DP-GLM and Posterior Sampling Steps

1 DP-GLM

$$y_i \mid z_i, x_i \sim K(y_i \mid z_i, x_i) = K(y_i \mid z_i), \quad y_i, z_i \in \mathcal{Y} \tag{1}$$

$$z_i \mid x_i = x, \widetilde{\theta}_x, \widetilde{\mu} \sim p_x(z_i) \propto \exp(\widetilde{\theta}_x z_i)\widetilde{\mu}(z_i) \tag{2}$$

$$\widetilde{\theta}_x \mid \theta_x \sim p(\widetilde{\theta}_x \mid \theta_x), \text{ with } b'(\theta_x) = \int_{\mathcal{Y}} z \frac{\exp(\theta_x z)\widetilde{\mu}(z)}{\int_{\mathcal{Y}} \exp(\theta_x u)\widetilde{\mu}(u)du} dz = g^{-1}(x'\beta) \tag{3}$$

$$\widetilde{\mu} \sim \text{ gamma CRM}(\nu), \text{ with } \nu(ds, dz) = \frac{e^{-s}}{s} ds \cdot \alpha G_0(dz) \tag{4}$$

$$\beta \sim \text{ MVN}(\mu_\beta, \Sigma_\beta). \tag{5}$$

1.1 Notes

- Eq. (4) implies $\frac{\widetilde{\mu}}{\widetilde{\mu}(\mathcal{Y})} \sim DP(\alpha, G_0)$.
- Symmetric $K(\cdot \mid z)$ in z respects the GLM mean regression. $\mathbb{E}(y_i \mid x_i) = \mathbb{E}_{z_i,\widetilde{\theta}_i|x_i}\mathbb{E}(y_i \mid x_i, z_i, \widetilde{\theta}_i) = \mathbb{E}_{z_i,\widetilde{\theta}_i|x_i}(z_i) = \mathbb{E}_{\widetilde{\theta}_i|x_i}\mathbb{E}_{z_i|x_i,\widetilde{\theta}_i}(z_i) = \mathbb{E}_{\widetilde{\theta}_i|x_i}(b'(\widetilde{\theta}_i)) = b'(\theta_i) = g^{-1}(x_i'\beta)$. The second last equality holds only if b' is a linear function.
- $p(\widetilde{\theta}_x \mid \theta_x) = N(\theta_x, \sigma_\theta^2)$ makes life easy! Closed form complete conditional for $\widetilde{\theta}_x$. We choose σ_θ to be small positive fraction. This ensures $\mathbb{E}(y_i \mid x_i) = g^{-1}(x_i'\beta) + R_i \approx g^{-1}(x_i'\beta)$, where $R_i \approx 0$.
- Choices to be made for $\alpha, G_0, K, \mu_\beta, \Sigma_\beta, g$.

1.2 Modeling fractional data

Here $\mathcal{Y} = [0, 1]$.

- $K(\cdot \mid z_i) =$ Uniform$(z_i - c_0, z_i + c_0)$.
- $G_0 =$ Uniform$(0, 1)$.
- $g(\mu) = \ln(\frac{\mu}{1-\mu})$ [logit link]

2 Posterior sampling steps

2.1 Notes

- Model parameters: $(\beta, \widetilde{\mu}, \widetilde{\theta}_i)$. Is $\widetilde{\theta}_i$ a model parameter or latent variable?
- Latent variables: $\{z_i, u_i\}_{i=1}^n$.
- Derived parameters: $\theta_i = b'^{-1}(g^{-1}(\eta_x))$, where $\eta_x = x^T\beta$ or $s(x)^T\beta$. We can go fully nonparametric here; bad thing is, we lose the β interpretation.
- M is the CRM finite truncation point. We use $M = 20$.

2.2 Steps

(A): $[\widetilde{\beta} \mid -] \sim N_p(\mu^\star_\beta, \Sigma^\star_\beta)1_A\left(\widetilde{\beta}\right)$ is the proposal for updating β, where $\mu^\star_\beta = \beta^{(t)}, \Sigma^\star_\beta = \rho\mathcal{I}^{-1}\left(\widehat{\beta}_{mle}; \widehat{\widetilde{\mu}}_{mle}\right)$, 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 $\mathcal{I}^{-1}\left(\widehat{\beta}_{mle}; \widehat{\widetilde{\mu}}_{mle}\right)$ is the inverse Fisher information matrix associated with β at $(\beta, \widetilde{\mu}) = \left(\widehat{\beta}_{mle}; \widehat{\widetilde{\mu}}_{mle}\right)$, where

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

$\mu_i = \mathbb{E}(y_i \mid x_i) = g^{-1}(x_i^T\beta)$ and $\theta_i = \theta(x_i; \beta, \widetilde{\mu})$ is the derived parameter. Hence the proposal distribution implicitly depends on $\widetilde{\mu}$ through dispersion matrix. We correct the MH step with an acceptance probability,

$$\delta(\widetilde{\beta} \mid \beta) = \frac{L(\widetilde{\beta})\pi(\widetilde{\beta})}{L(\beta)\pi(\beta)} \cdot \frac{q(\beta \mid \widetilde{\beta})}{q(\widetilde{\beta} \mid \beta)},$$

where

$$\begin{aligned}
L(\beta, \widetilde{\mu}) &= \prod_{i=1}^n p(y_i \mid x_i; \beta, \widetilde{\mu}) \\
&= \prod_{i=1}^n \int_{\widetilde{\theta}_i}\int_{z_i} p(y_i \mid z_i, \widetilde{\theta}_i, x_i; \beta, \widetilde{\mu}) \cdot p(z_i, \widetilde{\theta}_i \mid x_i; \beta, \widetilde{\mu}) dz_i d\widetilde{\theta}_i \\
&= \prod_{i=1}^n \int_{\widetilde{\theta}_i}\int_{z_i} p(y_i \mid z_i) \cdot p(z_i \mid \widetilde{\theta}_i, x_i; \beta, \widetilde{\mu}) \cdot p(\widetilde{\theta}_i \mid x_i; \beta, \widetilde{\mu}) dz_i d\widetilde{\theta}_i \\
&= \prod_{i=1}^n \int_{\widetilde{\theta}_i}\int_{z_i} K(y_i \mid z_i) \cdot p(z_i \mid \widetilde{\theta}_i; \widetilde{\mu}) \cdot p(\widetilde{\theta}_i \mid \theta_i) dz_i d\widetilde{\theta}_i,
\end{aligned}$$

with $K(y_i \mid z_i) = \frac{1}{2c_0}1(z_i - c_0 \leq y_i \leq z_i + c_0)$, $p(z_i = z \mid \widetilde{\theta}_i, \widetilde{\mu}) = \frac{\exp(\widetilde{\theta}_i z)\widetilde{\mu}(z)}{\int_{\mathcal{Y}} \exp(\widetilde{\theta}_i u)\widetilde{\mu}(u) du}$, $p(\widetilde{\theta}_i \mid \theta_i) = \text{Normal}(\widetilde{\theta}_i \mid \theta_i, \sigma^2_\theta)$, $\theta_i = b'^{-1}(g^{-1}(x_i^T\beta))$. Note that $L(\beta \mid \widetilde{\mu})$ is the conditional likelihood contribution of β given $\widetilde{\mu}$. So, given $\widetilde{\mu}(\cdot) = \sum_{m=1}^M J_m \delta_{s_m}(\cdot)$, $p(z_i = s_m \mid$

$\widetilde{\theta}_i, \widetilde{\mu}) = \frac{\exp(\widetilde{\theta}_i s_m) J_m}{\sum_{m'=1}^{M} \exp(\widetilde{\theta}_i s_{m'}) J_{m'}}$. Do we need to use marginal (instead of conditional) likelihood contribution of β? π and q are the prior and proposal density for β.

(B): $[\widetilde{\theta}_{x_i} \mid -]$ Normal$(\theta_{x_i} + z_i\sigma_\theta^2, \sigma_\theta^2)$, where $\theta_{x_i} = \theta(x_i; \beta, \widetilde{\mu})$ is the derived parameter.

(C): $[\mathbf{u} \mid -]$:

$$Pr\left(\mathbf{u} \mid \mathbf{z}^{(t)}, \theta^{(t)}\right) \propto \exp\left\{-\int_{\mathcal{Y}} \ln\left[1 + \sum_{i=1}^{n} u_i \exp(\theta_i^{(t)} z)\right] G_n(dz)\right\},$$

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.

Proposal for $\mathbf{u}$ [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.

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

(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 $\Pr(J_\ell^\star \mid -) \propto s^{n_\ell - 1} e^{-\left(\Psi\left(z_\ell^{\star(t)}\right)+1\right)s} \equiv Gamma\left(n_\ell, \Psi\left(z_\ell^{\star(t)}\right) + 1\right)$ $, \ell = 1, \ldots, r$.

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

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

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.*