Bayesian nonparametric models for ranked data: Supplementary Material

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

Yee Whye Teh
Department of Statistics
University of Oxford
Oxford, United Kingdom
y.w.teh@stats.ox.ac.uk

A   Proof of Theorem 1

The marginal probability (14) is obtained by taking the expectation of (13) with respect to G. Note however that (13) 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.

$$P((Y_\ell, Z_\ell)_{\ell=1}^L) = \mathbb{E}\left[P((Y_\ell, Z_\ell)_{\ell=1}^L | G)\right]$$

$$= \mathbb{E}\left[e^{-G(\mathbb{X})\sum_{\ell i} Z_{\ell i}} \prod_{k=1}^K G(\{X_k^*\})^{n_k} e^{-G(\{X_k^*\})\sum_{\ell i}(\delta_{\ell i k}-1)Z_{\ell i}}\right]$$

The gamma prior on $G = \sum_{j=1}^\infty w_j \delta_{X_j}$ is equivalent to a Poisson process prior on $N = \sum_{j=1}^\infty \delta_{(w_j, X_j)}$ defined over the space $\mathbb{R}^+ \times \mathbb{X}$ with mean intensity $\lambda(w)h(x)$. Then,

$$= \mathbb{E}\left[e^{-\int wN(dw,dx)\sum_{\ell i} Z_{\ell i}} \prod_{k=1}^K \sum_{j=1}^\infty w_j^{n_k} \mathbb{1}(X_j = X_k^*) e^{-w_j \sum_{\ell i}(\delta_{\ell i k}-1)Z_{\ell i}}\right]$$

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

$$= \int \mathbb{E}\left[e^{-\int w(N+\delta_{w_1^*, x_1^*})(dw,dx)\sum_{\ell i} Z_{\ell i}} \prod_{k=2}^K \sum_{j=1}^\infty w_j^{n_k} \mathbb{1}(X_j = X_k^*) e^{-w_j \sum_{\ell i}(\delta_{\ell i k}-1)Z_{\ell i}}\right]$$
$$\times (w_1^*)^{n_k} h(X_1^*) e^{-w_1^* \sum_{\ell i}(\delta_{\ell i 1}-1)Z_{\ell i}} \lambda(w_1^*) dw_1^*$$

$$= \mathbb{E}\left[e^{-\int wN(dw,dx)\sum_{\ell i} Z_{\ell i}} \prod_{k=2}^K \sum_{j=1}^\infty w_j^{n_k} \mathbb{1}(X_j = X_k^*) e^{-w_j \sum_{\ell i}(\delta_{\ell i k}-1)Z_{\ell i}}\right]$$
$$\times h(X_1^*) \int (w_1^*)^{n_k} e^{-w_1^* \sum_{\ell i} \delta_{\ell i 1} Z_{\ell i}} \lambda(w_1^*) dw_1^*$$

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

$$= \mathbb{E}\left[e^{-G(\mathbb{X})\sum_{\ell i} Z_{\ell i}}\right] \prod_{k=1}^K h(X_k^*) \int (w_k^*)^{n_k} e^{-w_k^* \sum_{\ell i} \delta_{\ell i k} Z_{\ell i}} \lambda(w_k^*) dw_k^*$$

$$= e^{-\psi(\sum_{\ell i} Z_{\ell i})} \prod_{k=1}^K h(X_k^*) \kappa\left(n_k, \sum_{\ell i} \delta_{\ell i k} Z_{\ell i}\right) \tag{1}$$

This completes the proof of Theorem 1.

B Proof of Theorem 2

Let $f : \mathbb{X} \to \mathbb{R}$ be measurable with respect to H. Then the characteristic functional of the posterior G is given by:

$$\mathbb{E}[e^{-\int f(x)G(dx)}|(Y_\ell, Z_\ell)_{\ell=1}^L] = \frac{\mathbb{E}[e^{-\int f(x)G(dx)}P((Y_\ell, Z_\ell)_{\ell=1}^L|G)]}{\mathbb{E}[P((Y_\ell, Z_\ell)_{\ell=1}^L|G)]} \tag{2}$$

The proof is essentially obtained by calculating the numerator and denominator of (2). The denominator is already given in Theorem 1. The numerator is obtained using the same technique with the inclusion of the term $e^{\int f(x)G(dx)}$, which gives:

$$\mathbb{E}\left[e^{-\int f(x)G(dx)}P((Y_\ell, Z_\ell)_{\ell=1}^L|G)\right]$$

$$=\mathbb{E}\left[e^{-\int(f(x)+\sum_{\ell i} Z_{\ell i})G(dx)}\right]\prod_{k=1}^K h(X_k^*)\int (w_k^*)^{n_k}e^{-w_k^*(f(X_k^*)+\sum_{\ell i}\delta_{\ell i k}Z_{\ell i})}\lambda(w_k^*)dw_k^*$$

By the Lévy-Khintchine Theorem (using the fact that G has a Poisson process representation N),

$$=\exp\left(-\int(1-e^{-w(f(x)+\sum_{\ell i} Z_{\ell i})})\lambda(w)h(x)dwdx\right)$$

$$\times\prod_{k=1}^K h(X_k^*)\int (w_k^*)^{n_k}e^{-w_k^*(f(X_k^*)+\sum_{\ell i}\delta_{\ell i k}Z_{\ell i})}\lambda(w_k^*)dw_k^* \tag{3}$$

Dividing the numerator (1) by the denominator (3), the characteristic functional of the posterior G is:

$$\mathbb{E}\left[e^{-\int f(x)G(dx)}|(Y_\ell, Z_\ell)_{\ell=1}^L\right]$$

$$=\exp\left(-\int(1-e^{-wf(x)})e^{-\sum_{\ell i} Z_{\ell i}}\lambda(w)h(x)dwdx\right)$$

$$\times\prod_{k=1}^K h(X_k^*)\frac{\int e^{-f(X_k^*)}(w_k^*)^{n_k}e^{-w_k^*\sum_{\ell i}\delta_{\ell i k}Z_{\ell i}}\lambda(w_k^*)dw_k^*}{\int(w_k^*)^{n_k}e^{-w_k^*\sum_{\ell i}\delta_{\ell i k}Z_{\ell i}}\lambda(w_k^*)dw_k^*} \tag{4}$$

Since the characteristic functional is the product of $K+1$ terms, we see that the posterior G consists of $K+1$ independent components, one corresponding to the first term above (G^*), and the others corresponding to the K terms in the product over k. Substituting the Lévy measure $\lambda(w)$ for a gamma process, we note that the first term shows that G^* is a gamma process with updated inverse scale τ^*. The kth term in the product shows that the corresponding component is an atom located at X_k^* with density $(w_k^*)^{n_k}e^{-w_k^*\sum_{\ell i}\delta_{\ell i k}Z_{\ell i}}\lambda(w_k^*)$; this is the density of the gamma distribution over w_k^* in Theorem 2. This completes the proof.

C Proof of Proposition 4

We have

$$P(G_t(X_{1k}) = 0|w_{t-1,,k}) = \exp(-\phi_{t-1}w_{t-1,k})$$

Assume that

$$P(G_t(X_{1k}) = 0|w_{sk}) = \exp(-y_{t|s}w_{sk})$$

then

$$P(G_t(X_{1k}) = 0|w_{s-1,k}) = \int \exp(-y_{t|s}w_{sk})p(w_{sk}|w_{s-1,k})dw_{sk}$$

$$= \sum_{c_{s-1,k}} \int \exp(-y_{t|s}w_{sk})p(w_{sk}|c_{s-1,k})p(c_{s-1,k}|w_{s-1,k})dw_{sk}$$

$$= \sum_{c_{s-1,k}} \exp\left[-c_{s-1,k}\log\left(1 + \frac{y_{t|s}}{\phi_{s-1} + \tau}\right)\right]p(c_{s-1,k}|w_{s-1,k})$$

$$= \exp\left(\frac{-y_{t|s}\phi_{s-1}}{\phi_{s-1} + \tau + y_{t|s}}w_{s-1,k}\right)$$

D Gibbs sampler for the dynamic nonparametric Plackett-Luce model

For ease of presentation, we assume that ϕ_t takes the same value ϕ at each time step. The Gibbs sampler will iterate between the following steps

1. a. For $t = 1, \ldots, T$, update $G_t(\mathbb{X})$ given $(G_{t-1}(\mathbb{X}), \alpha, \phi)$
   b. For $t = 1, \ldots, T$, update (c_t, c_{t*}) given $(w_t, w_{t*}, w_{t+1}, w_{t+1*}, \phi, \alpha)$
2. a. Update α given (Z, ϕ)
   b. For $t = 1, \ldots, T$
      Update w_{t*} given $(c_{t-1*}, Z, \phi, \alpha)$
      Update c_{t*} given $(w_{t*}, Z, \phi, \alpha)$
3. For $t = 1, \ldots, T$, update (w_t, w_{t*}) given $(c_{t-1}, c_{t-1*}, c_t, c_{t*}, Z_t, \alpha, \phi)$
4. For $t = 1, \ldots, T$, update Z_t given (w_t, w_{t*})
5. Update ϕ given w, w_*, α, ϕ

The steps are now fully described.

1.a) Sample $(G_t(\mathbb{X}))$ **given** (α, ϕ)

We have
$$G_1(\mathbb{X})|\alpha \sim \text{Gamma}(\alpha, \tau)$$
and for $t = 1, \ldots, T - 1$
$$G_{t+1}(\mathbb{X}) \sim \text{Gamma}(\alpha + M_t, \tau + \phi)$$
where $M_t \sim \text{Poisson}(\phi G_t(\mathbb{X}))$. The weights (w_t, w_{t*}) are then appropriately rescaled.

1.b) Sample (c, c_*) **given** (w, w_*, ϕ, α)

Consider first the sampling of $c_{1:T}$. We have, for $t = 1, \ldots, T$ and $k = 1, \ldots, K$
$$p(c_{tk}|w_{tk}, w_{t+1,k}) \propto p(c_{tk}|w_{tk})p(w_{t+1,k}|c_{tk})$$
where
$$p(c_{tk}|w_{tk}) = \text{Poisson}(c_{tk}; \phi w_{tk})$$
and
$$p(w_{t+1,k}|c_{tk}) = \begin{cases} \delta_0(w_{t+1,k}) & \text{if } w_{tk} = 0 \\ \text{Gamma}(w_{t+1,k}; c_{tk}, \tau + \phi) & \text{if } w_{tk} > 0 \end{cases}$$
Hence we can have the following MH update. If $w_{t+1,k} > 0$, then we necessarily have $c_{tk} > 0$. We sample $c_{tk}^* \sim \text{zPoisson}(\phi w_{tk})$ where $\text{zPoisson}(\phi w_{tk})$ denotes the zero-truncated Poisson distribution and accept c_{tk}^* w.p.
$$\min\left(1, \frac{\text{Gamma}(w_{t+1,k}; c_{tk}^*, \tau + \phi)}{\text{Gamma}(w_{t+1,k}; c_{tk}, \tau + \phi)}\right)$$

If $w_{t+1,k} = 0$, we only have two possible moves: $c_{tk} = 0$ or $c_{tk} = 1$, given by the following probabilities

$$P(c_{tk} = 0 | w_{t+1,k} = 0, w_{tk}) = \frac{\exp(-\phi w_{tk})}{\exp(-\phi w_{tk}) + \phi w_{tk} \exp(-\phi w_{tk})(\tau + \phi)} = \frac{1}{1 + \phi w_{tk}(\tau + \phi)}$$

$$P(c_{tk} = 1 | w_{t+1,k} = 0, w_{tk}) = \frac{\phi w_{tk} \exp(-\phi w_{tk})(\tau + \phi)}{\exp(-\phi w_{tk}) + \phi w_{tk} \exp(-\phi w_{tk})(\tau + \phi)} = \frac{\phi w_{tk}(\tau + \phi)}{1 + \phi w_{tk}(\tau + \phi)}$$

Note that the above Markov chain is not irreducible, as the probability is zero to go from a state $(c_{tk} > 0, w_{t+1,k} > 0)$ to a state $(c_{tk} = 0, w_{t+1,k} = 0)$, even though the posterior probability of this event is non zero in the case item k does not appear after time t. We can resolve that by the following procedure, that uses a backward forward recursion.

Assume that item k does not appear after time step τ_k^+ (the same procedure applies if item k does not appear before time step τ_k^-). Then we can sample jointly the whole sequence $(w_{k,t}, c_{k,t})_{t=\tau_k+1,\dots,T}$ using the following backward forward recursion.

Let

$$x_T = \sum_{k=1}^{m} Z_{Tk} \tag{5}$$

and for $t = T - 1, \dots, \tau_k^+$

$$x_t = \sum_{k=1}^{m} Z_{tk} + \frac{\phi x_{t+1}}{1 + \phi + x_{t+1}}$$

We have, for $k = 1, \dots, K$ and $t = \tau_k^+$

$$c_{tk} | (Z, \phi, w_{tk}) \sim \text{Poisson}\left(\frac{1 + \phi}{1 + \phi + x_t} \phi w_{tk}\right) \tag{6}$$

$$w_{t+1,k} | c_{tk}, Z \sim \text{Gamma}(c_{k,t}, \tau + \phi + x_{t+1}) \tag{7}$$

2.a) Sample α given (Z, ϕ)

We can sample from the full conditional which is given by

$$\alpha | (Z, \gamma, \phi) \sim \text{Gamma}(a + K, b + y_1 + \log(1 + x_1)) \tag{8}$$

where x_1 and y_1 are obtained with the following recursion

$$x_T = \sum_{k=1}^{m} Z_{Tk} \tag{9}$$

$$y_T = 0 \tag{10}$$

and for $t = T - 1, \dots, 1$

$$x_t = \sum_{k=1}^{m} Z_{tk} + \frac{\phi x_{t+1}}{1 + \phi + x_{t+1}}$$

$$y_t = y_{t+1} - \log\left(\frac{1 + \phi}{1 + \phi + x_{t+1}}\right)$$

.

2.b) Sample (c_*, w_*) given (Z, ϕ, α)

We can sample from the full conditional which is given by

$$w_{1*}|(Z, \phi, \alpha) \sim \text{Gamma}\left(\alpha, \tau + x_1\right) \qquad (11)$$

where x_1 is defined above. Then for $t = 2, \ldots, T$, let

$$c_{t-1*}|(Z, \phi, \alpha, w_{t-1*}) \sim \text{Poisson}\left(\frac{1+\phi}{1+\phi+x_t}\phi w_{t-1*}\right)$$

$$w_{t*}|c_{t-1*}, Z, \alpha \sim \text{Gamma}\left(\alpha + c_{t-1*}, \tau + \phi + x_t\right)$$

3) Sample (w, w_*) **given** $(Z, \alpha, c, c_*, \phi)$

For each time step $t = 1, \ldots, T$

- For each item $k = 1, \ldots, K$, sample

$$w_{tk}|c_{t-1,k}, c_{tk}, Z_t \sim \text{Gamma}\left(n_{tk} + c_{t-1,k} + c_{tk}, \tau + 2\phi + \sum_{i=1}^{m}\delta_{tik}Z_{ti}\right) \qquad (12)$$

  if $c_{tk} + c_{t-1,k} + n_{tk} > 0$, otherwise, set $w_{tk} = 0$. The occurence indicator δ_{tik} is defined as

$$\delta_{tik} = \begin{cases} 0 & \text{if } \exists j < i \text{ with } Y_{tj} = X_k^*; \\ 1 & \text{otherwise.} \end{cases} \qquad (13)$$

- Sample the total mass

$$w_{t*}|c_{t*}, c_{t-1*}, Z_t, \alpha \sim \text{Gamma}\left(\alpha + c_{t*} + c_{t-1*}, \tau + 2\phi + \sum_{i=1}^{m}Z_{ti}\right) \qquad (14)$$

4) Sample Z **given** (w, w_*)

For $t = 1, \ldots, T$ and $i = 1, \ldots m$, sample

$$Z_{ti}|w, w_* \sim \text{Exp}\left(w_{t*} + \sum_{k=1}^{K}\delta_{tik}w_{tk}\right) \qquad (15)$$

5) Sample ϕ **given** w, w_*, α, ϕ

We sample ϕ using a MH step. Propose $\widetilde{\phi} = \phi\exp(\sigma\varepsilon)$ where $\sigma > 0$ and $\varepsilon \sim \mathcal{N}(0, 1)$. And accept it with probability

$$\min\left(1, \frac{p(\widetilde{\phi})}{p(\phi)}\frac{\widetilde{\phi}}{\phi}\prod_{t=1}^{T-1}\left[\frac{p(w_{t+1*}|\widetilde{\phi}, w_{t*})}{p(w_{t+1*}|\phi, w_{t*})}\prod_{k=1}^{K}\frac{p(w_{t+1,k}|\widetilde{\phi}, w_{tk})}{p(w_{t+1,k}|\phi, w_{tk})}\right]\right) \qquad (16)$$