On the dependence structure in Bayesian nonparametric priors

Sulla dipendenza nelle distribuzioni a priori Bayesiane e non parametriche

Filippo Ascolani, Beatrice Franzolini, Antonio Lijoi, and Igor Prünster

Abstract Bayesian models for data grouped into distinct samples are typically defined within the framework of partial exchangeability. All currently known nonparametrics priors for partially exchangeable data induce positive correlation both between observations coming from different samples as well as between the underlying random probability measures. However, such property is not implied by partial exchangeability and may not be appropriate in some applications. Using σ-stable completely random measures and Clayton-Lévy copulas, we propose a nonparametric prior that may induce either negative or positive correlation. The contents of these pages summarize some of the results derived in [1].

Abstract *La parziale scambiabilità è un'assunzione spesso utilizzata nei modelli Bayesiani per dati suddivisi in campioni. Tutte le distribuzioni non parametriche note per dati parzialmente scambiabili inducono correlazione positiva sia tra le osservazioni in diversi campioni, sia tra le misure di probailità sottostanti. Tuttavia, la correlazione positiva non è implicata dalla parziale scambiabilità. In questo lavoro viene introdotta una distribuzione a priori nonparametrica che può indurre correlazione negativa o positiva e che fa uso delle misure completamente aleatorie σ-stabili e delle Clayton-Lévy copulas. Il contenuto di queste pagine riassume alcuni dei risultati derivati in [1].*

Key words: Bayesian nonparametrics, Completely random measure, Lévy copula, Negative correlation, Partial exchangeability

---

Filippo Ascolani
Bocconi University and BIDSA, e-mail: filippo.ascolani@phd.unibocconi.it

Beatrice Franzolini
Bocconi University and BIDSA, e-mail: beatrice.franzolini@phd.unibocconi.it

Antonio Lijoi
Bocconi University and BIDSA, e-mail: antonio.lijoi@unibocconi.it

Igor Prünster
Bocconi University and BIDSA, e-mail: igor.pruenster@unibocconi.it

1 Introduction

Traditional Bayesian models assume that data are exchangeable, which is a homogeneity condition implying the existence of a common underlying distribution from which observations have been sampled. More formally, a sequence of observations $X = (X_i)_{i\geq 1}$ is said exchangeable if and only if $\forall n \geq 1$, $(X_1,\dots,X_n)$ is equal in distribution to $(X_{\sigma(1)},\dots,X_{\sigma(n)})$ for any σ permutation of n elements. It should be clear that exchangeability is an appropriate assumption only when one would like to develop an inferential procedure which disregards any information that may be included in the order in which data were collected and stored.

However, this is not the case, for instance, when data are grouped into many samples corresponding to different experimental conditions or when discrete covariates information is available. In these situations a more plausible assumption is partial exchangeability. Two sequences of data $X_1 = (X_{i,1})_{i\geq 1}$ and $X_2 = (X_{i,2})_{i\geq 1}$, where $X_{j,i}$ is a random variable taking value in a Polish space $(\mathbb{X}, \mathscr{X})$, are said partially exchangeable if and only if for all $n_1 \geq 1$ and $n_2 \geq 1$:

$$(X_{1,1},\dots,X_{n_1,1},X_{1,2},\dots,X_{n_2,2}) \stackrel{d}{=} (X_{\sigma_1(1),1},\dots,X_{\sigma_1(n_1),1},X_{\sigma_2(1),2},\dots,X_{\sigma_2(n_2),2})$$

for any σ_1 and σ_2 permutations of respectively n_1 and n_2 elements. Thanks to de Finetti's representation theorem for partial exchangeability [3], we know that X_1 and X_2 are partial exchangeable if and only if there exist two (possibly dependent) random probability measure $\tilde{p}_1$ and $\tilde{p}_2$ such that:

$$X_{i,j} \mid (\tilde{p}_1, \tilde{p}_2) \stackrel{ind}{\sim} \tilde{p}_j \quad \text{for } j = 1,2 \qquad\qquad (\tilde{p}_1, \tilde{p}_2) \sim Q$$

and Q plays the role of the prior.

In the last two decades there has been a growing interest in developing nonparametric priors for partially exchangeable data. See [5, 11] and references therein.

However, all existing and used nonparametric priors induce a non-negative correlation both between $\tilde{p}_1(A)$ and $\tilde{p}_2(A)$, for every $A \in \mathscr{X}$, and between $X_{i,1}$ and $X_{i',2}$ for any i, i'. Such property is not implied by partial exchangeability and does not fit those applications where one has a priori information regarding negative correlation between obsevarbles in different groups.

In this work, after some preliminaries regarding completely random measures (Section 2), we introduce a novel nonparametric prior (Section 3) over $(\tilde{p}_1, \tilde{p}_2)$ that may induce either negative or positive correlation between the observables. Lastly (Section 4), we develop an algorithm for sampling from the proposed prior and use it to show the conditional behaviour of $\tilde{p}_2$ given $\tilde{p}_1$. The focus of this work is the prior law of $(\tilde{p}_1, \tilde{p}_2)$. For what concerns posterior inference, a comment can be found at the end of Section 3, while further details will be provided in forthcoming works.

2 Preliminaries on completely random measures

Consider a Polish space $(\mathbb{X}, \mathscr{X})$ endowed with its Borel σ-algebra and $(\mathbb{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ the space of boundedly finite measures on $\mathbb{X}$.

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

If $\tilde{\mu}$ is a CRM without deterministic component and fixed points of discontinuity, then $\tilde{\mu}$ is almost surely discrete, i.e.

$$\tilde{\mu} \overset{a.s.}{=} \sum_{j=1}^{\infty} J_j \delta_{X_j}$$

and $\tilde{\mu}$ is characterized by the following Laplace functional transform. For any measurable positive-valued function f,

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

where $\tilde{\nu}$ is called Lévy intensity and uniquely identifies the law of $\tilde{\mu}$. Finally, we assume that the jumps $(J_j)_{j\geq 1}$ and the locations $(X_j)_{j\geq 1}$ are independent, so that $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\mathrm{d}s\,\alpha(\mathrm{d}x)$. For more details on CRM, we refer to [8, 9]. CRMs have been proven a useful tool for prior specification. In particular, they may be normalized to obtain random probability measures, called normalized random measures with independent increments (NRMI), introduced in [12]. The notion of CRM can be extended to a vector of measures as follows:

Definition 2. 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) on $(\mathbb{X}, \mathscr{X})$ if, for every collection of pairwise disjoint sets $(A_i)_{i\geq 1}^n$ in $\mathscr{X}$, 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.

The Laplace functional transform of $\underline{\mu}$ is

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

for measurable $f_1, f_2 : \mathbb{X} \to \mathbb{R}^+$, where ν is called joint Lévy intensity and uniquely identifies the law of $(\tilde{\mu}_1, \tilde{\mu}_2)$.

3 Atom-dependent σ-stable normalized completely random measures

Definition 3. Let $\underline{\xi} = (\xi_1, \xi_2)$ be a CRV on $(\mathbb{X}\times\mathbb{X}, \mathscr{X}\otimes\mathscr{X})$ with Lévy intensity $\nu(\mathrm{d}s_1, \mathrm{d}s_2, \mathrm{d}x_1, \mathrm{d}x_2) = \rho(s_1, s_2)\mathrm{d}s_1\mathrm{d}s_2\,\alpha(\mathrm{d}x_1, \mathrm{d}x_2)$ such that

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

Then $\tilde{\mu}_1(\cdot) = \xi_1(\cdot\times\mathbb{X})$ and $\tilde{\mu}_2(\cdot) = \xi_2(\mathbb{X}\times\cdot)$ are called atom-dependent σ-stable CRMs with underlying joint Levy intensity ν.

Proposition 1. *Consider $\tilde{\mu}_1$ and $\tilde{\mu}_2$ atom-dependent σ-stable CRMs, as defined in Definition 3, then $\tilde{\mu}_j$ is a σ-stable CRM, for $j = 1,2$ and the a.s. discrete representation of $\tilde{\mu}_1$ and $\tilde{\mu}_2$ is:*

$$\tilde{\mu}_1 \overset{a.s.}{=} \sum_{k\geq 1} W_{1,k}\delta_{(\theta_{1,k})} \qquad\qquad \tilde{\mu}_2 \overset{a.s.}{=} \sum_{k\geq 1} W_{2,k}\delta_{(\theta_{2,k})}$$

where the two sequences of weights $(W_{1,k})_{k\geq 1}$ and $(W_{2,k})_{k\geq 1}$ are inherited from the underlying measures ξ_1 and ξ_2 and $(\theta_{1,k}, \theta_{2,k}) \overset{iid}{\sim} G_0 \equiv \alpha/\alpha(\mathbb{X})$.

Definition 4. The random probability measures $\tilde{p}_1$ and $\tilde{p}_2$ obtained normalizing two atom-dependent σ-stable CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$ with underlying joint Levy intensity ν:

$$\tilde{p}_1(\cdot) = \frac{\tilde{\mu}_1(\cdot)}{\tilde{\mu}_1(\mathbb{X})} \qquad\qquad \tilde{p}_2(\cdot) = \frac{\tilde{\mu}_2(\cdot)}{\tilde{\mu}_2(\mathbb{X})}$$

are called atom-dependent σ-stable NRMIs.

In order to obtain a working model which makes use of atom-dependent σ-stable NRMIs, the underlying joint Lévy intensity ν has to be specified. A useful stategy to serve the purpose is to use Lévy copulas. See [2, 7, 10]. A popular Lévy copula is the Clayton's one, which is given by the following expression:

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

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 ξ_1 and ξ_2. As consequence, when Clayton's copula is used to specify the law of two atom-dependent NRMIs, 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 (see [4]):

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

Theorem 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 atom-dependent* σ*-stable NRMIs with underlying joint Lévy intensity* (1)*, then:*

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

where $g:\mathbb{R}^+\to(0,(1-\sigma))$ *and* ρ *is the correlation between two random variables jointly sampled from* G_0.

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

Lastly, concerning the possibility of deriving posterior inference, it is important to note that the representation of $(\tilde{\mu}_1,\tilde{\mu}_2)$ in terms of the CRV (ξ_1,ξ_2) is crucial. Indeed, it allows to obtain posterior representation theorems generalizing the results provided in [6] for the exchangeable case.

4 Prior algorithm and simulations

We conclude this work with a simulation study, which shows 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, unit variances and correlation ρ. To this end we need an algortihm to sample the infinite dimensional parameters $\tilde{p}_1$ and $\tilde{p}_2$ for different values of the hyperparameters θ and ρ. Algorithm 1 serves the purpose and it has been obtained adapting the Algorithm 6.15 in [2] 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 1 shows the results in terms of cumulative distribution functions. The plots in the first and second row ($\rho=-1$ and $\rho=-0.5$) show a strong and

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 Uniform(0, 1);
- Compute $S_k^{(2)} = S_k^{(1)}\left(U_k^{-\theta/(1+\theta)}-1\right)^{-1/\theta}$;
- Compute $W_{j,k} = (S_k^{(j)}\,\sigma\,\Gamma(1-\sigma))^{-1/\sigma}$ 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 $\tilde{p}_2 \approx \sum_{k=1}^K \bar{W}_{2,k}\delta_{\theta_{2,k}}$

mild negative correlation between the observables, represented by the opposite behaviour of $\tilde{p}_2$ and $\tilde{p}_1$. While $\tilde{p}_1$ associates high probabilities to positive values, $\tilde{p}_2$ tends to associate high probabilities to negative values. While ρ increases, first the conditional distribution of $\tilde{p}_2$ becomes independent from $\tilde{p}_1$ ($\rho = 0$) and then shows a behaviour similar to that of $\tilde{p}_1$ ($\rho = 0.5$ and $\rho = 1$), corresponding to positive correlation of the observables.

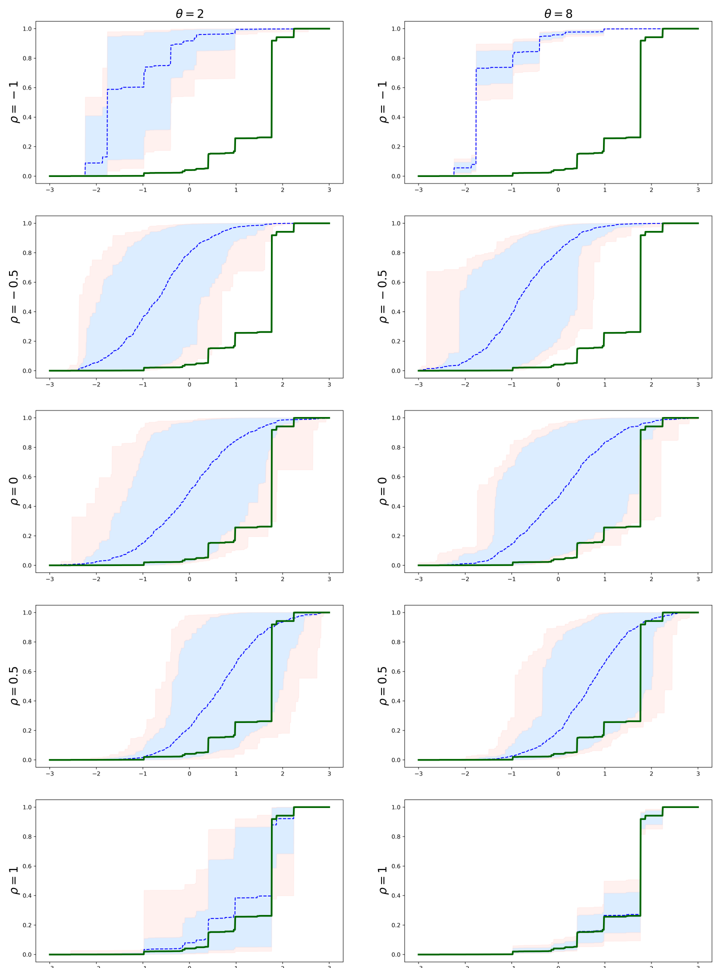

Fig. 1 Green solid line: a realization of the cumulative distribution function (cdf) corresponding to $\tilde{p}_1$, i.e. $\int_{-\infty}^{x} \tilde{p}_1(\mathrm{d}x)$. Blue dashed lines: conditional expected value of the cdf corresponding to $\tilde{p}_2$, given the realization of $\tilde{p}_1$, i.e. $\mathbb{E}\left[\int_{-\infty}^{x} \tilde{p}_2(\mathrm{d}x) \mid \tilde{p}_1\right]$. Light blue shaded area: 95% pointwise credible interval for the cdf corresponding to $\tilde{p}_2$. Pink shaded area: 99% pointwise credible interval for the cdf corresponding to $\tilde{p}_2$.

References

1. Ascolani, F., Franzolini, B., Lijoi, A., Prünster, I.: Dependent nonparametric models inducing negative correlation. Technical report. (2021)
2. Cont, R., Tankov, P.: Financial modelling with jump processes. Chapman & Hall/CRC, Boca Raton, FL (2004)
3. de Finetti, B.: Sur la condition d'equivalence partielle. Actual. Sci. Ind. **739**, 5–18 (1938)
4. Epifani, I., Lijoi, A.: Nonparametric priors for vectors of survival functions. Stat. Sin. **20**, 1455–1484 (2010)
5. Foti, N.J., Williamson, S.A.: A survey of non-exchangeable priors for Bayesian nonparametric models. IEEE Trans. Pattern Anal. Mach. Intell. **37**, 359–371 (2013)
6. James, L.F., Lijoi, A., Prünster, I.: Posterior analysis for normalized random measures with independent increments. Scand. J. Stat. **36**, 76–97 (2009)
7. Kallsen, J., Tankov, P.: Characterization of dependence of multidimensional Lévy processes using Lévy copulas. J. Multivar. Anal. **97**, 1551–1572 (2006)
8. Kingman, J.F.: Completely random measures. Pac. J. Math. **21**, 59–78 (1967)
9. Kingman, J.F.: Poisson Processes. Oxford Studies in Probability. Oxford University Press, Oxford (1993)
10. Palacio, A. R., Leisen, F.: Bayesian nonparametric estimation of survival functions with multiple-samples information. Electron. J. Stat., **12**, 1330–1357 (2018)
11. Quintana, F. A., Mueller, P., Jara, A., MacEachern, S. N.: The dependent Dirichlet process and related models. arXiv preprint arXiv:2007.06129. (2020)
12. Regazzini, E., Lijoi, A., Prünster, I.: Distributional results for means of normalized random measures with independent increments. Ann. Stat. **31**, 560–585 (2003)