i)$, for $j = 1, ..., k$, which are distributed according to a multinomial distribution, i.e.

$$\mathbb{P}[N_1 = n_1, ..., N_k = n_k \mid p] = \begin{cases} \frac{\Gamma\left(\sum_{j=1}^{k} n_j + 1\right)}{\prod_{j=1}^{k} \Gamma(n_j + 1)} \prod_{j=1}^{k} p_j^{n_j} & \text{for } (n_1, \ldots, n_j) \;:\; \sum_{j=1}^{k} n_j = n \\ 0 & \text{otherwise} \end{cases}$$

and we write $N_1, ..., N_k \mid p \sim \text{Multinomial}(n, p)$. The conjugate prior distribution to the multinomial distribution is the Dirichlet distribution

$$p \sim \Delta_{k-1}(\alpha_1, \ldots, \alpha_k)$$

The marginal likelihood of the model is called Dirichlet-Multinomial distribution and is given by

$$\mathbb{P}[N_1 = n_1, ..., N_k = n_k] = \begin{cases} \frac{\Gamma\left(\sum_{j=1}^{k} n_j + 1\right)\Gamma(\alpha_0)}{\Gamma\left(\sum_{j=1}^{k} n_j + \alpha_0\right)} \prod_{j=1}^{k} \frac{\Gamma(n_j + \alpha_j)}{\Gamma(n_j + 1)\Gamma(\alpha_j)} & \text{if } \sum_{j=1}^{k} n_j = n \\ 0 & \text{otherwise} \end{cases}$$

Following the Bayesian paradigm, one gets that the posterior distribution for p is given by

$$p \mid N_1 = n_1, ..., N_k = n_k \sim \Delta_{k-1}(\alpha_1 + n_1, \alpha_2 + n_2, ..., \alpha_k + n_k)$$

where n_j, for $j = 1, ..., k$, is the value assumed by N_j in the observed sample, i.e. the observed absolute frequency for the j-th group. So clearly the point estimate p under a quadratic loss function is

$$\mathbb{E}[p \mid N_1 = n_1, ..., N_k = n_k] = \left(\frac{\alpha_1 + n_1}{\alpha_0 + n}, ..., \frac{\alpha_k + n_k}{\alpha_0 + n} \right)$$

which implies that the posterior point estimate is a weighted average of the prior expected value and the frequentist estimate

$$\mathbb{E}[p_j \mid N_1 = n_1, ..., N_k = n_k] = \frac{\alpha_j}{\alpha_0} \frac{\alpha_0}{\alpha_0 + n} + \frac{n_j}{n} \frac{n}{\alpha_0 + n}$$

As usually, letting n going to infinity, the Bayesian point estimate converges to the frequentist estimate of p_j: $\hat{p_j} = \frac{n_j}{n}$. Moreover, the higher α_0, the smaller the pace of converge, which is a result coherent to the fact that α_0 can be seen as reflecting our "confidence" in the prior guess.

A.3 Finite Mixture Model with Dirichlet prior

A finite mixture model is defined as follows

$$X_i \mid \pi, \theta_1^*, ..., \theta_k^* \overset{iid}{\sim} \sum_{j=1}^{k} \pi_j \kappa(\cdot; \theta_j^*)$$

where $\pi = (\pi_1, \ldots, \pi_k)$ is a random parameter taking value in the $(k-1)$-dimensional probability simplex and $\kappa(\cdot; \cdot)$ is a transition kernel. The most classical prior for such model is provided by

$$\begin{aligned} (\pi_1, \ldots, \pi_k) &\sim D_{k-1}(\alpha_1, \ldots, \alpha_k) \\ \theta_j^* &\overset{iid}{\sim} P_0 \end{aligned}$$

An analogous way to describe the model requires the use of n latent variables c_1, c_2, ..., c_n, each taking value in $\{1, 2, \ldots, k\}$, called *cluster labels*. They lead to the following representation.

$$\begin{aligned} X_i \mid c_1, c_2, ..., c_n, \theta_1^*, ..., \theta_k^* &\overset{ind}{\sim} \kappa(\cdot; \theta_{c_i}^*) \\ c_i \mid \pi &\overset{iid}{\sim} \textit{Discrete}(\pi) \\ \pi &\sim D_{k-1}(\alpha_1, ..., \alpha_k) \\ \theta_j^* &\overset{iid}{\sim} P_0 \end{aligned}$$

where *Discrete*(π) is used to denote a probability mass function f with support on $\{1, 2, \ldots, k\}$ and such that $f(j) = \pi_j$, for $j = 1, \ldots, k$.
Finite mixture models are mainly used to perform clustering (i.e. unsupervised learning). However, their main limitation resides in the fact that they require to define the number of clusters k a priori even when it is unknown. This drawback can be overcome mainly in three alternative ways. One is to develop a modeling selection procedure to choose the appropriate value for k. See, for instance, Ishwaran et al. (2001). Another strategy is setting a prior over the parameter k. When this is done, the resulting model is called mixture of finite mixture. See, for instance, Richardson & Green (1997), Miller & Harrison (2018), Argiento & De Iorio (2019) and Frühwirth-Schnatter et al. (2020). Finally, one can assume an infinite number of latent components, as happens in Dirichlet process mixture models, which are reviewed in Chapter 1 of this thesis.

Appendix B

Moments of Functional of CRMs

Proposition 6.2. *Consider the object*

$$\mu(f) = \int_{\mathbb{Y}} f(y)\tilde{\mu}(dy)$$

where $f : \mathbb{Y} \to \mathbb{R}^+$ and $\tilde{\mu}$ is a completely random measure on $(\mathbb{Y}, \mathcal{Y})$ without fixed jumps and with Lévy intensity given by $v(ds, dy) = \rho(s)ds\, c\, P_0(dy)$ with P_0 a non-atomic probability measure on $(\mathbb{Y}, \mathcal{Y})$. Then

$$\mathbb{E}[\mu(f)] = \int_{\mathbb{Y}} f(y)\, s\, v(ds, dy)$$

Proof.

$$\begin{aligned}
\mathbb{E}[\lambda(t)] &= -\frac{\partial}{\partial u}\left(\mathbb{E}\left[e^{-u\int_{\mathbb{Y}} f(y)\tilde{\mu}(dy)}\right]\right)\Bigg|_{u=0} = \\
&= -\frac{\partial}{\partial u}\left(\exp\left\{\int_{\mathbb{R}^+\times\mathbb{Y}} (1 - e^{-u\, s\, f(y)})\, v(ds, dy)\right\}\right)\Bigg|_{u=0} = \\
&= \int_{\mathbb{R}^+\times\mathbb{Y}} f(y)\, s\, v(ds, dy)
\end{aligned}$$

□

Proposition 6.3. *Consider the two following objects*

$$\mu(f_1) = \int_{\mathbb{Y}} f_1(y)\tilde{\mu}(dy) \qquad\qquad \mu(f_2) = \int_{\mathbb{Y}} f_2(y)\tilde{\mu}(dy)$$

where $f_l : \mathbb{Y} \to \mathbb{R}^+$, for $l = 1, 2$, and $\tilde{\mu}$ is a completely random measure on $(\mathbb{Y}, \mathcal{Y})$ without fixed jumps and with Lévy intensity given by $v(ds, dy) = \rho(s)ds\, c\, P_0(dy)$ with P_0 a non-atomic

probability measure on $(\mathbb{Y}, \mathcal{Y})$*. Then*

$$\mathbb{E}[\mu(f_1)\,\mu(f_2)] = \int\limits_{\mathbb{R}^+\times\mathbb{Y}} f_1(y)\,f_2(y)\,s^2\,v(ds,dy) + \\ + \int\limits_{\mathbb{R}^+\times\mathbb{Y}} f_1(y)\,s\,v(ds,dy) \int\limits_{\mathbb{R}^+\times\mathbb{Y}} f_2(y)\,s\,v(ds,dy)$$

Proof.

$$\mathbb{E}[\lambda_1(t_1)\,\lambda_2(t_2)] = \frac{\partial''}{\partial u_1 \partial u_2}\left(\mathbb{E}\Big[e^{-\int_{\mathbb{Y}}(u_1\,f_1(y)+u_2\,f_2(y))\tilde{\mu}(dy)}\Big]\right)\Bigg|_{\substack{u_1=0\\u_2=0}} = \\ = \frac{\partial''}{\partial u_1 \partial u_2}\left(\exp\left\{\int\limits_{\mathbb{R}^+\times\mathbb{Y}}(1-e^{-u_1\,s\,f_1(y)-u_2\,s\,f_2(y)})\,v(ds,dy)\right\}\right)\Bigg|_{\substack{u_1=0\\u_2=0}} = \\ = \int\limits_{\mathbb{R}^+\times\mathbb{Y}} f_1(y)\,f_2(y)\,s^2\,v(ds,dy) \; + \left(\int\limits_{\mathbb{R}^+\times\mathbb{Y}} f_1(y)\,s\,v(ds,dy)\times\right. \\ \left.\times \int\limits_{\mathbb{R}^+\times\mathbb{Y}} f_2(y)\,s\,v(ds,dy)\right)$$

□

Appendix C

Faà di Bruno's Formula

Faà di Bruno's Formula

The formula of Faa di Bruno (Faà di Bruno, 1857) provides an explicit expression for the n-th derivative of a composition of functions as follows. Let $g(x)$ be defined on a neighborhood of x_0 and have derivatives up to order n at x_0; let $f(y)$ be defined on a neighborhood of $y_0 = g(x_0)$ and have derivatives up to order n at y_0. Then the n-th derivative of the composition $h(x) = f[g(x)]$ at x_0 is given by the formula

$$\frac{\partial^n}{\partial x^n} h(x)\bigg|_{x=x_0} = \sum_{k=1}^{n} f_k \sum_{p(n,k)} n! \prod_{i=1}^{n} \frac{g_i^{\lambda_i}}{\lambda_i! i!^{\lambda_i}}$$

where

$$f_k = \frac{\partial^k}{\partial y^k} f(y)\bigg|_{y=g(x_0)} \qquad g_i = \frac{\partial^i}{\partial x^i} g(x)\bigg|_{x=x_0}$$

$$p(n,k) = \{(\lambda_1, \ldots, \lambda_n) \,:\, \lambda_i \in \mathbb{N}_0, \sum_{i=1}^{n} \lambda_i = k, \sum_{i=1}^{n} i\lambda_i = n\}$$

$p(n,k)$ encodes the set of partitions of n identical elements into k groups such that λ_i is the number of groups with i elements.

From Equation (5.9) to Equation (5.12)

$$\psi_z(f_1, f_2) = z \int_{\mathbb{R}^+ \times \mathbb{Y}} (1 - e^{-s\, f_1(y)}) + (1 - e^{-s\, f_2(y)}) \rho(s) ds\, P_0(dy) +$$
$$+ (1 - z) \int_{\mathbb{R}^+ \times \mathbb{Y}} (1 - e^{-s\,(f_1(y) + f_2(y))}) \rho(s) ds\, P_0(dy)$$

$$\psi_z(\gamma \mathbb{1}_{y^*}, K_N(y)\mathbb{1}_{y^*}) = P_0(dy^*)\Bigg\{ z \int\limits_{\mathbb{R}^+} \left[(1-e^{-s\,\gamma}) + (1-e^{-s\,K_N(y^*)})\right]\rho(s)ds + \\ + (1-z)\int\limits_{\mathbb{R}^+} (1-e^{-s\,(\gamma+K_N(y^*))})\rho(s)ds \Bigg\}$$

$$\frac{\partial}{\partial\gamma}\psi_z(\gamma \mathbb{1}_{y^*}, K_N(y)\mathbb{1}_{y^*}) = P_0(dy^*)\Bigg\{ z \int\limits_{\mathbb{R}^+} s\, e^{-s\,\gamma}\rho(s)ds + \\ + (1-z)\int\limits_{\mathbb{R}^+} s\, e^{-s\,(\gamma+K_N(y^*))}\rho(s)ds \Bigg\}$$

$$\frac{\partial^i}{\partial\gamma^i}\psi_z(\gamma \mathbb{1}_{y^*}, K_N(y)\mathbb{1}_{y^*}) = P_0(dy^*)\Bigg\{ z \int\limits_{\mathbb{R}^+} (-1)^{i-1}s^i\, e^{-s\,\gamma}\rho(s)ds + \\ + (1-z)\int\limits_{\mathbb{R}^+} (-1)^{i-1}s^i\, e^{-s\,(\gamma+K_N(y^*))}\rho(s)ds \Bigg\}$$

By Faà di Bruno's formula, setting $f(y) = e^y$ and $g(x) = -c\,\psi_z(x\mathbb{1}_{y^*_{h,T}}, K_T(y)\mathbb{1}_{y^*_{h,T}})$, we have

$$(-1)^{n_{h,T}}\frac{\partial^{n_{h,T}}}{\partial\gamma^{n_{h,T}}} e^{-c\psi_z(\gamma \mathbb{1}_{dy^*_{h,T}}, K_N(y)\mathbb{1}_{dy^*_{h,T}})}\Bigg|_{\gamma=K_T(y^*_{h,T})} =$$

$$= (-1)^{n_{h,T}}\sum_{k=1}^{n_{h,T}} e^{-c\psi_z(K_T(y^*_{h,T}),K_N(y^*_{h,T}))}c^k\, P_0(dy^*_{h,T})^k \times \\ \times \sum_{p(n_{h,T},k)} n_{h,T}! \prod_{i=1}^{n_{h,T}} \frac{(-1)^{\lambda_i}}{\lambda_i! i!^{\lambda_i}} \Bigg\{ z\int\limits_{\mathbb{R}^+} (-1)^{i-1}s^i\, e^{-s\,K_T(y^*_{h,T})}\rho(s)ds + \\ + (1-z)\int\limits_{\mathbb{R}^+} (-1)^{i-1}s^i\, e^{-s\,(K_T(y^*_{h,T})+K_N(y^*_{h,T}))}\rho(s)ds \Bigg\}^{\lambda_i}$$

and, thus,

$$(-1)^{n_{h,T}}\frac{\partial^{n_{h,T}}}{\partial\gamma^{n_{h,T}}}e^{-c\psi_z(\gamma\mathbb{1}_{dy^*_{h,T}},K_N(y)\mathbb{1}_{dy^*_{h,T}})}\Big|_{\gamma=K_T(y^*_{h,T})}=$$

$$\begin{aligned}&= e^{-c\psi_z(K_T(y^*_{h,T}),K_N(y^*_{h,T}))}\times\\&\times c\,P_0(dy^*_{h,T})\Bigg\{z\int_{\mathbb{R}^+}s^{n_{h,T}}\,e^{-s\,K_T(y^*_{h,T})}\rho(s)ds+\\&+(1-z)\int_{\mathbb{R}^+}s^{n_{h,T}}\,e^{-s\,(K_T(y^*_{h,T})+K_N(y^*_{h,T}))}\rho(s)ds\Bigg\}+o(P_0(dy^*_{h,T}))\end{aligned}$$

From Equation (5.16) to Equation (5.17)

$$\psi^{(l)}(\gamma)=\int_{\mathbb{R}^+}[1-e^{-s\gamma}]\rho_l(s)ds$$

$$\frac{\partial^i}{\partial\gamma^i}\psi^{(l)}(\gamma)=\int_{\mathbb{R}^+}(-1)^{i-1}s^i\,e^{-s\,\gamma}\rho_l(s)ds$$

By Faà di Bruno's formula, setting $f(y)=e^y$ and $g(x)=-\psi^{(T)}(\gamma)\tilde{\mu}_0(dy^*_h)$, we have

$$\begin{aligned}(-1)^{n_h}\frac{\partial^{n_{h,T}}}{\partial\gamma^{n_{h,T}}}e^{-\psi^{(T)}(\gamma)\tilde{\mu}_0(dy^*_h)}\Big|_{\gamma=K_T(y^*_h)}&=\cancel{(-1)^{n_h}}\sum_{r=1}^{n_{h,T}}e^{-\psi^{(T)}(K_T(y^*_h))\tilde{\mu}_0(dy^*_h)}\times\\&\times\sum_{p(n_{h,T},r)}n_{h,T}!\prod_{i=1}^{n_{h,T}}\frac{\cancel{(-1)^{\lambda_i}}}{\lambda_i!i!^{\lambda_i}}\Bigg\{\tilde{\mu}_0(dy^*_h)\int_{\mathbb{R}^+}\cancel{(-1)^{i-1}}s^i\,e^{-s\,K_T(y^*_h)}\rho_T(s)ds\Bigg\}^{\lambda_i}=\end{aligned}$$

$$\begin{aligned}&= e^{-\psi^{(T)}(K_T(y^*_h))\tilde{\mu}_0(dy^*_h)}\sum_{r=1}^{n_{h,T}}\tilde{\mu}_0(dy^*_h)^r\times\\&\times\sum_{p(n_{h,T},r)}n_{h,T}!\prod_{i=1}^{n_{h,T}}\frac{1}{\lambda_i!i!^{\lambda_i}}\Bigg\{\int_{\mathbb{R}^+}s^i\,e^{-s\,K_T(y^*_h)}\rho_T(s)ds\Bigg\}^{\lambda_i}\end{aligned}$$

Defining $\xi_{n_{h,T},T,r}(K_T(y^*_h))=\sum_{p(n_{h,T},r)}n_{h,T}!\prod_{i=1}^{n_{h,T}}\frac{1}{\lambda_i!i!^{\lambda_i}}\Bigg\{\int_{\mathbb{R}^+}s^i\,e^{-s\,K_T(y^*_h)}\rho_T(s)ds\Bigg\}^{\lambda_i}$, we get

$$(-1)^{n_h}\frac{\partial^{n_h}}{\partial\gamma^{n_h}}e^{-\psi^{(T)}(\gamma)\tilde{\mu}_0(dy_h^*)}\Big|_{\gamma=K_T(y_h^*)} =$$
$$=\sum_{r=1}^{n_{h,T}}\xi_{n_{h,T},T,r}(K_T(y_h^*))\,e^{-\psi^{(T)}(K_T(y_h^*))\tilde{\mu}_0(dy_h^*)}\tilde{\mu}_0(dy_h^*)^r$$

Notice that $\xi_{n_{h,T},T,r}(K_T(y_h^*))=\sum\limits_{(*)}\binom{n_{h,T}}{q_1,\cdots,q_r}\frac{1}{r!}\tau_{q_1}^{(T)}(K_T(y_h^*))\cdots\tau_{q_r}^{(T)}(K_T(y_h^*))$, where the sum runs over all vectors $(q_1,\ldots,q_r)$ of positive integers such that $\sum_{j=1}^r q_j=n_{h,T}$.
Analogously, we have that

$$(-1)^{n_{h,N}}\frac{\partial^{n_{h,N}}}{\partial\gamma^{n_{h,N}}}e^{-\psi^{(N)}(\gamma)\tilde{\mu}_0(dy_h^*)}\Big|_{\gamma=K_N(y_h^*)} =$$
$$=\sum_{r=1}^{n_{h,N}}\xi_{n_{h,N},N,r}(K_N(y_h^*))\,e^{-\psi^{(N)}(K_N(y_h^*))\tilde{\mu}_0(dy_h^*)}\tilde{\mu}_0(dy_h^*)^r$$

where $\xi_{n_{h,N},N,r}(K_N(y_h^*))=\sum\limits_{p(n_{h,N},r)}n_{h,N}!\prod\limits_{i=1}^{n_{h,N}}\frac{1}{\lambda_i!i!^{\lambda_i}}\left\{\int\limits_{\mathbb{R}^+}s^i\,e^{-s\,K_N(y_h^*)}\rho_N(s)ds\right\}^{\lambda_i}$ or, equivalently, $\xi_{n_{h,N},N,r}(K_N(y_h^*))=\sum\limits_{(*)}\binom{n_{h,N}}{q_1,\cdots,q_r}\frac{1}{r!}\tau_{q_1}^{(N)}(K_N(y_h^*))\cdots\tau_{q_r}^{(N)}(K_N(y_h^*))$, where the sum runs over all vectors $(q_1,\ldots,q_r)$ of positive integers such that $\sum_{j=1}^r q_j=n_{h,N}$.

From Equation (5.18) to Equation (5.19)

$$e^{-\psi^{(T)}(K_T(y_h^*))-\psi^{(N)}(K_N(y_h^*))\tilde{\mu}_0(dy_h^*)}\tilde{\mu}_0(dy_h^*)^{r_{T,h}+r_{N,h}} =$$
$$=(-1)^{r_{T,h}+r_{N,h}}\frac{\partial^{r_{T,h}+r_{N,h}}}{\partial\gamma^{r_{T,h}+r_{N,h}}}e^{-\gamma\tilde{\mu}_0(dy_h^*)}\Big|_{\gamma=\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))}$$

$$\mathbb{E}\left[e^{-\psi^{(T)}(K_T(y_h^*))-\psi^{(N)}(K_N(y_h^*))\tilde{\mu}_0(dy_h^*)}\tilde{\mu}_0(dy_h^*)^{r_{T,h}+r_{N,h}}\right] =$$
$$=(-1)^{r_{T,h}+r_{N,h}}\frac{\partial^{r_{T,h}+r_{N,h}}}{\partial\gamma^{r_{T,h}+r_{N,h}}}e^{-c_0\psi^{(0)}(\gamma)P_0(dy_h^*)}\Big|_{\gamma=\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))}$$

By Faà di Bruno's formula, setting $f(y)=e^y$ and $g(x)=-c_0\psi^{(0)}(\gamma)P_0(dy_h^*)$, we have

$$(-1)^{r_{T,h}+r_{N,h}}\frac{\partial^{r_{T,h}+r_{N,h}}}{\partial\gamma^{r_{T,h}+r_{N,h}}}e^{c_0-\psi^{(0)}(\gamma)P_0(dy_h^*)}\Big|_{\gamma=\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))}=$$

$$=\cancel{(-1)^{r_{T,h}+r_{N,h}}}e^{-c_0\psi^{(0)}(\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*)))P_0(dy_h^*)}\times$$

$$\times\sum_{r'=1}^{r_{T,h}+r_{N,h}}c_0^{r'}P_0(dy_h^*)^{r'}\sum_{p(r_{T,h}+r_{N,h},r')}(r_{T,h}+r_{N,h})!$$

$$\prod_{i=1}^{r_{T,h}+r_{N,h}}\frac{\cancel{(-1)^{\lambda_i}}}{\lambda_i!i!^{\lambda_i}}\left\{\int_{\mathbb{R}^+}\cancel{(-1)^{i-1}}s^i\,e^{-s\,\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))}\rho_0(s)ds\right\}^{\lambda_i}$$

$$(-1)^{r_{T,h}+r_{N,h}}\frac{\partial^{r_{T,h}+r_{N,h}}}{\partial\gamma^{r_{T,h}+r_{N,h}}}e^{c_0-\psi^{(0)}(\gamma)P_0(dy_h^*)}\Big|_{\gamma=\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))}=$$

$$=e^{-\psi^{(0)}(\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*)))P_0(dy_h^*)}\times$$

$$\times c_0P_0(dy_h^*)\int_{\mathbb{R}^+}s^{r_{T,h}+r_{N,h}}\,e^{-s\,\psi^{(T)}(K_T(y_h^*))+\psi^{(N)}(K_N(y_h^*))}\rho_0(s)ds+o(P_0(dy_h^*))$$

This Page Intentionally Left Blank

Bibliography

AKSU, E., CUGLAN, B., TOK, A., CELIK, E., DOGANER, A., SOKMEN, A. & SOKMEN, G. (2021). Cardiac electrical and structural alterations in preeclampsia. *The Journal of Maternal-Fetal & Neonatal Medicine* , 1–10.

ALDOUS, D. J. (1985). Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII—1983*. Springer, pp. 1–198.

AMBROŽIC, J., LUCOVNIK, M., PROKŠELJ, K., TOPLIŠEK, J. & CVIJIC, M. (2020). Dynamic changes in cardiac function before and early postdelivery in women with severe preeclampsia. *Journal of hypertension* **38**, 1367–1374.

ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.

ARBEL, J., LIJOI, A. & NIPOTI, B. (2016). Full Bayesian inference with hazard mixture models. *Computational Statistics & Data Analysis* **93**, 359–372.

ARBEL, J. & PRÜNSTER, I. (2017). A moment-matching Ferguson & Klass algorithm. *Statistics and Computing* **27**, 3–17.

ARGIENTO, R., CREMASCHI, A. & VANNUCCI, M. (2020). Hierarchical normalized completely random measures to cluster grouped data. *Journal of the American Statistical Association* **115**, 318–333.

ARGIENTO, R. & DE IORIO, M. (2019). Is infinity that far? a Bayesian nonparametric perspective of finite mixture models. *arXiv preprint arXiv:1904.09733* .

ASCOLANI, F., FRANZOLINI, B., PRÜNSTER, I. & LIJOI, A. (2021). On the dependence structure in Bayesian nonparametric priors. In *Book of Short Papers SIS2021*.

BALSHAW, R. F. & DEAN, C. (2002). A semiparametric model for the analysis of recurrent-event panel data. *Biometrics* **58**, 324–331.

BARRIOS, E., LIJOI, A., NIETO-BARAJAS, L. E., PRÜNSTER, I. et al. (2013). Modeling with normalized random measure mixture models. *Statistical Science* **28**, 313–334.

BELLAMY, L., CASAS, J.-P., HINGORANI, A. D. & WILLIAMS, D. J. (2007). Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ* **335**, 974.

BERAHA, M., GUGLIELMI, A. & QUINTANA, F. A. (2021). The semi-hierarchical dirichlet process and its application to clustering homogeneous distributions. *Bayesian Analysis* **1**, 1–33.

BERGER, M. & TUTZ, G. (2018). Tree-structured clustering in fixed effects models. *Journal of Computational and Graphical Statistics* **27**, 380–392.

BERNARDO, J. M. & SMITH, A. F. (2009). *Bayesian theory*, vol. 405. John Wiley & Sons.

BHARDWAJ, G. & DUNSBY, A. (2013). The business cycle and the correlation between stocks and commodities. *Journal of Investment Consulting* **14**, 14–25.

BLACKWELL, D. & MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

BRILLINGER, D. R. (2002). John W. Tukey: his life and professional contributions. *The Annals of Statistics* **30**, 1535–1575.

BUSH, C. A. & MACEACHERN, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–285.

CAMERLENGHI, F., DUNSON, D. B., LIJOI, A., PRÜNSTER, I. & RODRIGUEZ, A. (2019a). Latent nested nonparametric priors. *Bayesian Analysis* **14**, 1303–1356.

CAMERLENGHI, F., LIJOI, A., ORBANZ, P., PRÜNSTER, I. et al. (2019b). Distribution theory for hierarchical processes. *The Annals of Statistics* **47**, 67–92.

CAMERLENGHI, F., LIJOI, A. & PRÜNSTER, I. (2018). Bayesian nonparametric inference beyond the Gibbs-type framework. *Scandinavian Journal of Statistics* **45**, 1062–1091.

CAMERLENGHI, F., LIJOI, A. & PRÜNSTER, I. (2021). Survival analysis via hierarchically dependent mixture hazards. *The Annals of Statistics* **49**, 863–884.

CATALANO, M., LIJOI, A. & PRÜNSTER, I. (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *The Annals of Statistics* **49**, 2916–2947.

CHRISTENSEN, J. & MA, L. (2020). A Bayesian hierarchical model for related densities by using Pólya trees. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 127–153.

CIFARELLI, D. M. & REGAZZINI, E. (1978). Problemi statistici non parametrici in condizioni di scambialbilita parziale e impiego di medie associative. *Quaderni Istituto Matematica Finanziaria, Torino* .

CIPOLLI, W., HANSON, T. & MCLAIN, A. C. (2016). Bayesian nonparametric multiple testing. *Computational Statistics & Data Analysis* **101**, 64–79.

CONT, R. & TANKOV, P. (2004). *Financial modelling with jump processes*. CRC press.

DAHL, D. B. & NEWTON, M. A. (2007). Multiple hypothesis testing by clustering treatment effects. *Journal of the American Statistical Association* **102**, 517–526.

DALAL, S. (1979a). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Processes and their Applications* **9**, 99–107.

DALAL, S. R. (1979b). Nonparametric and robust Bayes estimation of location. *Optimizing Methods in Statistics* **9**, 141–166.

DALEY, D. J. & VERE-JONES, D. (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer.

DAVIS, E. F., LAZDAM, M., LEWANDOWSKI, A. J., WORTON, S. A., KELLY, B., KENWORTHY, Y., ADWANI, S., WILKINSON, A. R., MCCORMICK, K. & SARGENT, I. (2012). Cardiovascular risk factors in children and young adults born to preeclamptic pregnancies: a systematic review. *Pediatrics* **129**, 1552–1561.

DE BLASI, P., FAVARO, S., LIJOI, A., MENA, R. H., PRÜNSTER, I. & RUGGIERO, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE transactions on pattern analysis and machine intelligence* **37**, 212–229.

DE BLASI, P., PECCATI, G., PRÜNSTER, I. et al. (2009). Asymptotics for posterior hazards. *The Annals of Statistics* **37**, 1906–1945.

DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, vol. 7.

DE FINETTI, B. (1938). Sur la condition d'equivalence partielle. *Actualités Scientifiques et Industrielles* **739**, 5–18, Translated In: Studies in Inductive and Probability, II. Jeffrey, R. (ed.) University of California Press: Berkeley 1980.

DE IORIO, M., MÜLLER, P., ROSNER, G. L. & MACEACHERN, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

DEMARTELLY, V. A., DREIXLER, J., TUNG, A., MUELLER, A., HEIMBERGER, S., FAZAL, A. A., NASEEM, H., LANG, R., KRUSE, E., YAMAT, M. et al. (2021). Long-term postpartum cardiac function and its association with preeclampsia. *Journal of the American Heart Association* **10**, e018526.

DENTI, F., GUINDANI, M., LEISEN, F., LIJOI, A., WADSWORTH, W. D. & VANNUCCI, M. (2020). Two-group Poisson-Dirichlet mixtures for multiple testing. *Biometrics* **76**, 10.1111/biom.13314.

DIACONIS, P. & FREEDMAN, D. (1986). On inconsistent Bayes estimates of location. *The Annals of Statistics* **14**, 68–87.

DO, K.-A., MÜLLER, P. & TANG, F. (2005). A Bayesian mixture model for differential gene expression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**, 627–644.

DOLEA, C. & ABOUZAHR, C. (2003). Global burden of hypertensive disorders of pregnancy in the year 2000. Tech. rep., GBD 2000 Working Paper, World Health Organization, Geneva.

DOSS, H. (1984). Bayesian estimation in the symmetric location problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **68**, 127–147.

DUNSON, D. & PARK, J. (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307–323.

DUNSON, D. B. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian nonparametrics (Hjort, N.L., Holmes, C.C., Müller, P., Walker, S.G. Eds.)*. Cambridge University Press, Cambridge, pp. 223–273.

DYKSTRA, R. & LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics* **9**, 356–367.

EFRON, B. & MORRIS, C. (1977). Stein's paradox in statistics. *Scientific American* **236**, 119–127.

EPIFANI, I. & LIJOI, A. (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica* **20**, 1455–1484.

ESCOBAR, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277.

ESCOBAR, M. D. & WEST, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

FAÀ DI BRUNO, F. (1857). Note sur une nouvelle formule de calcul différentiel. *Quarterly Journal of Pure and Applied Mathematics* **1**, 99.

FAVARO, S., LIJOI, A., NAVA, C., NIPOTI, B., PRUENSTER, I., TEH, Y. W. et al. (2016). On the stick-breaking representation for homogeneous NRMIs. *Bayesian Analysis* **11**, 697–724.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

FERGUSON, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*. Elsevier, pp. 287–302.

FERGUSON, T. S. & KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics* **43**, 1634–1643.

FERGUSON, T. S., PHADIA, E. G. & TIWARI, R. C. (1992). Bayesian nonparametric inference. *Lecture Notes-Monograph Series* **17**, 127–150.

FOTI, N. J. & WILLIAMSON, S. A. (2015). A survey of non-exchangeable priors for Bayesian nonparametric models. *IEEE transactions on pattern analysis and machine intelligence* **37**, 359–371.

FRÜHWIRTH-SCHNATTER, S., MALSINER-WALLI, G. & GRÜN, B. (2020). Generalized mixtures of finite mixtures and telescoping sampling. *arxiv:2005.09918* .

GARCIA-GONZALEZ, C., GEORGIOPOULOS, G., AZIM, S. A., MACAYA, F., KAMETAS, N., NIHOYANNOPOULOS, P., NICOLAIDES, K. H. & CHARAKIDA, M. (2020). Maternal cardiac assessment at 35 to 37 weeks improves prediction of development of preeclampsia. *Hypertension* **76**, 514–522.

GELFAND, A. E., DEY, D. K. & CHANG, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. Tech. rep., STANFORD UNIV CA DEPT OF STATISTICS.

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. & RUBIN, D. B. (2013). *Bayesian data analysis*. CRC press.

GHOSAL, S., GHOSH, J. K. & RAMAMOORTHI, R. (1999). Consistent semiparametric Bayesian inference about a location parameter. *Journal of Statistical Planning and Inference* **77**, 181–193.

GHOSAL, S. & VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference*, vol. 44. Cambridge University Press.

GOPALAN, R. & BERRY, D. A. (1998). Bayesian multiple comparisons using Dirichlet process priors. *Journal of the American Statistical Association* **93**, 1130–1139.

GRIFFIN, J. E., KOLOSSIATIS, M. & STEEL, M. F. (2013). Comparing distributions by using dependent normalized random-measure mixtures. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology* **75**, 499–529.

GRIFFIN, J. E. & LEISEN, F. (2017). Compound random measures and their use in Bayesian non-parametrics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2**, 525–545.

GRIFFITHS, R. C. & MILNE, R. K. (1978). A class of bivariate poisson processes. *Journal of Multivariate Analysis* **8**, 380–395.

GUGLIELMI, A., LEVA, F., PAGANONI, A. M., RUGGERI, F. & SORIANO, J. (2014). Semiparametric Bayesian models for clustering and classification in the presence of unbalanced in-hospital survival. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* , 25–46.

GUINDANI, M., MÜLLER, P. & ZHANG, S. (2009). A Bayesian discovery procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 905–925.

GUTIÉRREZ, L., BARRIENTOS, A. F., GONZÁLEZ, J., TAYLOR-RODR´IGUEZ, D. et al. (2019). A Bayesian nonparametric multiple testing procedure for comparing several treatments against a control. *Bayesian Analysis* **14**, 649–675.

HALL, M. E., GEORGE, E. M. & GRANGER, J. P. (2011). The heart during pregnancy. *Revista Española de Cardiolog´ia (English Edition)* **64**, 1045–1050.

HANNUM, R. & HOLLANDER, M. (1983). Robustness of Ferguson's Bayes estimator of a distribution function. *The Annals of Statistics* , 632–639.

HE, X., TONG, X. & SUN, J. (2009). Semiparametric analysis of panel count data with correlated observation and follow-up times. *Lifetime Data Analysis* **15**, 177–196.

IGBERASE, G. & EBEIGBE, P. (2006). Eclampsia: ten-years of experience in a rural tertiary hospital in the Niger delta, Nigeria. *Journal of Obstetrics and Gynaecology* **26**, 414–417.

IGLESIAS, P. L., ORELLANA, Y. & QUINTANA, F. A. (2009). Nonparametric Bayesian modelling using skewed Dirichlet processes. *Journal of Statistical Planning and Inference* **139**, 1203–1214.

ISHWARAN, H. & JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.

ISHWARAN, H., JAMES, L. F. & SUN, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association* **96**, 1316–1332.

JAMES, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *The Annals of Statistics* **33**, 1771–1799.

JAMES, L. F., LIJOI, A. & PRÜNSTER, I. (2006). Conjugacy as a distinctive feature of the dirichlet process. *Scandinavian Journal of Statistics* **33**, 105–120.

JAMES, L. F., LIJOI, A. & PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics* **36**, 76–97.

JAMES, L. F., LIJOI, A. & PRÜNSTER, I. (2010). On the posterior distribution of classes of random means. *Bernoulli* **16**, 155–180.

KALLSEN, J. & TANKOV, P. (2006). Characterization of dependence of multidimensional lévy processes using lévy copulas. *Journal of Multivariate Analysis* **97**, 1551–1572.

KINGMAN, J. (1967). Completely random measures. *Pacific Journal of Mathematics* **21**, 59–78.

KINGMAN, J. (1993). Poisson Processes.

KLEINMAN, K. P. & IBRAHIM, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics* , 921–938.

LEE, J., QUINTANA, F. A., MÜLLER, P. & TRIPPA, L. (2013). Defining predictive probability functions for species sampling models. *Statistical Science* **28**, 209–222.

LIANG, Y., LI, Y. & ZHANG, B. (2018). Bayesian nonparametric inference for panel count data with an informative observation process. *Biometrical Journal* **60**, 583–596.

LIJOI, A., MENA, R. H. & PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association* **100**, 1278–1291.

LIJOI, A. & NIPOTI, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *Journal of the American Statistical Association* **109**, 802–814.

LIJOI, A., NIPOTI, B. & PRÜNSTER, I. (2014a). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260–1291.

LIJOI, A., NIPOTI, B. & PRÜNSTER, I. (2014b). Dependent mixture models: clustering and borrowing information. *Computational Statistics & Data analysis* **71**, 417–433.

LIJOI, A. & PRÜNSTER, I. (2010). Models beyond the dirichlet process. In *Bayesian nonparametrics (Hjort, N.L., Holmes, C.C., Müller, P., Walker, S.G. Eds.)*. pp. 80–136.

LIJOI, A., PRÜNSTER, I. & REBAUDO, G. (2020). Flexible clustering via hidden hierarchical Dirichlet priors. *Carlo Alberto notebooks* **634**.

LINDLEY, D. V. (1972). *Bayesian statistics: A review*. SIAM.

LO, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics* **12**, 351–357.

LO, A. Y. & WENG, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Annals of the Institute of Statistical Mathematics* **41**, 227–245.

MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA proceedings of the section on Bayesian statistical science*, vol. 1. Alexandria, Virginia. Virginia: American Statistical Association; 1999.

MACEACHERN, S. N. (2000). Dependent Dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University.

MAJUMDAR, S. (1992). On topological support of Dirichlet prior. *Statistics & Probability Letters* **15**, 385–388.

MALIK, A., JEE, B. & GUPTA, S. K. (2019). Preeclampsia: disease biology and burden, its management strategies with reference to India. *Pregnancy Hypertension* **15**, 23–31.

MARTIN, R. & TOKDAR, S. T. (2012). A nonparametric empirical Bayes framework for large-scale multiple testing. *Biostatistics* **13**, 427–439.

MCCLURE, E. M., SALEEM, S., PASHA, O. & GOLDENBERG, R. L. (2009). Stillbirth in developing countries: a review of causes, risk factors and prevention strategies. *The Journal of Maternal-fetal & Neonatal Medicine* **22**, 183–190.

MILLER, J. W. & HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *Journal of the American Statistical Association* **113**, 340–356.

MORENO, A., WU, Z., YAP, J., LAM, C., WETTER, D. W., NAHUM-SHANI, I., DEMPSEY, W. & REHG, J. M. (2020). A robust functional EM algorithm for incomplete panel count data. In *Advances in Neural Information Processing Systems - NIPS*.

MOSER, S., RODRIGUEZ, A. & LOFLAND, C. L. (2021). Multiple ideal points: Revealed preferences in different domains. *Political Analysis* **29**, 139–166.

MULIERE, P. & PETRONE, S. (1978). A Bayesian predictive approach to sequential search for an optimal dose: parametric and nonparametric models. *Journal of the Italian Statistical Society* **2**, 349–364.

MÜLLER, P., QUINTANA, F. & ROSNER, G. (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 735–749.

MÜLLER, P., QUINTANA, F. A., JARA, A. & HANSON, T. (2015). *Bayesian nonparametric data analysis*. Springer.

MÜLLER, P., QUINTANA, F. A. & PAGE, G. (2018). Nonparametric Bayesian inference in applications. *Statistical Methods & Applications* **27**, 175–206.

NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics* **9**, 249–265.

NIETO-BARAJAS, L. E., PRÜNSTER, I. & WALKER, S. G. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics* **32**, 2343–2360.

PECCATI, G., PRÜNSTER, I. et al. (2008). Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Annals of Applied Probability* **18**, 1910–1943.

PEDERSEN, S. S., VON KÄNEL, R., TULLY, P. J. & DENOLLET, J. (2017). Psychosocial perspectives in cardiovascular disease. *European Journal of Preventive Cardiology* **24**, 108–115.

PETRALIA, F., RAO, V. & DUNSON, D. B. (2012). Repulsive mixtures. In *Advances in Neural Information Processing Systems - NIPS*.

PITMAN, J. (1996). Some developments of the Blackwell-Macqueen urn scheme. *Lect. Notes-Monograph Ser.* **30**, 245–267.

QUINLAN, J. J., QUINTANA, F. A. & PAGE, G. L. (2017). Parsimonious hierarchical modeling using repulsive distributions. *arXiv preprint arXiv:1701.04457* .

QUINTANA, F. A., MÜLLER, P., JARA, A. & MACEACHERN, S. N. (2020). The dependent Dirichlet process and related models. *arXiv preprint arXiv:2007.06129* .

REGAZZINI, E., LIJOI, A., PRÜNSTER, I. et al. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics* **31**, 560–585.

RICHARDSON, S. & GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)* **59**, 731–792.

RODRIGUEZ, A. & DUNSON, D. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis* **6**.

RODRIGUEZ, A., DUNSON, D. B. & GELFAND, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association* **103**, 1131–1154.

SCOTT, J. G. & BERGER, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference* **136**, 2144–2162.

SCOTT, J. G. & BERGER, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics* , 2587–2619.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica sinica* **4**, 639–650.

SETHURAMAN, J. & TIWARI, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In *Statistical decision theory and related topics III*. Elsevier, pp. 305–315.

SHAH, A., FAWOLE, B., M'IMUNYA, J. M., AMOKRANE, F., NAFIOU, I., WOLOMBY, J.-J., MUGERWA, K., NEVES, I., NGUTI, R. & KUBLICKAS, M. (2009). Cesarean delivery outcomes from the WHO global survey on maternal and perinatal health in Africa. *International Journal of Gynecology & Obstetrics* **107**, 191–197.

SKLAR, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* **8**, 229–231.

SORIANO, J. & MA, L. (2017). Probabilistic multi-resolution scanning for two-sample differences. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 547–572.

SUN, J. (2013). *Statistical analysis of panel count data*. Springer.

TANKOV, P. (2016). Lévy copulas: review of recent results. In *The fascination of probability, statistics and their applications (Podolskij, M., Stelzer, R., Thorbjørnsen, S., Veraart, A.E.D. Eds.* Springer, pp. 127–151.

TATAPUDI, R. & PASUMARTHY, L. R. (2017a). Data for: Maternal cardiac function in gestational hypertension, mild and severe preeclampsia and normal pregnancy: A comparative study. https://data.mendeley.com/datasets/d72zr4xggx/1. Licensed under a Creative Commons Attribution 4.0 International licence.

TATAPUDI, R. & PASUMARTHY, L. R. (2017b). Maternal cardiac function in gestational hypertension, mild and severe preeclampsia and normal pregnancy: A comparative study. *Pregnancy Hypertension* **10**, 238–241.

TEH, Y. W., JORDAN, M. I., BEAL, M. J. & BLEI, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**, 1566–1581.

THALL, P. F. & LACHIN, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association* **83**, 339–347.

TIMOKHINA, E., KUZMINA, T., STRIZHAKOV, A., PITSKHELAURI, E., IGNATKO, I. & BELOUSOVA, V. (2019). Maternal cardiac function after normal delivery, preeclampsia, and eclampsia: A prospective study. *Journal of Pregnancy* **2019**, 2090–2727.

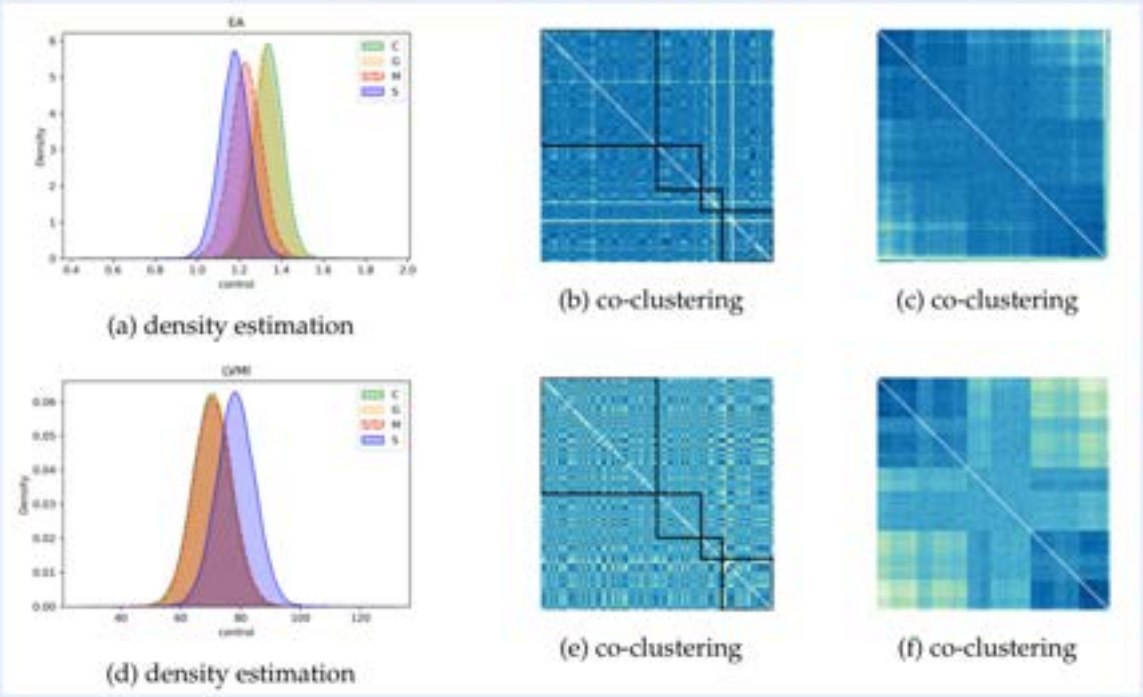
TIWARI, R. C. (1988). Convergence of Dirichlet invariant measures and the limits of Bayes estimates. *Communications in Statistics-Theory and Methods* **17**, 375–393.

XIE, F. & XU, Y. (2020). Bayesian repulsive Gaussian mixture model. *Journal of the American Statistical Association* **115**, 187–203.

ZUANETTI, D. A., MÜLLER, P., ZHU, Y., YANG, S. & JI, Y. (2018). Clustering distributions with the marginalized nested Dirichlet process. *Biometrics* **74**, 584–594.

This Page Intentionally Left Blank

With the term dependent processes, we refer to two or more infinite dimensional random objects, i.e., random probability measures, completely random measures, and random partitions, whose joint probability law does not factorize and, thus, encodes non-trivial dependence. We investigate properties and limits of existing nonparametric dependent priors and propose new dependent processes that fill gaps in the existing literature. To do so, we first define a class of priors, namely multivariate species sampling processes, which encompasses many dependent processes used in Bayesian nonparametrics. Then, in light of our theoretical findings, as well as considering specific motivating applications, we develop novel prior processes outside this class, enlarging the types of data structures and prior information that can be handled by the Bayesian nonparametric approach. We propose three new classes of dependent processes: full-range borrowing of information priors, invariant dependent priors, and dependent priors for panel count data.



(a) density estimation

(b) co-clustering

(c) co-clustering

(d) density estimation

(e) co-clustering

(f) co-clustering