
Bayesian nonparametric models for bipartite graphs: Supplementary Material

François Caron
 INRIA
 IMB - University of Bordeaux
 Talence, France
 Francois.Caron@inria.fr

A Properties of the generalized gamma process

Process	$\lambda(w)$	$\psi(t)$	$\kappa(n, z)$
Generalized gamma	$\frac{\alpha}{\Gamma(1-\sigma)} w^{-\sigma-1} e^{-w\tau}$	$\frac{\alpha}{\sigma} ((t + \tau)^\sigma - \tau^\sigma)$	$\frac{\alpha}{(z+\tau)^{n-\sigma}} \frac{\Gamma(n-\sigma)}{\Gamma(1-\sigma)}$
Gamma ($\sigma = 0$)	$\alpha w^{-1} e^{-w\tau}$	$\alpha \log(1 + \frac{t}{\tau})$	$\frac{\alpha}{(z+\tau)^n} \Gamma(n)$
Inverse Gaussian ($\sigma = \frac{1}{2}$)	$\frac{\alpha}{\sqrt{\pi}} w^{-\frac{3}{2}} e^{-w\tau}$	$2\alpha(\sqrt{t + \tau} - \sqrt{\tau})$	$\frac{\alpha}{(z+\tau)^{n-1/2}} \frac{\Gamma(n-1/2)}{\sqrt{\pi}}$
Stable ($\tau = 0$)	$\frac{\alpha}{\Gamma(1-\sigma)} w^{-\sigma-1}$	$\frac{\alpha t^\sigma}{\sigma}$	$\frac{\alpha}{z^{n-\sigma}} \frac{\Gamma(n-\sigma)}{\Gamma(1-\sigma)}$

Table 1: Expressions of $\lambda(w)$, $\psi(t)$ and $\kappa(n, z)$ for the generalized gamma, gamma, inverse Gaussian and stable processes.

B Proof of Proposition 1

Z_i is clearly a Poisson process, as obtained from transformations of Poisson processes. From Campbell theorem [3], as the Lévy intensity λ verifies conditions (2), $Z_i(\Theta)$ is finite with probability one. Moreover, we have

$$\begin{aligned}
 \mathbb{E}[Z_i(\Theta)] &= \mathbb{E} \left[\sum_{j=1}^{\infty} z_{ij} \right] = \mathbb{E} \left[\sum_{j=1}^{\infty} \mathbb{E}[z_{ij} | w_j] \right] \\
 &= \mathbb{E} \left[\sum_{j=1}^{\infty} (1 - \exp(-\gamma_i w_j)) \right] = \int_0^{\infty} (1 - \exp(-\gamma_i w)) \lambda(w) dw
 \end{aligned}$$

C Proof of Proposition 2

The proof is the same as that of Theorem 1 in [1] and is included here for completeness.

The marginal probability (11) is obtained by taking the expectation of (10) with respect to G . Note however that (10) is a density, so to be totally precise here we need to work with the *probability* of infinitesimal neighborhoods around the observations instead, which introduces significant notational complexity. To keep the notation simple, we will work with densities, leaving it to the careful reader

to verify that the calculations indeed carry over to the case of probabilities.

$$\begin{aligned} P(U_1, \dots, U_n) &= \mathbb{E}[P(U_1, \dots, U_n|G)] \\ &= \left(\prod_{i=1}^n \gamma_i^{K_i} \right) \mathbb{E} \left[\left[\prod_{j=1}^K w_j^{m_j} \exp \left(-w_j \sum_{i=1}^n \gamma_i (u_{ij} - 1) \right) \right] \exp \left(- \left(\sum_{i=1}^n \gamma_i \right) G(\Theta) \right) \right] \end{aligned}$$

The gamma prior on $G = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}$ is equivalent to a Poisson process prior on $N = \sum_{j=1}^{\infty} \delta_{(w_j, \theta_j)}$ defined over the space $\mathbb{R}^+ \times \Theta$ with mean intensity $\lambda(w)h(\theta)$. Then, dividing by the constant term $\left(\prod_{i=1}^n \gamma_i^{K_i} \right)$

$$\propto \mathbb{E} \left[e^{-\int w N(dw, d\theta)(\sum_{i=1}^n \gamma_i)} \prod_{k=1}^K \sum_{j=1}^{\infty} w_j^{m_k} \mathbf{1}_{\theta_j = \theta_k} e^{-w_j \sum_{i=1}^n \gamma_i (u_{ij} - 1)} \right]$$

Applying the Palm formula for Poisson processes to pull the $k = 1$ term out of the expectation,

$$\begin{aligned} &= \int \mathbb{E} \left[e^{-\int w (N + \delta_{w_1, x_1})(dw, d\theta)(\sum_{i=1}^n \gamma_i)} \prod_{k=2}^K \sum_{j=1}^{\infty} w_j^{m_k} \mathbf{1}_{\theta_j = \theta_k} e^{-w_j \sum_{i=1}^n \gamma_i (u_{ij} - 1)} \right] \\ &\quad \times (w_1)^{m_1} h(\theta_1) e^{-w_1 \sum_{i=1}^n \gamma_i (u_{i1} - 1)} \lambda(w_1) dw_1 \\ &= \mathbb{E} \left[e^{-\int w N(dw, d\theta)(\sum_{i=1}^n \gamma_i)} \prod_{k=2}^K \sum_{j=1}^{\infty} w_j^{m_k} \mathbf{1}_{\theta_j = \theta_k} e^{-w_j \sum_{i=1}^n \gamma_i (u_{ij} - 1)} \right] \\ &\quad \times h(\theta_1) \int (w_1)^{m_1} e^{-w_1 \sum_{i=1}^n \gamma_i u_{i1}} \lambda(w_1) dw_1 \end{aligned}$$

Now iteratively pull out terms $k = 2, \dots, K$ using the same idea, and we get:

$$\begin{aligned} &= \mathbb{E} \left[e^{-G(\Theta)(\sum_{i=1}^n \gamma_i)} \right] \prod_{k=1}^K h(\theta_k) \int (w_k)^{m_k} e^{-w_k \sum_{i=1}^n \gamma_i u_{ik}} \lambda(w_k) dw_k \\ &= e^{-\psi(\sum_{i=1}^n \gamma_i)} \prod_{k=1}^K h(\theta_k) \kappa \left(m_k, \sum_{i=1}^n \gamma_i u_{ik} \right) \end{aligned}$$

This completes the proof of Proposition 2.

D Proof of Proposition 3

The proof is the same as that of Theorem 2 in [1] and is included here for completeness. Let $f : \mathbb{X} \rightarrow \mathbb{R}$ be measurable with respect to H . Then the characteristic functional of the posterior G is given by:

$$\mathbb{E}[e^{-\int f(\theta)G(d\theta)} | U_1, \dots, U_n] = \frac{\mathbb{E}[e^{-\int f(\theta)G(d\theta)} P(U_1, \dots, U_n|G)]}{\mathbb{E}[P(U_1, \dots, U_n|G)]}$$

The denominator is as given in Proposition 2, while the numerator is obtained using the same Palm formula technique as Proposition 2, with the inclusion of the term $e^{-\int f(\theta)G(d\theta)}$. Some algebra shows that the resulting characteristic functional of the posterior G coincides with that of (12).

E Conditional distribution of u for the GGP

For the generalized gamma process, we have

$$p(u_{n+1, j} | \text{other}) = \frac{1}{C} \frac{1}{(u_{n+1, j} \gamma_{n+1} + \sum_{i=1}^n \gamma_i u_{ij} + \tau)^{m_j + 1 - \sigma}} \quad (1)$$

where we can compute the normalizing constant, which is given by

$$C = \begin{cases} \frac{(\sum_{i=1}^n \gamma_i u_{ij} + \tau)^{-m_j + \sigma} - (\sum_{i=1}^n \gamma_i u_{ij} + \tau + \gamma_{n+1})^{-m_j + \sigma}}{\gamma_{n+1} (m_j - \sigma)} & \text{if } m_j - \sigma \neq 0 \\ \frac{1}{\gamma_{n+1}} \log \left(1 + \frac{\gamma_{n+1}}{\tau + \sum_{i=1}^n \gamma_i u_{ij}} \right) & \text{if } m_j - \sigma = 0 \end{cases}$$

We can sample exactly from the above equation using the inverse cdf transform.

Let $s_{jn} = \sum_{i=1}^n \gamma_i u_{ij}$. Let $P(u_{n+1,j} < x | \text{rest}) = F(x)$. We have for GGP

$$F(x) = \begin{cases} \frac{(s_{jn} + \tau)^{-m_j + \sigma} - (s_{jn} + \tau + x \gamma_{n+1})^{-m_j + \sigma}}{(s_{jn} + \tau)^{-m_j + \sigma} - (s_{jn} + \tau + \gamma_{n+1})^{-m_j + \sigma}} & \text{if } m_j - \sigma \neq 0 \\ \frac{\log \left(1 + \frac{x \gamma_{n+1}}{\tau + s_{jn}} \right)}{\log \left(1 + \frac{\gamma_{n+1}}{\tau + s_{jn}} \right)} & \text{if } m_j - \sigma = 0 \end{cases}$$

and

$$F^{-1}(y) = \frac{1}{\gamma_{n+1}} \left[\left((\tau + s_{jn})^{-m_j + \sigma} + y \left((\tau + s_{jn} + \gamma_{n+1})^{-m_j + \sigma} - (\tau + s_{jn})^{-m_j + \sigma} \right) \right)^{1/(-m_j + \sigma)} - (\tau + s_{jn}) \right]$$

if $m_j - \sigma \neq 0$ and

$$F^{-1}(y) = \frac{\tau + s_{jn}}{\gamma_{n+1}} \left(\left(1 + \frac{\gamma_{n+1}}{\tau + s_{jn}} \right)^y - 1 \right)$$

otherwise.

F Inference in latent factor models

Let assume here that the binary matrix Z is unknown and that we have some likelihood function $p(\mathcal{D}|Z)$ where \mathcal{D} is the set of data. The Gibbs sampler for approximating the posterior distribution $p(Z, U, w, G^*(\Theta) | \mathcal{D})$ iterates between the following steps

- Update $U|Z, w$ as in Section 2.5
- Update $Z|w$
- Update $(w, G^*(\Theta))|U, Z$ as in Section 2.5

We now describe the update of $Z|w$. For $i = 1, \dots, n$, let K_{-i} be the total number of features in $\{Z_k\}_{k \neq i}$. Given $\{w_j\}_{j=1, \dots, K_{-i}}$ and $G^*(\Theta)$, we have for $j = 1, \dots, K_{-i}$

$$p(z_{ij} | w_j, \mathcal{D}, z_{-ij}) \propto (1 - \exp(-\gamma_i w_j))^{z_{ij}} \exp(-(1 - z_{ij}) \gamma_i w_j) p(\mathcal{D}|Z)$$

and the total number of new features K_i^+ for object i is obtained by

$$p(K_i^+ | \mathcal{D}, \{Z_k\}_{k \neq i}) \propto \text{Poisson} \left(K_i^+; \tilde{\psi}_\lambda \left(\gamma_i, \sum_{k \neq i} \gamma_k \right) \right) p(\mathcal{D}|Z)$$

from which we can sample by truncating the infinite sum on the right hand-side, as for the IBP model [2]. Note that exact sampling techniques based on slice sampling may also be employed [5].

G Marginal distribution

When $\gamma_i = \gamma$, it is possible to marginalize over the CRM G as well as the latent variables U so as to obtain an analytical expression for the joint distribution of $P(Z_1, \dots, Z_n)$ as well as the predictive $P(Z_{n+1} | Z_1, \dots, Z_n)$. Note that in this case, we have a Poisson degree distribution for readers. We provide the proof in the following section, which relies on properties of Poisson processes. In the case of the gamma process, it is also possible to derive this result by taking the limit of a finite model. As this construction may appear more intuitive, we also include it here for completeness.

G.1 Construction using CRM

We have

$$\begin{aligned}
P(Z_1, \dots, Z_n | G) &= \prod_{j=1}^K [(1 - \exp(-\gamma w_j))^{m_j} \exp(-\gamma w_j(n - m_j))] \times \exp(-\gamma n G(\Theta \setminus \{\theta_1, \dots, \theta_K\})) \\
&= \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \exp(-k w_j \gamma) \right] \exp(-\gamma w_j(n - m_j)) \times \exp(-\gamma n G(\Theta \setminus \{\theta_1, \dots, \theta_K\})) \\
&= \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \exp(-w_j \gamma(n - m_j + k)) \right] \times \exp(-\gamma n G(\Theta \setminus \{\theta_1, \dots, \theta_K\})) \\
&= \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^{k+1} (1 - \exp(-w_j \gamma(n - m_j + k))) \right] \times \exp(-\gamma n G(\Theta \setminus \{\theta_1, \dots, \theta_K\}))
\end{aligned}$$

Applying the Palm formula to the above yields

$$P(Z_1, \dots, Z_n) = \prod_{j=1}^K \left[h(\theta_j) \sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^{k+1} \psi_\lambda(\gamma(n - m_j + k)) \right] \times \exp(-\psi_\lambda(n\gamma))$$

Hence the conditional distribution of Z_{n+1} given Z_1, \dots, Z_n is

$$Z_{n+1} = Z_{n+1}^* + \sum_{j=1}^K z_{n+1,j} \delta_{\theta_j}$$

where

$$z_{n+1,j} | Z_1, \dots, Z_n \sim \text{Ber} \left(\frac{\sum_{k=0}^{m_j+1} \binom{m_j+1}{k} (-1)^{k+1} \psi_\lambda(\gamma(n - m_j + k))}{\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^{k+1} \psi_\lambda(\gamma(n - m_j + k))} \right)$$

while the number of new elements Z_{n+1}^* is a Poisson process over Θ of intensity measure $(\psi_\lambda((n+1)\gamma) - \psi_\lambda(n\gamma))h(\theta)$ as shown in corollary 4.

G.2 Construction by taking the limit of a finite model

In the gamma process case, we can derive the above construction from the limit of a finite model, similarly to the original construction of the Indian buffet process [2]. Let us consider the finite model defined for $i = 1, \dots, n$ and $j = 1, \dots, p$ by

$$z_{ij} | w_j \sim \text{Ber}(1 - \exp(-\gamma w_j))$$

and for $j = 1, \dots, p$

$$w_j \sim \text{Gamma} \left(\frac{\alpha}{p}, \tau \right)$$

Then we have

$$p(z|w) = \prod_{j=1}^p (1 - \exp(-w_j \gamma))^{m_j} (\exp(-w_j \gamma))^{n - m_j}$$

Using the identity $(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k$, we obtain

$$\begin{aligned}
p(z|w) &= \prod_{j=1}^p \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \exp(-k w_j \gamma) \right] (\exp(-w_j \gamma))^{n - m_j} \\
&= \prod_{j=1}^p \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \exp(-w_j \gamma(n - m_j + k)) \right]
\end{aligned}$$

and then

$$\begin{aligned}
p(z) &= \mathbb{E}_w[P(z|w)] \\
&= \prod_{j=1}^p \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \mathbb{E}_{w_j} [\exp(-w_j \gamma(n - m_j + k))] \right] \\
&= \prod_{j=1}^p \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right)^{-\alpha/p} \right]
\end{aligned}$$

Similarly to the Indian buffet construction [2], we take the equivalence class $[z]$ defined by left-ordered binary matrices. We write $j = 1, \dots, K$ for the features j such that $m_j > 0$ having at least one feature and obtain

$$\begin{aligned}
p([z]) &= \frac{p!}{\prod_{h=0}^{2^n-1} p_h} \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right)^{-\alpha/p} \right] \prod_{j=K+1}^p \left(1 + \frac{\gamma n}{\tau} \right)^{-\alpha/p} \\
&= \frac{p!}{\prod_{h=0}^{2^n-1} p_h} \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right)^{-\alpha/p} \right] \left(1 + \frac{\gamma n}{\tau} \right)^{-\alpha(p-K)/p}
\end{aligned}$$

where p_h is the count of the number of columns with full history h , see [2] for details. When p is large, $\left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right)^{-\alpha/p} \simeq 1 - \frac{\alpha}{p} \log \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right)$, and using the fact that $\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^k = 0$, we obtain, for p large

$$\begin{aligned}
p([z]) &= \frac{p!}{\prod_{h=0}^{2^n-1} p_h} \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^{k+1} \frac{\alpha}{p} \log \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right) \right] \left(1 + \frac{\gamma n}{\tau} \right)^{-\alpha} \\
&= \frac{\alpha^K}{\prod_{h=1}^{2^n-1} p_h} \frac{p!}{p_0 p^K} \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^{k+1} \log \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right) \right] \left(1 + \frac{\gamma n}{\tau} \right)^{-\alpha} \\
&= \frac{\alpha^K}{\prod_{h=1}^{2^n-1} p_h} \prod_{j=1}^K \left[\sum_{k=0}^{m_j} \binom{m_j}{k} (-1)^{k+1} \log \left(1 + \frac{\gamma(n - m_j + k)}{\tau} \right) \right] \left(1 + \frac{\gamma n}{\tau} \right)^{-\alpha}
\end{aligned}$$

as $\frac{p!}{p_0 p^K} \rightarrow 1$ when $p \rightarrow \infty$, which completes the proof.

H Derivation of power-law properties

In this section, we derive power-law properties of the proposed model, in the case $\gamma_i = \gamma$. The proofs are similar to those for the stable IBP (see Appendix A of [5]).

The total number of books follows a Poisson distribution of rate $\psi_\lambda(n\gamma)$. For the GGP, we have

$$\psi_\lambda(n\gamma) = \frac{\alpha}{\sigma} ((n\gamma + \tau)^\sigma - \tau^\sigma)$$

which for large n , is of order n^σ .

For the degree distribution, we are interested in the joint distribution of (M_1, \dots, M_n) , where M_m is the number of books read by exactly m readers. There are $\frac{K!}{\prod_{m=1}^n M_m!} \prod_{m=1}^n \left(\frac{n!}{m!(n-m)!} \right)^{M_m}$ configurations of the model with the same statistics (M_1, \dots, M_n) . Hence, by using Eq. (11) and the fact that as $m \ll n$, we have $\gamma \sum_{i=1}^n u_{ij} \simeq n\gamma$, we obtain

$$P(M_1, \dots, M_n) \propto \frac{K!}{\prod_{m=1}^n M_m!} \prod_{m=1}^n \left(\frac{n!}{m!(n-m)!} \gamma^m \kappa(m, n\gamma) \right)^{M_m}$$

Conditioning on $\sum_{m=1}^n M_m$, (M_1, \dots, M_n) is multinomial with the probability of having a book read by m readers being proportional to the term in parentheses. For large m and even larger n , it simplifies to $O\left(\frac{\Gamma(m-\sigma)}{\Gamma(m+1)}\right) = O(m^{-1-\sigma})$.

I Derivation of the Gibbs sampler from the limit of a finite model

In this section, we derive the Gibbs sampler, in the gamma process case, as the limit of a finite model. Similar constructions were given, e.g. for the Dirichlet process[4] or the beta-Bernoulli process [2]. Let us consider the finite model defined for $i = 1, \dots, n$ and $j = 1, \dots, p$ by

$$z_{ij}|w_j \sim \text{Ber}(1 - \exp(-\gamma_i w_j))$$

and for $j = 1, \dots, p$

$$w_j \sim \text{Gamma}\left(\frac{\alpha}{p}, \tau\right)$$

For $i = 1, \dots, n$ and $j = 1, \dots, p$, we introduce latent variables u_{ij} such that

$$u_{ij} \sim \text{rExp}(\gamma_i w_j, 1)$$

if $z_{ij} = 1$, and 1 otherwise. We therefore have

$$\begin{aligned} p(z_{ij} = 0, u_{ij}|w_j) &= \exp(-\gamma_i w_j) \\ p(z_{ij} = 1, u_{ij}|w_j) &= \gamma_i w_j \exp(-\gamma_i w_j u_{ij}) \end{aligned}$$

Let $K = \sum_{i=1}^n \sum_{j=1}^p z_{ij}$ and $m_j = \sum_{i=1}^n z_{ij}$. Without loss of generality, assume that $m_j > 0$ for $j = 1, \dots, K$ and $m_j = 0$ for $j = K+1, \dots, p$. Let $w_* = \sum_{j=K+1}^p w_j$, then we have the following updates. For $j = 1, \dots, K$

$$w_j|z_{ij}, u_{ij} \sim \text{Gamma}\left(\frac{\alpha}{p} + m_j, \tau + \sum_{i=1}^n \gamma_i u_{ij}\right)$$

and

$$w_*|z, u \sim \text{Gamma}\left(\alpha \frac{p-K}{p}, \tau + \sum_{i=1}^n \gamma_i\right)$$

Taking the limit when $p \rightarrow \infty$ yields

$$w_j|z_{ij}, u_{ij} \sim \text{Gamma}\left(m_j, \tau + \sum_{i=1}^n \gamma_i u_{ij}\right)$$

and

$$w_*|z, u \sim \text{Gamma}\left(\alpha, \tau + \sum_{i=1}^n \gamma_i\right)$$

and one recovers the Gibbs sampler for the gamma process described in Section 2.5.

References

- [1] F. Caron and Y. W. Teh. Bayesian nonparametric models for ranked data. In *Neural Information Processing Systems (NIPS)*, 2012.
- [2] T Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *NIPS*, 2005.
- [3] J.F.C. Kingman. *Poisson processes*, volume 3. Oxford University Press, USA, 1993.
- [4] R.M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, pages 249–265, 2000.
- [5] Y.W. Teh and D. Görür. Indian buffet processes with power-law behavior. In *NIPS*, 2009.