Pitman-Yor process with dependence on predictors

Claudio Del Sole

February 20, 2024

Consider a random measure $\tilde{\mu}$ on $\mathbb{X}$, and define the measure-valued stochastic process $\tilde{p} = \{\, \tilde{p}_x \colon x \in \mathbb{X} \,\}$ such that, for each $x \in \mathbb{X}$,

$$\tilde{p}_x(dz) = \tilde{p}(dz|x) := \frac{h(x,z)\, \tilde{\mu}(dz)}{\displaystyle\int_{\mathbb{X}} h(x,y)\, \tilde{\mu}(dy)},$$

is a random probability measure on $\mathbb{X}$, for a positive bounded kernel $h \colon \mathbb{X} \times \mathbb{X} \mapsto [0,1]$.

Introduce a collection of n observations, that is, predictor-response pairs,

$$(X_1, Z_1), \ldots, (X_n, Z_n),$$

where predictors $X^{(n)} = (X_1, \ldots, X_n)$ can possibly be all distinct, while responses $Z^{(n)} = (Z_1, \ldots, Z_n)$ assume k distinct values $Z_1^*, \ldots, Z_k^*$ with multiplicities $n_1 + \cdots + n_k = n$. Moreover, let $A^{(n)} = (A_1, \ldots, A_n)$ be the vector of labels corresponding to the responses $Z^{(n)}$, that is, $A_i = j$ if and only if $Z_i = Z_j^*$.

The likelihood function takes the form

$$\begin{aligned}
\mathcal{L}(\tilde{\mu}; X^{(n)}, Z^{(n)}) &= \prod_{i=1}^{n} \tilde{p}_{X_i}(dZ_i) = \prod_{i=1}^{n} \tilde{p}(dZ_i|X_i) = \prod_{i=1}^{n} \frac{h(X_i, Z_i)\, \tilde{\mu}(dZ_i)}{\displaystyle\int_{\mathbb{X}} h(X_i, y)\, \tilde{\mu}(dy)} \\
&= \prod_{i=1}^{n} h(X_i, Z_i)\, \tilde{\mu}(dZ_i) \int_{\mathbb{R}^+} \exp\left\{ - \int_{\mathbb{X}} u_i\, h(X_i, y)\, \tilde{\mu}(dy) \right\} du_i,
\end{aligned}$$

which is conveniently rewritten, in terms of labels $A^{(n)}$ and distinct values Z^*, as

$$\left(\prod_{i=1}^{n} h(X_i, Z_{A_i}^*) \right) \left(\prod_{j=1}^{k} \tilde{\mu}(dZ_j^*)^{n_j} \right) \int_{(\mathbb{R}^+)^n} \exp\left\{ - \int_{\mathbb{X}} u^{(n)} h(X^{(n)}, y)\, \tilde{\mu}(dy) \right\} du^{(n)},$$

where $u^{(n)} = (u_1, \ldots, u_n)$ and the notation $u^{(n)} h(X^{(n)}, y) = \sum_{i=1}^{n} u_i\, h(X_i, y)$ is adopted for convenience.

Prior specification Consider a σ-stable CRM, denoted by $\tilde{\mu}_\sigma$, having Lévy intensity measure

$$\nu(ds, dz) = \frac{\sigma}{\Gamma(1-\sigma)}\, s^{-1-\sigma}\, ds\, P_0(dz),$$

where P_0 is a diffuse probability measure on $\mathbb{X}$.

The random measure $\tilde{\mu}$ is defined through a change of measure with respect to $\tilde{\mu}_\sigma$. Specifically, the Radon-Nikodym derivative of the probability distribution of $\tilde{\mu}$ with respect to the probability distribution of $\tilde{\mu}_\sigma$ is

$$\begin{aligned}\frac{d\mathcal{L}(\tilde{\mu})}{d\mathcal{L}(\tilde{\mu}_\sigma)}(m) &= \frac{\sigma\Gamma(\theta)}{\Gamma(\theta/\sigma)}\left(\int_{\mathbb{X}} G_0(z)^\sigma P_0(dz)\right)^{\theta/\sigma}\left(\int_{\mathbb{X}} G_0(z)\, m(dz)\right)^{-\theta}\\ &= \frac{\sigma}{\Gamma(\theta/\sigma)}\left(\int_{\mathbb{X}} G_0(z)^\sigma P_0(dz)\right)^{\theta/\sigma}\int_{\mathbb{R}^+} u_0^{\theta-1}\exp\left\{-\int_{\mathbb{X}} u_0\, G_0(z)\, m(dz)\right\} du_0,\end{aligned}$$

where $G_0 \colon \mathbb{X} \mapsto \mathbb{R}^+$ is a non-negative measurable function. Note that the change of measure defining the Pitman-Yor process is obtained by choosing $G_0(z) = 1$.

Marginal distribution

Define the function

$$G(z; U_0, U^{(n)}) = U_0\, G_0(z) + U^{(n)} h(X^{(n)}, z).$$

The marginal distribution of responses $Z^{(n)}$ given predictors $X^{(n)}$ is

$$\begin{aligned}
&\mathbb{P}(Z^{(n)} \mid X^{(n)}) = \mathbb{P}(A^{(n)}, Z^* \mid X^{(n)}) \\
&= \frac{\sigma^{k+1}}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy)\right)^{\theta/\sigma} \left(\prod_{i=1}^n h(X_i, Z^*_{A_i})\right) \left(\prod_{j=1}^k \frac{\Gamma(n_j-\sigma)}{\Gamma(1-\sigma)}\, P_0(dZ^*_j)\right) \\
&\int_{(\mathbb{R}^+)^n} \exp\left\{-\int_{\mathbb{X}} G(y; u_0, u^{(n)})^\sigma P_0(dy)\right\} \left(\prod_{j=1}^k G(Z^*_j; u_0, u^{(n)})^{-n_j+\sigma}\right) u_0^{\theta-1}\, du_0\, du^{(n)}.
\end{aligned}$$

Note that the conditioning on predictors $X^{(n)}$ is not a formal conditioning, since predictors are not random variables. The integration above can be considered a marginalization with respect to a vector of additional non-negative latent variables $(U_0, U_1, \ldots, U_n)$, so that the augmented marginal distribution is

$$\begin{aligned}
&\mathbb{P}(Z^{(n)}, U_0, U^{(n)} \mid X^{(n)}) = \mathbb{P}(A^{(n)}, Z^*, U_0, U^{(n)} \mid X^{(n)}) \\
&= \frac{\sigma^{k+1}}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy)\right)^{\theta/\sigma} \left(\prod_{i=1}^n h(X_i, Z^*_{A_i})\right) \left(\prod_{j=1}^k \frac{\Gamma(n_j-\sigma)}{\Gamma(1-\sigma)}\, P_0(dZ^*_j)\right) \\
&\qquad \exp\left\{-\int_{\mathbb{X}} G(y; U_0, U^{(n)})^\sigma P_0(dy)\right\} \left(\prod_{j=1}^k G(Z^*_j; U_0, U^{(n)})^{-n_j+\sigma}\right) U_0^{\theta-1}.
\end{aligned}$$

Consider the change of variables $(u_0, u^{(n)}) \to (t, v_0, v^{(n)})$ defined as

$$t = u_0 + \sum_{i=1}^n u_i, \qquad v_i = \frac{u_i}{t}, \qquad du_0\, du^{(n)} = t^n\, dt\, dv_0\, dv^{(n)}.$$

Computing the integral with respect to t, the marginal distribution above is rewritten as

$$\begin{aligned}
&\mathbb{P}(Z^{(n)} \mid X^{(n)}) = \mathbb{P}(A^{(n)}, Z^* \mid X^{(n)}) = \sigma^k \frac{\Gamma(k+\theta/\sigma)}{\Gamma(\theta/\sigma)} \frac{\Gamma(\theta)}{\Gamma(n+\theta)} \left(\prod_{j=1}^k \frac{\Gamma(n_j-\sigma)}{\Gamma(1-\sigma)}\, P_0(dZ^*_j)\right) \\
&\left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy)\right)^{\theta/\sigma} \left(\prod_{i=1}^n h(X_i, Z^*_{A_i})\right) \int_{\Delta^{n+1}} \prod_{j=1}^k G(Z^*_j; v_0, v^{(n)})^{\sigma-n_j} \\
&\qquad \left(\int_{\mathbb{X}} G(y; v_0, v^{(n)})^\sigma P_0(dy)\right)^{-k-\theta/\sigma} \underbrace{\frac{\Gamma(n+\theta)}{\Gamma(\theta)}\, v_0^{\theta-1}\, dv_0\, dv^{(n)}}_{\text{Dirichlet}(\theta,1,\ldots,1)}.
\end{aligned}$$

The expression above factorizes into the product of the marginal distribution induced by the Pitman-Yor prior for exchangeable data, that is

$$\sigma^k \frac{\Gamma(k+\theta/\sigma)}{\Gamma(\theta/\sigma)} \frac{\Gamma(\theta)}{\Gamma(n+\theta)} \left(\prod_{j=1}^k \frac{\Gamma(n_j-\sigma)}{\Gamma(1-\sigma)}\, P_0(dZ^*_j)\right),$$

and a term involving both predictors and responses, which is a mixture over the Dirichlet distribution with parameters $(1, \ldots, 1, \theta)$ on the $(n+1)$-dimensional unit simplex.

Prior predictive probability measure The marginal probability measure for the observation (X, Z) is

$$\begin{aligned}
\mathbb{P}(Z \in dz \mid X = x) &= \mathbb{E}\big[\tilde{p}(dz|x)\big] \\
&= h(x,z)\, P_0(dz) \left(\int_{\mathbb{X}} G_0(y)^{\sigma} P_0(dy) \right)^{\theta/\sigma} \int_0^1 (v\, G_0(z) + (1-v)\, h(x,z))^{\sigma-1} \\
&\qquad \left(\int_{\mathbb{X}} (v\, G_0(y) + (1-v)\, h(x,y))^{\sigma}\, P_0(dy) \right)^{-1-\theta/\sigma} \theta\, v^{\theta-1}\, dv.
\end{aligned}$$

Co-clustering probability The prior probability for two observations (X_1, Z_1) and (X_2, Z_2) to be in the same cluster is

$$\begin{aligned}
\mathbb{P}(Z_1 = Z_2 \mid X_1, X_2) &= \mathbb{E}\left[\int_{\mathbb{X}} \tilde{p}(dz|X_1)\, \tilde{p}(dz|X_2) \right] \\
&= \frac{1-\sigma}{1+\theta} \left(\int_{\mathbb{X}} G_0(y)^{\sigma} P_0(dy) \right)^{\theta/\sigma} \int_{\Delta^3} \int_{\mathbb{X}} h(X_1, z)\, h(X_2, z)\, G(z; v_0, v^{(2)})^{\sigma-2} P_0(dz) \\
&\qquad \left(\int_{\mathbb{X}} G(y; v_0, v^{(2)})^{\sigma} P_0(dy) \right)^{-1-\theta/\sigma} (1+\theta)\, \theta\, v_0^{\theta-1}\, dv_0\, dv^{(2)}.
\end{aligned}$$

Posterior distribution

Consider the vector of latent variables $(U_0, U_1, \dots, U_n)$ having density function with respect to the Lebesgue measure on $(\mathbb{R}^+)^{n+1}$ proportional to

$$f(u_0, u^{(n)}) = f(u_0, u^{(n)} \mid X^{(n)}, A^{(n)}, Z^*) \\ \propto u_0^{\theta-1} \exp\left\{ - \int_{\mathbb{X}} G(y; u_0, u^{(n)})^\sigma P_0(dy) \right\} \prod_{j=1}^k G(Z_j^*; u_0, u^{(n)})^{-n_j+\sigma}.$$

Considering the transformation of variables introduced before, one can alternatively consider the latent variables T, V_0 and $V^{(n)} = (V_1, \dots, V_n)$, with density function with respect to the Lebesgue measure on the product space $\mathbb{R}^+ \times \Delta^{n+1}$ proportional to

$$f(t, v_0, v^{(n)}) = f(t, v_0, v^{(n)} \mid X^{(n)}, A^{(n)}, Z^*) \\ \propto v_0^{\theta-1}\, t^{k\sigma+\theta-1} \exp\left\{ - t^\sigma \int_{\mathbb{X}} G(y; v_0, v^{(n)})^\sigma P_0(dy) \right\} \prod_{j=1}^k G(Z_j^*; v_0, v^{(n)})^{-n_j+\sigma}.$$

Note that this density can be factorized into the product $f(t, v_0, v^{(n)}) = f(t \mid v_0, v^{(n)})\, f(v_0, v^{(n)})$, where

$$f(v_0, v^{(n)}) = f(v_0, v^{(n)} \mid X^{(n)}, A^{(n)}, Z^*) \\ \propto v_0^{\theta-1} \left(\int_{\mathbb{X}} G(y; v_0, v^{(n)})^\sigma P_0(dy) \right)^{-k-\theta/\sigma} \prod_{j=1}^k G(Z_j^*; v_0, v^{(n)})^{\sigma-n_j},$$

and T^σ has Gamma distribution

$$T^\sigma \mid V_0, V^{(n)}, X^{(n)} \;\sim\; \text{Gamma}\left(k + \frac{\theta}{\sigma}, \int_{\mathbb{X}} G(y; V_0, V^{(n)})^\sigma P_0(dy) \right).$$

The posterior distribution of $\tilde{\mu}$ given responses $Z^{(n)}$, encoded into the labels $A^{(n)}$ and atom Z^*, predictors $X^{(n)}$ and latent variables T, V_0 and $V^{(n)}$ is

$$\tilde{\mu}(dz) \mid X^{(n)}, A^{(n)}, Z^*, T, V_0, V^{(n)} \\ \sim\; \tilde{\mu}^*(dz \mid X^{(n)}, T, V_0, V^{(n)}) + \sum_{j=1}^k J_j(n_j, X^{(n)}, T, V_0, V^{(n)}, Z_j^*)\, \delta_{Z_j^*}(dx),$$

where $\tilde{\mu}^*$ is a generalized gamma completely random measure, with Lévy intensity measure

$$\nu^*(ds, dz \mid X^{(n)}, T, V_0, V^{(n)}) = \frac{\sigma}{\Gamma(1-\sigma)}\, s^{-1-\sigma} \exp\left\{ -T\, G(z; V_0, V^{(n)}) s \right\} ds\, P_0(dz),$$

and $\{ J_j \}_{j=1}^k$ are independent random variables having Gamma distribution

$$J_j \mid n_j, X^{(n)}, T, V_0, V^{(n)}, Z_j^* \;\sim\; \text{Gamma}\left(n_j - \sigma,\; T\, G(Z_j^*; V_0, V^{(n)}) \right).$$

Note that, according to the change of variables above, it holds that

$$T\, G(z; V_0, V^{(n)}) = G(z; U_0, U^{(n)}).$$

Continuous component Given predictors $X^{(n)}$ and latent variables V_0 and $V^{(n)}$, but marginalizing with respect to T, the random measure $\tilde{\mu}^*$ has Laplace transform

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}} f(z)\,\tilde{\mu}^*(dz)\right\} \,\Big|\, X^{(n)}, Z^{(n)}, V_0, V^{(n)}\right] \\
&\qquad = \frac{\sigma\,\Gamma(\theta+k\sigma)}{\Gamma(\theta/\sigma+k)} \left(\int_{\mathbb{X}} G(z;V_0,V^{(n)})^{\sigma} P_0(dz)\right)^{k+\theta/\sigma} \\
&\qquad\qquad \mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}} f(z)\,\tilde{\mu}_\sigma(dz)\right\}\left(\int_{\mathbb{X}} G(z;V_0,V^{(n)})\tilde{\mu}_\sigma(dz)\right)^{-k\sigma-\theta}\right].
\end{aligned}
$$

Therefore, the probability distribution of $\tilde{\mu}^*$, given predictors $X^{(n)}$ and latent variables V_0 and $V^{(n)}$, has Radon-Nikodym derivative with respect to a σ-stable CRM $\tilde{\mu}_\sigma$ given by

$$
\frac{d\mathcal{L}(\tilde{\mu}^*)}{d\mathcal{L}(\tilde{\mu}_\sigma)}(m) = \frac{\sigma\,\Gamma(\theta+k\sigma)}{\Gamma(\theta/\sigma+k)}\,\frac{\left(\int_{\mathbb{X}} G(z;V_0,V^{(n)})^{\sigma} P_0(dz)\right)^{\theta/\sigma+k}}{\left(\int_{\mathbb{X}} G(z;V_0,V^{(n)}) m(dz)\right)^{\theta+k\sigma}}
$$

which is analogous to the Radon-Nikodym derivative defining the prior random measure $\tilde{\mu}$, with updated parameter $\theta^* = \theta + k\sigma$ and updated function

$$
G^*(z) = G(z;V_0,V^{(n)}) = V_0\,G_0(z) + V^{(n)} h(X^{(n)}, z).
$$

Dirichlet distribution Consider the following random variables,

$$
Z_0 = \int_{\mathbb{X}} G(z;V_0,V^{(n)})\,\tilde{\mu}^*(dz), \qquad Z_j = G(Z_j^*;V_0,V^{(n)})\,J_j, \quad j=1,\dots,k,
$$

and their normalization, obtained dividing by their sum, that is,

$$
W_0 = \frac{Z_0}{Z_0 + \sum_{h=1}^{k} Z_h} = \frac{\int_{\mathbb{X}} G(z;V_0,V^{(n)})\,\tilde{\mu}^*(dz)}{\int_{\mathbb{X}} G(y;V_0,V^{(n)})\,\tilde{\mu}^*(dy) + \sum_{h=1}^{k} G(Z_h^*;V_0,V^{(n)})\,J_h},
$$

$$
W_j = \frac{Z_j}{Z_0 + \sum_{h=1}^{k} Z_h} = \frac{G(Z_j^*;V_0,V^{(n)})\,J_j}{\int_{\mathbb{X}} G(y;V_0,V^{(n)})\,\tilde{\mu}^*(dy) + \sum_{h=1}^{k} G(Z_h^*;V_0,V^{(n)})\,J_h}, \qquad j=1,\dots,k.
$$

Given predictors $X^{(n)}$ and latent variables V_0 and $V^{(n)}$, but marginalizing with respect to T, the random vector $W = (W_0, W_1, \dots, W_k)$ has Dirichlet distribution on the $(k+1)$-dimensional unit simplex,

$$
W = (W_0, W_1, \dots, W_k) \ \sim\ \text{Dirichlet}\left(\theta + k\sigma, n_1 - \sigma, \dots, n_k - \sigma\right).
$$

Posterior random probability measure For some predictor $x \in \mathbb{X}$, consider the posterior random probability measure,

$$\tilde{p}^*_x(dz) = \tilde{p}^*(dz|x) := \frac{h(x,z)\,\tilde{\mu}^*(dz) + \sum_{j=1}^{k} h(x, Z^*_j)\, J_j\, \delta_{Z^*_j}(dz)}{\int_{\mathbb{X}} h(x,y)\,\tilde{\mu}^*(dy) + \sum_{h=1}^{k} h(x, Z^*_h)\, J_h}.$$

Specifically, consider the following random variables,

$$S_0 = \int_{\mathbb{X}} h(x,z)\,\tilde{\mu}^*(dz), \qquad S_j = h(x, Z^*_j)\, J_j, \quad j = 1, \ldots, k,$$

and their normalization, obtained dividing by their sum, that is,

$$W_0 = \frac{S_0}{S_0 + \sum_{h=1}^{k} S_h} = \frac{\int_{\mathbb{X}} h(x,z)\,\tilde{\mu}^*(dz)}{\int_{\mathbb{X}} h(x,y)\,\tilde{\mu}^*(dy) + \sum_{h=1}^{k} h(x, Z^*_h)\, J_h} = \tilde{p}^*_x(\mathbb{X} \setminus \{\, Z^*_1, \ldots, Z^*_k \,\}),$$

$$W_j = \frac{S_j}{S_0 + \sum_{h=1}^{k} S_h} = \frac{h(x, Z^*_j)\, J_j}{\int_{\mathbb{X}} h(x,y)\,\tilde{\mu}^*(dy) + \sum_{h=1}^{k} h(x, Z^*_h)\, J_h} = \tilde{p}^*_x(Z^*_j), \qquad j = 1, \ldots, k.$$

Given predictors $X^{(n)}$ and latent variables V_0 and $V^{(n)}$, but marginalizing with respect to T, the random vector $W = (W_0, W_1, \ldots, W_k)$ has density function

$$\begin{aligned} f(w_0, w_1, \ldots, w_k) = \sigma\, \frac{\Gamma(\theta + k\sigma)}{\Gamma(k + \theta/\sigma)} \left(\int_{\mathbb{X}} G(y; V_0, V^{(n)})^{\sigma} P_0(dy) \right)^{k+\theta/\sigma} \\ \mathbb{E}\Bigg[\left(w_0 \frac{\int_{\mathbb{X}} G(y; V_0, V^{(n)}) \tilde{\mu}_\sigma(dy)}{\int_{\mathbb{X}} h(x,y) \tilde{\mu}_\sigma(dy)} + \sum_{j=1}^{k} w_j \frac{G(Z^*_j; V_0, V^{(n)})}{h(x, Z^*_j)} \right)^{-n-\theta} \\ \left(\int_{\mathbb{X}} h(x,y) \tilde{\mu}_\sigma(dy) \right)^{-\theta-k\sigma} \Bigg] f_{\mathrm{Dir}}(w_0, w_1, \ldots, w_k), \end{aligned}$$

where $\tilde{\mu}_\sigma$ is a σ-stable CRM and $f_{\mathrm{Dir}}(w_0, w_1, \ldots, w_k)$ is the density function of the Dirichlet distribution on the $(k+1)$-dimensional unit simplex with parameters $(\theta+k\sigma, n_1-\sigma, \ldots, n_k-\sigma)$, that is

$$f_{\mathrm{Dir}}(w_0, w_1, \ldots, w_k) = \Gamma(n+\theta) \left(\frac{w_0^{\theta+k\sigma-1}}{\Gamma(\theta+k\sigma)} \prod_{j=1}^{k} \frac{w_j^{n_j-\sigma-1}}{\Gamma(n_j - \sigma)} \right).$$

Marginal sampler

Consider the joint marginal distribution of the responses $Z^{(n)}$, encoded into the labels $A^{(n)}$ and atoms Z^*, and the latent variables V_0 and $V^{(n)}$, supported on the unit simplex, that is

$$\mathbb{P}(A^{(n)}, Z^*, V_0, V^{(n)} \mid X^{(n)}) = \sigma^k \frac{\Gamma(k+\theta/\sigma)}{\Gamma(\theta/\sigma)} \frac{\Gamma(\theta)}{\Gamma(n+\theta)} \left(\prod_{j=1}^{k} \frac{\Gamma(n_j - \sigma)}{\Gamma(1-\sigma)} P_0(dZ_j^*) \right)$$
$$\left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy) \right)^{\theta/\sigma} \left(\prod_{i=1}^{n} h(X_i, Z_{A_i}^*) \right) \prod_{j=1}^{k} G(Z_j^*; V_0, V^{(n)})^{\sigma - n_j}$$
$$\left(\int_{\mathbb{X}} G(y; V_0, V^{(n)})^\sigma P_0(dy) \right)^{-k-\theta/\sigma} \frac{\Gamma(n+\theta)}{\Gamma(\theta)} V_0^{\theta-1}.$$

At each iteration of the marginal Gibbs sampler, the following steps are performed.

1. Sample the latent variables V_0 and $V^{(n)}$ from their joint full conditional distribution on the unit simplex, that is, proportionally to

$$V_0^{\theta-1} \left(\int_{\mathbb{X}} G(y; V_0, V^{(n)})^\sigma P_0(dy) \right)^{-k-\theta/\sigma} \prod_{j=1}^{k} G(Z_j^*; V_0, V^{(n)})^{\sigma - n_j}.$$

   Sampling from this distribution is performed using a Metropolis-Hastings strategy.

2. For each $i = 1, \ldots, n$, sample the label A_i proportionally to the discrete distribution

$$\mathbb{P}(A_i = j) \propto (n_j - \sigma) \frac{h(X_i, Z_j^*)}{G(Z_j^*; V_0, V^{(n)})}, \qquad j = 1, \ldots, k,$$
$$\mathbb{P}(A_i = k+1) \propto (k\sigma + \theta) \int_{\mathbb{X}} \frac{h(X_i, z)}{G(z; V_0, V^{(n)})} Q_0(dz),$$

   where Q_0 is the diffuse probability distribution proportional to

$$Q_0(dz) \propto G(z; V_0, V^{(n)})^\sigma P_0(dz).$$

   In case $A_i = k+1$, sample the new atom Z_{k+1}^* proportionally to

$$\frac{h(X_i, z)}{G(z; V_0, V^{(n)})} Q_0(dz).$$

3. For each $j = 1, \ldots, k$, sample the atom Z_j^* proportionally to

$$\mathbb{P}(Z_j^* \in dz) \propto \left(\prod_{i:\ A_i = j} h(X_i, z) \right) G(z; V_0, V^{(n)})^{-n_j} Q_0(dz).$$

Conditional truncated sampler

Consider the likelihood function introduced at the beginning,

$$\begin{aligned}\mathbb{P}(Z^{(n)} \mid X^{(n)}, \tilde{\mu}) &= \mathbb{P}(A^{(n)}, Z^* \mid X^{(n)}, \tilde{\mu}) \\ &= \left(\prod_{i=1}^{n} h(X_i, Z^*_{A_i})\, \tilde{\mu}(dZ^*_{A_i})\right) \int_{(\mathbb{R}^+)^n} \exp\left\{-\int_{\mathbb{X}} u^{(n)} h(X^{(n)}, y)\, \tilde{\mu}(dy)\right\} du^{(n)}.\end{aligned}$$

The integration above can be considered a marginalization with respect to a vector of additional non-negative latent variables $U^{(n)} = (U_1, \ldots, U_n)$, so that the augmented likelihood is

$$\begin{aligned}&\mathbb{P}(A^{(n)}, Z^*, U^{(n)} \mid X^{(n)}, \tilde{\mu}) \\ &\qquad = \left(\prod_{i=1}^{n} h(X_i, Z^*_{A_i})\, \tilde{\mu}(dZ^*_{A_i})\right) \exp\left\{-\int_{\mathbb{X}} U^{(n)} h(X^{(n)}, y)\, \tilde{\mu}(dy)\right\}.\end{aligned}$$

The probability distribution of the random measure $\tilde{\mu}$ is absolutely continuous with respect to the probability distribution of a σ-stable CRM $\tilde{\mu}_\sigma$, that is

$$\frac{d\mathcal{L}(\tilde{\mu})}{d\mathcal{L}(\tilde{\mu}_\sigma)} = \frac{\sigma}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(z)^\sigma P_0(dz)\right)^{\theta/\sigma} \int_{\mathbb{R}^+} u_0^{\theta-1} \exp\left\{-\int_{\mathbb{X}} u_0\, G_0(z)\, \tilde{\mu}_\sigma(dz)\right\} du_0.$$

The random measure $\tilde{\mu}_\sigma$ has almost-surely discrete realization, and can be represented as a sequence of atom $Z^* = (Z^*_h)_{h\geq 1}$ and jumps $J = (J_h)_{h\geq 1}$, with the correspondence

$$\tilde{\mu}_\sigma(dz) = \sum_{h\geq 1} J_h\, \delta_{Z^*_h}(dz).$$

In this sampling algorithm, consider a truncation level H for the infinite series above, so that $\tilde{\mu}_\sigma$ is replaced by the truncated measure

$$\tilde{\mu}^T_\sigma(dz) = \sum_{h=1}^{H} J_h\, \delta_{Z^*_h}(dz).$$

For convenience of notation, assume that the k distinct values $Z^*_1, \ldots, Z^*_k$ assumed by the responses $Z^{(n)}$ correspond to the first k labels in the ordering, so that $1 \leq A_i \leq k < H$.

Therefore, introducing a further non-negative latent variable U_0, the joint probability distribution of the random quantities involved, namely

- the responses $Z^{(n)}$, encoded into the labels $A^{(n)}$;
- the non-negative latent variables U_0 and $U^{(n)}$;
- the random truncated measure $\tilde{\mu}^T_\sigma$, encoded into the truncated sequences of atoms Z^* and jumps J;

takes the final form

$$\begin{aligned}&\frac{\sigma}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy)\right)^{\theta/\sigma} \left(\prod_{i=1}^{n} h(X_i, Z^*_{A_i})\, J_{A_i}\right) \prod_{h=1}^{H} P_0(dZ^*_h) \\ &\qquad\qquad U_0^{\theta-1} \exp\left\{-\sum_{h=1}^{H} G(Z^*_h; U_0, U^{(n)})\, J_h\right\} d\mathcal{L}(J),\end{aligned}$$

where $\mathcal{L}(J)$ is the probability distribution of the truncated sequence of jumps, that is $(J_1, \ldots, J_H)$. This sequence is sampled according to the procedure described hereunder for the initialization of the sampler.

The conditional truncated sampling algorithm is initialized as follows.

1. Sample H atoms $Z_1^*, \ldots, Z_H^*$ independently from the diffuse probability measure P_0.
2. For each $h = 1, \ldots, H$, sample the random jump J_h according to the following procedure:
   (a) sample $W_h \sim \text{Exp}(1)$ and set $S_h = S_{h-1} + \Gamma(1-\sigma)\, W_h$, with $S_0 = 0$;
   (b) define J_h as the value solving the equation
   $$S_h = \Gamma(1-\sigma) \int_{\mathbb{X}} \int_{J_h}^{\infty} \nu(ds, dz) = J_h^{-\sigma},$$
   that is $J_h = (S_h)^{-1/\sigma}$.

An alternative to the last step is the following.

2bis. (a) Sample the random variable S_H from the Gamma distribution
$$S_H \sim \text{Gamma}\left(H, \frac{1}{\Gamma(1-\sigma)}\right);$$
(b) for each $h = 1, \ldots, H-1$, sample the random jump S_h from the uniform distribution,
$$S_h \sim \text{Unif}\,[0, S_H]$$
(c) for each $h = 1, \ldots, H$, define $J_h = (S_h)^{-1/\sigma}$.

At each iteration of the Gibbs sampler, the following steps are performed.

1. Sample the non-negative latent variable U_0 from its full conditional distribution, that is,
$$U_0 \sim \text{Gamma}\left(\theta, \sum_{h=1}^{H} G_0(Z_h^*)\, J_h\right).$$
2. For each $i = 1, \ldots, n$, sample the non-negative latent variables U_i from its full conditional distribution, that is,
$$U_i \sim \text{Exp}\left(\sum_{h=1}^{H} h(X_i, Z_h^*)\, J_h\right).$$
3. For each $j = 1, \ldots, k$, sample the atom Z_j^* proportionally to
$$\mathbb{P}(Z_j^* \in dz) \propto \left(\prod_{i:\, A_i = j} h(X_i, z)\right) \exp\left\{-G(z; U_0, U^{(n)})\, J_j\right\} P_0(dz).$$

4. For each $j = 1, \dots, k$, sample the jump J_j from the Gamma distribution

$$J_j \sim \text{Gamma}\left(n_j - \sigma,\, G(Z_j^*; U_0, U^{(n)})\right)$$

5. For each $h = k+1, \dots, H$, sample the atom Z_h^* from the probability distribution P_0 and sample the random jump J_h according to the following procedure:

   (a) sample $W_h \sim \text{Exp}(1)$ and set $S_h = S_{h-1} + W_h$, with $S_k = 0$;

   (b) define J_h as the value solving the equation

   $$S_h = \int_{J_h}^{\infty} \rho^*(ds \mid Z_h^*) = G(Z_h^*; U_0, U^{(n)})\, \Gamma\left(-\sigma, J_h\, G(Z_h^*; U_0, U^{(n)})\right),$$

   where Γ is the incomplete gamma function, that is

   $$\Gamma\left(-\sigma, J_h\, G(Z_h^*; U_0, U^{(n)})\right) = \frac{S_h}{G(Z_h^*; U_0, U^{(n)})}.$$

6. For each $i = 1, \dots, n$, sample the label A_i from the corresponding discrete distribution, that is, proportionally to

   $$\mathbb{P}(A_i = h) \propto h(X_i, Z_h^*)\, J_h, \qquad h = 1, \dots, H;$$

   for convenience, reorder the labels so that observations are assigned to the first k labels in the ordering, so that $1 \le A_i \le k$.

Conditional slice sampler

Consider the augmented likelihood function introduced before,

$$\mathbb{P}(A^{(n)}, Z^*, U^{(n)} \mid X^{(n)}, \tilde{\mu}) \\ = \left(\prod_{i=1}^{n} h(X_i, Z^*_{A_i})\, \tilde{\mu}(dZ^*_{A_i}) \right) \exp\left\{ - \int_{\mathbb{X}} U^{(n)} h(X^{(n)}, y)\, \tilde{\mu}(dy) \right\}.$$

Introduce a vector of auxiliary non-negative slice variables $S^{(n)} = (S_1, \ldots, S_n)$, so that the augmented likelihood is expressed as

$$\mathbb{P}(A^{(n)}, Z^*, U^{(n)}, S^{(n)} \mid X^{(n)}, \tilde{\mu}) = \exp\left\{ - \int_{\mathbb{X}} U^{(n)} h(X^{(n)}, y)\, \tilde{\mu}(dy) \right\} \\ \left(\prod_{i=1}^{n} \big(S_i < G(Z^*_{A_i}; V_0, V^{(n)})\, \tilde{\mu}(dZ^*_{A_i}) \big)\, \frac{h(X_i, Z^*_{A_i})}{G(Z^*_{A_i}; V_0, V^{(n)})} \right);$$

note that the marginalization with respect to S_i, for each $i = 1, \ldots, n$, reads

$$\int_{\mathbb{R}^+} \big(s_i < G(Z^*_{A_i}; V_0, V^{(n)})\, \tilde{\mu}(dZ^*_{A_i}) \big)\, \frac{h(X_i, Z^*_{A_i})}{G(Z^*_{A_i}; V_0, V^{(n)})}\, ds_i = h(X_i, Z^*_{A_i})\, \tilde{\mu}(dZ^*_{A_i}).$$

The probability distribution of the random measure $\tilde{\mu}$ is absolutely continuous with respect to the probability distribution of a σ-stable CRM $\tilde{\mu}_\sigma$, that is

$$\frac{d\mathcal{L}(\tilde{\mu})}{d\mathcal{L}(\tilde{\mu}_\sigma)} = \frac{\sigma}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(z)^\sigma P_0(dz) \right)^{\theta/\sigma} \int_{\mathbb{R}^+} u_0^{\theta-1} \exp\left\{ - \int_{\mathbb{X}} u_0\, G_0(z)\, \tilde{\mu}_\sigma(dz) \right\} du_0.$$

For a threshold function $L\colon z \mapsto L(z)$, the σ-stable CRM can be decomposed into the sum of two independent completely random measures

$$\tilde{\mu}_\sigma(dz) = \tilde{\mu}^+_\sigma(dz) + \tilde{\mu}^-_\sigma(dz),$$

having Lévy intensity measures, respectively,

$$\nu^+(ds, dz) = \nu(ds, dz)(s > L(z)), \qquad \nu^-(ds, dz) = \nu(ds, dz)(0 \le s \le L(z)).$$

The random measure $\tilde{\mu}^+_\sigma$ has almost-surely discrete realizations, and can be represented as a sequence of atom $Z^* = (Z^*_1, \ldots, Z^*_H)$ and jumps $J = (J_1, \ldots, J_H)$, with the correspondence

$$\tilde{\mu}^+_\sigma(dz) = \sum_{h=1}^{H} J_h\, \delta_{Z^*_h}(dz).$$

The random number H of atoms and jumps is an almost-surely finite random variable, having Poisson distribution

$$H \;\sim\; \text{Poisson}\left(\int_{\mathbb{X}} \int_{L(z)}^{\infty} \nu(ds, dz) \right) = \text{Poisson}\left(\int_{\mathbb{X}} \frac{L(z)^{-\sigma}}{\Gamma(1-\sigma)} P_0(dz) \right);$$

the atoms $(Z^*_1, \ldots, Z^*_H)$ are then sampled independently from the diffuse probability measure P_0, while the jumps $(J_1, \ldots, J_H)$ are sampled independently from the probability distribution

$$f_{J_h}(s) = \sigma L(Z^*_h)^\sigma\, s^{-1-\sigma}\, (s > L(Z^*_h)), \qquad h = 1, \ldots, H.$$

Notice that, setting $L < \min_i S_i$, the jumps $\tilde{\mu}_\sigma(dZ^*_{A_1}), \dots, \tilde{\mu}_\sigma(dZ^*_{A_n})$ are such that

$$G(Z^*_{A_i}; V_0, V^{(n)})\, \tilde{\mu}_\sigma(dZ^*_{A_i}) > L, \qquad i = 1, \dots, n,$$

and thus are included in $\tilde{\mu}^+_\sigma$ for the threshold function

$$L(z) = \frac{L}{G(z; V_0, V^{(n)})} = \frac{\min_i S_i}{G(z; V_0, V^{(n)})}.$$

For convenience of notation, assume that the k distinct values $Z^*_1, \dots, Z^*_k$ assumed by the responses $Z^{(n)}$ correspond to the first k labels in the ordering, so that $1 \le A_i \le k \le H$.

Therefore, introducing a further non-negative latent variable U_0, the joint probability distribution of the random quantities involved, namely

- the responses $Z^{(n)}$, encoded into the labels $A^{(n)}$;
- the non-negative latent variables U_0 and $U^{(n)}$;
- the auxiliary slice variables $S^{(n)}$;
- the random measure $\tilde{\mu}^+_\sigma$, encoded into the sequences of atoms Z^* and jumps J;
- the random measure $\tilde{\mu}^-_\sigma$;

takes the final form

$$\frac{\sigma}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy) \right)^{\theta/\sigma} \left(\prod_{i=1}^n \left(S_i < G(Z^*_{A_i}; V_0, V^{(n)})\, J_{A_i}\right) \frac{h(X_i, Z^*_{A_i})}{G(Z^*_{A_i}; V_0, V^{(n)})} \right)$$
$$U_0^{\theta-1} \exp\left\{ -\sum_{h=1}^H G(Z^*_h; U_0, U^{(n)})\, J_h \right\} \left(\prod_{h=1}^H P_0(dZ^*_h) \right) d\mathcal{L}(J_1, \dots, J_H)$$
$$\exp\left\{ -\int_{\mathbb{X}} G(y; U_0, U^{(n)}) \tilde{\mu}^-_\sigma(dy) \right\} d\mathcal{L}(\tilde{\mu}^-_\sigma).$$

The last step towards the description of a conditional slice sampling algorithm is the marginalization with respect to the random measure $\tilde{\mu}^-_\sigma$, which leads to

$$\frac{\sigma}{\Gamma(\theta/\sigma)} \left(\int_{\mathbb{X}} G_0(y)^\sigma P_0(dy) \right)^{\theta/\sigma} \left(\prod_{i=1}^n \left(S_i < G(Z^*_{A_i}; V_0, V^{(n)})\, J_{A_i}\right) \frac{h(X_i, Z^*_{A_i})}{G(Z^*_{A_i}; V_0, V^{(n)})} \right)$$
$$U_0^{\theta-1} \exp\left\{ -\sum_{h=1}^H G(Z^*_h; U_0, U^{(n)})\, J_h \right\} \left(\prod_{h=1}^H P_0(dZ^*_h) \right) d\mathcal{L}(J_1, \dots, J_H)$$
$$\exp\left\{ -\int_{\mathbb{X}} \int_{\mathbb{R}^+} \left(1 - e^{-G(y; U_0, U^{(n)})\, s}\right) \nu^-(ds, dy) \right\}.$$

Considering the threshold function $L(z) = L\, G(z; V_0, V^{(n)})^{-1}$, the double integral can be rewritten as

$$\begin{aligned}
&\int_{\mathbb{X}} \int_{\mathbb{R}^+} \left(1 - e^{-G(y; U_0, U^{(n)})\, s}\right) \nu^-(ds, dy) \\
&\qquad = \frac{\sigma}{\Gamma(1-\sigma)} \int_{\mathbb{X}} \int_0^{L(y)} \left(1 - e^{-G(y; U_0, U^{(n)})\, s}\right) s^{-\sigma-1}\, ds\, P_0(dy) \\
&\qquad\qquad = \gamma(TL) \int_{\mathbb{X}} G(y; U_0, U^{(n)})^\sigma P_0(dy),
\end{aligned}$$

where Γ_N is the normalized lower incomplete gamma function and

$$\gamma(t) = \Gamma_N(1-\sigma, t) - \frac{1-e^{-t}}{\Gamma(1-\sigma)\, t^{\sigma}}, \qquad t \geq 0;$$

note that γ is a monotonically increasing function such that

$$\gamma(t) \in [0,1], \qquad \gamma(0) = 0, \qquad \lim_{t\to\infty} \gamma(t) = 1.$$

The conditional slice sampling algorithm is initialized as follows.

1. Set a truncation level $L > 0$.
2. Sample H from the Poisson distribution

$$H \sim \text{Poisson}\left(\frac{L^{-\sigma}}{\Gamma(1-\sigma)}\right);$$

3. Sample H atoms $Z_1^*, \ldots, Z_H^*$ independently from the diffuse probability measure P_0.
4. Sample H jumps $J_1, \ldots, J_H$ independently from the probability distribution

$$f_{J_h}(s) = \sigma L^{\sigma}\, s^{-1-\sigma}\,(s > L), \qquad h = 1, \ldots, H,$$

that is, for each $h = 1, \ldots, H$, sample $W_h \sim \text{Unif}\,[0,1]$ and set $J_h = L\, W_h^{-1/\sigma}$.

5. Sample the labels $A^{(n)}$ from the discrete uniform distribution over the set $\{1, \ldots, H\}$,

$$A_i \sim \text{Unif}\,\{1, \ldots, H\}, \qquad i = 1, \ldots, n;$$

for convenience, reorder the labels so that observations are assigned to the first k labels in the ordering, so that $1 \leq A_i \leq k \leq H$.

At each iteration of the Gibbs sampler, the following steps are performed.

1. Sample the non-negative latent variables U_0 and $U^{(n)}$ from their joint full conditional distribution, that is, proportionally to

$$U_0^{\theta-1} \exp\left\{-\sum_{h=1}^{H} G(Z_h^*; U_0, U^{(n)})\, J_h\right\} \exp\left\{-\gamma(TL) \int_{\mathbb{X}} G(y; U_0, U^{(n)})^{\sigma} P_0(dy)\right\}.$$

Sampling from this distribution may be performed via rejection sampling:

(a) propose values for U_0 and $U^{(n)}$ from the independent distributions

$$U_0 \sim \text{Gamma}\left(\theta,\ \sum_{h=1}^{H} G_0(Z_h^*)\, J_h\right), \qquad U_i \sim \text{Exp}\left(\sum_{h=1}^{H} h(X_i, Z_h^*)\, J_h\right);$$

(b) accept the proposed values with probability

$$\exp\left\{-\gamma(TL) \int_{\mathbb{X}} G(y; U_0, U^{(n)})^{\sigma} P_0(dy)\right\}.$$

2. For each $j = 1, \ldots, k$, sample the atom Z_j^* proportionally to

$$\mathbb{P}(Z_j^* \in dz) \propto \left(\prod_{i:\ A_i = j} h(X_i, z) \right) \exp\left\{ -G(z; U_0, U^{(n)})\, J_j \right\} P_0(dz).$$

3. For each $j = 1, \ldots, k$, sample the jump J_j from the Gamma distribution

$$J_j \sim \text{Gamma}\left(n_j - \sigma,\ G(Z_j^*; U_0, U^{(n)})\right)$$

4. Sample the auxiliary slice variables $S^{(n)}$ from the uniform distribution

$$S_i \sim \text{Unif}\,[0,\ G(Z_{A_i}^*; V_0, V^{(n)})\, J_{A_i}], \qquad i = 1, \ldots, n,$$

   and set $L < \min_i S_i$.

5. Sample k' from the Poisson distribution

$$k' \sim \text{Poisson}\left(\frac{\sigma\, \Gamma(-\sigma, TL)}{\Gamma(1-\sigma)}\ T^\sigma \int_{\mathbb{X}} G(z; V_0, V^{(n)})^\sigma P_0(dz) \right),$$

   where Γ is the upper incomplete gamma function, and set $H = k + k'$.

6. For each $h = k+1, \ldots, H$, sample the atom Z_h^* from the probability distribution P_0 and sample the corresponding jump J_h according to the following procedure:

   (a) sample $W_h \sim \text{Unif}[0, 1]$;

   (b) define J_h as the value solving the equation

$$\Gamma(-\sigma, T\, G(Z_h^*; V_0, V^{(n)})\, J_h) = W_h\, \Gamma(-\sigma, T\, L).$$

7. For each $i = 1, \ldots, n$, sample the label A_i from the corresponding discrete distribution, that is, proportionally to

$$\mathbb{P}(A_i = h) \propto \frac{h(X_i, Z_{A_i}^*)}{G(Z_{A_i}^*; V_0, V^{(n)})} \left(S_i < G(Z_{A_i}^*; V_0, V^{(n)})\, J_{A_i} \right);$$

   for convenience, reorder the labels so that observations are assigned to the first k labels in the ordering, so that $1 \leq A_i \leq k \leq H$.

Density regression framework

The measure-valued stochastic process $\tilde{p}$ defined above can be used within a nonparametric density regression model. Specifically, the probability distribution of each observation Y_i, given the corresponding predictors X_i, for $i = 1, \dots, n$, is a countable mixture of parametric regression models, with weights given by the stochastic process $\tilde{p}$,

$$\mathbb{P}(Y_i \in dy \mid X_i) = \sum_{h \geq 1} F(X_i; \phi_h)(dy)\, \tilde{p}_{X_i}(dZ_h^*),$$

where F is a parametric regression model, parameterized by ϕ_h. The fundamental example is the standard linear regression model, that is

$$F(x; \phi = (\beta, \tau))(dy) = k(y \mid \beta^T x, \tau),$$

where $k(\cdot \mid \mu, \tau)$ is the density of a Gaussian random variable with mean μ and standard deviation τ. Moreover, consider independent prior distributions for each parameter $\phi_h = (\beta_h, \tau_h)$, for $h \geq 1$, that is,

$$\phi_h = (\beta_h, \tau_h) \overset{\text{i.i.d}}{\sim} \pi_\phi(d\beta_h, d\tau_h) = \pi_\beta(d\beta_h)\, \pi_\tau(d\tau_h), \qquad h \geq 1.$$

The only differences in the sampling algorithms described above are in the labels sampling steps.

Marginal sampler

2. For each $i = 1, \dots, n$, sample the label A_i proportionally to the discrete distribution

$$\mathbb{P}(A_i = j) \propto (n_j - \sigma)\, \frac{k(Y_i \mid \beta_j^T X_i, \tau_j)\, h(X_i, Z_j^*)}{G(Z_j^*; V_0, V^{(n)})}, \qquad j = 1, \dots, k,$$

$$\mathbb{P}(A_i = k+1) \propto (k\sigma + \theta) \int_{\mathbb{X}} \frac{h(X_i, z)}{G(z; V_0, V^{(n)})}\, Q_0(dz) \int k(Y_i \mid \beta^T X_i, \tau)\, \pi_\beta(d\beta)\, \pi_\tau(d\tau),$$

   where Q_0 is the diffuse probability distribution proportional to

$$Q_0(dz) \propto G(z; V_0, V^{(n)})^\sigma P_0(dz).$$

   In case $A_i = k+1$, sample the new atom Z_{k+1}^* proportionally to

$$\frac{k(Y_i \mid \beta_h^T X_i, \tau_h)\, h(X_i, z)}{G(z; V_0, V^{(n)})}\, Q_0(dz),$$

   and the new parameters $\phi_{k+1} = (\beta_{k+1}, \tau_{k+1})$ proportionally to

$$k(Y_i \mid \beta^T X_i, \tau)\, \pi_\beta(d\beta)\, \pi_\tau(d\tau).$$

Conditional truncated sampler

7. For each $i = 1, \dots, n$, sample the label A_i from the corresponding discrete distribution, that is, proportionally to

$$\mathbb{P}(A_i = h) \propto k(Y_i \mid \beta_h^T X_i, \tau_h)\, h(X_i, Z_h^*)\, J_h, \qquad h = 1, \dots, H.$$

Conditional slice sampler

7. For each $i = 1, \ldots, n$, sample the label A_i from the corresponding discrete distribution, that is, proportionally to

$$\mathbb{P}(A_i = h) \propto k(Y_i \mid \beta_h^T X_i, \tau_h) \, \frac{h(X_i, Z^*_{A_i})}{G(Z^*_{A_i}; V_0, V^{(n)})} \left(S_i < G(Z^*_{A_i}; V_0, V^{(n)}) \, J_{A_i}\right).$$

Resampling parameters An additional step for resampling the regression parameters is added at the end of each Gibbs sampling algorithm:

(). For each $j = 1, \ldots, k$, sample the parameter $\phi_j = (\beta_j, \tau_j)$ proportionally to

$$\mathbb{P}(\beta_j \in d\beta, \tau_j \in d\tau) \propto \left(\prod_{i:\, A_i = j} k(Y_i \mid \beta^T X_i, \tau) \right) \pi_\beta(d\beta) \, \pi_\tau(d\tau).$$