Posterior Sampling Steps

1 DP-SPGLM

Final hierarchical model —

$$
\begin{aligned}
y_i \mid z_i, x_i &\sim K(y_i \mid z_i); \\
z_i \mid x_i = x, \theta_x, \widetilde{\mu} &\sim p_x(z_i) \propto \exp(\theta_x z_i)\widetilde{\mu}(z_i) \\
\theta_x \mid \beta, \widetilde{\mu} &\sim p(\theta_x \mid \widetilde{\theta}_x), \text{ with } b'(\widetilde{\theta}_x) = g^{-1}(x'\beta) = \lambda(x) \\
\widetilde{\mu} &\sim \text{ gamma CRM}(\nu), \text{ with } \nu(ds, dz) = \frac{e^{-s}}{s} ds \cdot \alpha G_0(dz) \\
\beta &\sim \text{ MVN}(\mu_\beta, \Sigma_\beta)
\end{aligned}
$$

where $g_i(z) = \exp(\theta_i z)$, $\frac{\widetilde{\mu}}{T} \sim DP(\alpha, G_0)$, $K(\cdot \mid z_i) = N(\cdot \mid z_i, \sigma_0^2)$ (or some other distribution). Here, if we assume θ_i does not depend on $\widetilde{\mu}$, we have a (n?) non-homogeneous CRM(s?) case with n (deterministic) perturbing functions $g_i, i = 1, \ldots, n$.
Question: Should we make it more general and use normalized generalized gamma (NGG) as a prior?

1.1 Posterior sampling steps

Parameters, $(\beta, \widetilde{\mu})$ and latent variables, $\{z_i\}_{i=1}^n, \{u_i\}_{i=1}^n$.
Choices to make for $\alpha, G_0, K, \sigma_0, \mu_\beta, \Sigma_\beta$. We can also put a prior on σ_0.

1.1.1 β (and θ) update —

$[\widetilde{\beta} \mid \widetilde{\mu}^{(t)}, —] \sim N_p(\mu_\beta^\star, \Sigma_\beta^\star) 1_A\left(\widetilde{\beta}\right)$, where $\mu_\beta^\star = \beta^{(t)}, \Sigma_\beta^\star = \rho \mathcal{I}^{-1}\left(\widehat{\beta}_{mle}; \widehat{\widetilde{\mu}}_{mle}\right)$ or $\rho \mathcal{I}^{-1}\left(\beta^{(t)}; \widetilde{\mu}^{(t)}\right)$, where $A = \left\{\beta \in \mathbb{R}^p : g^{-1}\left(x_i^T \beta\right) \in \mathcal{Y}\right\}$, $\rho \in (0, 1]$ is a tuning parameter, and $\mathcal{I}^{-1}(\beta^{(t)}; \widetilde{\mu}^{(t)})$ is the inverse Fisher information matrix associated with β at $\beta = \beta^{(t)}$. That is,

$$
\mathcal{I}\left(\beta^{(t)}; \widetilde{\mu}^{(t)}\right) = \sum_{i=1}^n \frac{x_i x_i^T}{\left(g'\left(\mu_i^{(t)}\right)\right)^2 b''\left(\theta_i^{(t)}\right)} ,
$$

$g'\left(\mu_i^{(t)}\right) = g'\left(g^{-1}\left(x_i^T \beta^{(t)}\right)\right)$, $b''\left(\theta_i^{(t)}\right) = {\sigma_i^{(t)}}^2$. Note that $b(\theta) = \ln T(\theta)$. Writing $\widetilde{\mu}^{(t)}(\cdot) = \sum_\ell \widetilde{J}_\ell^{(t)} \delta_{\widetilde{z}_\ell^{(t)}}(\cdot)$ and hence, we solve for $\theta_i^{(t)}$ from $g^{-1}\left(x_i^T \beta^{(t)}\right) = \frac{\sum_\ell \widetilde{z}_\ell^{(t)} \exp\left(\theta_i^{(t)} \widetilde{z}_\ell^{(t)}\right) \widetilde{J}_\ell^{(t)}}{\sum_\ell \exp\left(\theta_i^{(t)} \widetilde{z}_\ell^{(t)}\right) \widetilde{J}_\ell^{(t)}}$. Omitting (t)

superscript from CRM for better readability, we have:

$$b''\left(\theta_i^{(t)}\right) = \frac{\sum_\ell \widetilde{z}_\ell^{\,2} \exp\left(\theta_i^{(t)} \widetilde{z}_\ell\right) \widetilde{J}_\ell}{\sum_\ell \exp\left(\theta_i^{(t)} \widetilde{z}_\ell\right) \widetilde{J}_\ell} - \left[\frac{\sum_\ell \widetilde{z}_\ell \exp(\theta_i^{(t)} \widetilde{z}_\ell) \widetilde{J}_\ell}{\sum_\ell \exp(\theta_i^{(t)} \widetilde{z}_\ell) \widetilde{J}_\ell}\right]^2 .$$

1.1.2 $\widetilde{\mu}$ (also, $\{z_i\}_{i=1}^n, \{u_i\}_{i=1}^n$) updates —

We assume θ_i not depending on $\widetilde{\mu}$ and consider $\theta_i = \theta_i^{(t)} = \theta(\beta^{(t+1)}, \widetilde{\mu}^{(t)}, x_i)$ to be fixed for simulating $\widetilde{\mu}^{(t+1)}$. We sample $\widetilde{\mu}$ from $[\widetilde{\mu} \mid \mathbf{y}, \mathbf{z}] \stackrel{d}{=} [\widetilde{\mu} \mid \mathbf{z}]$, and $\mathbf{z}$ from $[\mathbf{z} \mid \widetilde{\mu}, \mathbf{y}]$. For simulating $\widetilde{\mu}$, we further introduce latent variables $u_1, \ldots, u_n$ and sample $\widetilde{\mu}$ from $[\widetilde{\mu} \mid \mathbf{u}, \mathbf{z}]$ after we sample $\mathbf{u}$ from $[\mathbf{u} \mid \mathbf{z}]$.

1. $[\mathbf{u} \mid \mathbf{z}^{(t)}]$: $\Pr\left(\mathbf{u} \mid \mathbf{z}^{(t)}\right) \propto e^{-\psi(\mathbf{u})} \prod_{i=1}^n \int_{\mathbb{R}^+} s e^{-s\left\{\sum_{i=1}^n u_i \exp(\theta_i^{(t)} z_i^{(t)})\right\}} \rho(ds \mid z_i^{(t)})$, with

$$\begin{aligned} e^{-\psi(\mathbf{u})} &= \exp\left\{-\int_{\mathbb{R}^+ \times \mathcal{Y}} \left[1 - e^{-\left\{\sum_{i=1}^n u_i \exp(\theta_i^{(t)} z)\right\} s}\right] \nu(ds, dz)\right\} \\ &= \exp\left\{-\int_{\mathcal{Y}} \int_{\mathbb{R}^+} \left[1 - e^{-\left\{\sum_{i=1}^n u_i \exp(\theta_i^{(t)} z)\right\} s}\right] \alpha \frac{e^{-s}}{s} ds G_0(dz)\right\} \\ &= \exp\left\{-\alpha \int_{\mathcal{Y}} \ln\left[1 + \sum_{i=1}^n u_i \exp(\theta_i^{(t)} z)\right] G_0(dz)\right\} \end{aligned}$$

Similarly, $\prod_{j=1}^n \int_{\mathbb{R}^+} s e^{-s\left\{\sum_{i=1}^n u_i \exp(\theta_i^{(t)} z_j^{(t)})\right\}} \rho(ds \mid z_j^{(t)}) = \cdots \frac{e^{-s}}{s} ds = \prod_{j=1}^n \int_{\mathbb{R}^+} e^{-s\left\{1 + \sum_{i=1}^n u_i \exp(\theta_i^{(t)} z_j^{(t)})\right\}} ds = \prod_{j=1}^n \frac{1}{\left\{1 + \sum_{i=1}^n u_i \exp(\theta_i^{(t)} z_j^{(t)})\right\}} = \exp\left[-\sum_{j=1}^n \ln\left\{1 + \sum_{i=1}^n u_i \exp(\theta_i^{(t)} z_j^{(t)})\right\}\right]$. Hence,

$$\begin{aligned} Pr\left(\mathbf{u} \mid \mathbf{z}^{(t)}\right) &\propto \exp\left\{-\int_{\mathcal{Y}} \ln\left[1 + \sum_{i=1}^n u_i \exp(\theta_i^{(t)} z)\right] \left(\alpha G_0(dz) + \sum_{j=1}^n \delta_{z_j^{(t)}}(dz)\right)\right\} \\ &\propto \exp\left\{-\int_{\mathcal{Y}} \ln\left[1 + \sum_{i=1}^n u_i \exp(\theta_i^{(t)} z)\right] G_n(dz)\right\} \end{aligned}$$

where $G_n(\cdot) = \alpha G_0(\cdot) + \sum_{j=1}^n \delta_{z_j^{(t)}}(\cdot)$. The second term can also be written as $\sum_{\ell=1}^{\widetilde{n}} n_\ell \delta_{z_\ell^\star}(\cdot)$, where $\{z_\ell^\star\}_{\ell=1}^{\widetilde{n}}$ are the unique values, among $\{z_j^{(t)}, j = 1, \ldots, n\}$, with $\{n_\ell\}_{\ell=1}^{\widetilde{n}}$ ties.

Three possible approaches:

1. We generate $u_j \sim \Pr(u_j \mid u_{1:j-1}^{(t+1)}, u_{j+1:n}^{(t)}, \mathbf{z}^{(t)}), j = 1, \ldots, n$. In practice, we generate from an exact distribution (calculation depends on the choice of G_0) or in case of difficulty, we can always approximate it using Monte Carlo, $\approx \exp\left(-(1/R) \sum_{r=1}^R \Psi(z_r; u_j)\right), \quad z_r \sim G_n$, where $\Psi(z_r; u_j) = \ln\left[1 + \sum_{i=1}^{j-1} u_i^{(t+1)} \exp(\theta_i^{(t)} z_r) + \sum_{i=j+1}^n u_i^{(t)} \exp(\theta_i^{(t)} z_r) + u_j \exp(\theta_j^{(t)} z_r)\right]$. We can then followup with an MH acceptance-rejection step.

2. [Idea from Igor's Stat Sci 2013 paper] We generate proposal using random walk, $u_j \sim Gamma\left(\delta, \delta / u_j^{(t)}\right), j = 1, \ldots, n$ and followed up with an MH acceptance-rejection step. The

hyperparameter $\delta(\geq 1)$ controls the acceptance rate of the M-H step being higher for larger values.

3. Should we generate $v = \sum_{i=1}^{n} u_i \exp(\theta_i^{(t)} z)$ and then use that somehow to generate u_j's? Think about proposal: some transformation eg., log? something more clever?

Note: $P(u_j, T_j) = T_j e^{-u_j T_j} P(T_j) \implies$ the prior on $u_j \mid T_j$ is gamma$(1, T_j)$. Hence for initializing u_j, we generate from Gamma(1, T_j).

2. $[\widetilde{\mu} \mid \mathbf{u}^{(t+1)}, \mathbf{z}^{(t)}]$:

   (a) $\widetilde{\mu}^{\left(\mathbf{u}^{(t+1)}\right)} \sim$ CRM $\left(\nu^{(\mathbf{u})}\right)$ with $\nu^{(\mathbf{u})}(ds, dz) = e^{-\left\{\sum_{i=1}^{n} u_i^{(t+1)} \exp(\theta_i^{(t+1)} z)\right\}s} \rho(ds \mid z) G_0(dz) = e^{-\Psi(z)s} \rho(ds \mid z) G_0(dz) = \alpha \frac{e^{-(\Psi(z)+1)s}}{s} ds G_0(dz)$. $\widetilde{\mu}^{\left(\mathbf{u}^{(t+1)}\right)}(\cdot) = \sum_{\ell=1}^{M} J_\ell^{(t+1)} \delta_{z_\ell^{(t+1)}}(\cdot)$ (random locations z_ℓ and jumps J_ℓ are generated using Ferguson and Klass (1972) algorithm; M is used for finite approximation of the CRM). Omitting superscript u for better readability.

$$N(v) = \nu([v, \infty], \mathcal{Y}) = \int_v^\infty \int_{\mathcal{Y}} \nu(ds, dz) = \alpha \int_v^\infty \int_{\mathcal{Y}} \frac{e^{-(\Psi(z)+1)s}}{s} ds G_0(dz)$$

   (b) $\Pr\left(J_j^{\left(\mathbf{u}^{(t+1)}, z_j^{(t)}\right)}\right) \propto s e^{-s\left\{\sum_{i=1}^{n} u_i^{(t+1)} \exp(\theta_i^{(t+1)} z_j^{(t)})\right\}} \rho(ds \mid z_j^{(t)}) = e^{-\left(\Psi\left(z_j^{(t)}\right)+1\right)s} \sim EXP$ $\left(mean = 1 \Big/ \left(\Psi\left(z_j^{(t)}\right) + 1\right)\right)$. Let's assume we have r unique fixed locations $(z_1^\star, \ldots, z_r^\star)$. Then for jumps, we generate from $\Pr(J_\ell^\star \mid -) \propto s^{n_\ell - 1} e^{-\left(\Psi\left(z_\ell^{\star(t)}\right)+1\right)s} \equiv Gamma\left(n_\ell, \Psi\left(z_\ell^{\star(t)}\right) + 1\right)$ $, \ell = 1, \ldots, r$.

3. $[\mathbf{z} \mid \widetilde{\mu}^{(t+1)}, \mathbf{y}]$:

$$\Pr[z_i \mid \widetilde{\mu}^{(t+1)}, \mathbf{y}] \propto \sum_{\ell} K(y_i \mid z_i) \exp(\theta_i z_i) \widetilde{J}_\ell^{(t+1)} \delta_{\widetilde{z}_\ell^{(t+1)}}(z_i) \ ,$$

   where $\widetilde{J} := \{\widetilde{J}_\ell^{(t+1)}\}_{\ell \geq 1} = \{J_1^{\star(t+1)}, \ldots, J_r^{\star(t+1)}, {J_1}^{(t+1)}, \ldots, \}$ [first r jumps are from 2(b) and rest from 2(a)]. Similarly, $\widetilde{Z} := \{\widetilde{z}_\ell^{(t+1)}\}_{\ell \geq 1} = \{z_1^{\star(t+1)}, \ldots, z_r^{\star(t+1)}, {z_1}^{(t+1)}, \ldots, \}$. Hence, $z_i \sim$ *a categorical distribution.*

4. (Optional) Resample fixed locations $z_i(or,\ z_\ell^\star)$ to reduce 'sticky clusters effect'.