Independent Metropolis using Laplace approximations for latent Gaussian process models

Matt Simpson

October 18, 2016

1 Expected acceptance probability

Let $\pi(\theta)$ denote the target posterior density and $J(\theta)$ the independent Metropolis proposal. Then the acceptance probability to move from θ to θ' is $a^*(\theta, \theta') = \min(a(\theta, \theta'), 1)$ where

$$a(\theta, \theta') = \frac{\pi(\theta')J(\theta)}{\pi(\theta)J(\theta')}.$$

Suppose the Markov chain has converged. Then $\theta' \sim p \perp\!\!\!\perp \theta \sim \pi$ and the expected acceptance probability can be expressed as[1]

$$\begin{aligned}
\mathrm{E}[a^*(\theta,\theta')] &= \mathrm{E}[a(\theta,\theta')\mathbb{1}\{a(\theta,\theta') \leq 1\}] + \mathrm{E}[\mathbb{1}\{a(\theta,\theta') > 1\}] \\
&= \iint\limits_{\{a(\theta,\theta')\leq 1\}} \frac{\pi(\theta')J(\theta)}{\pi(\theta)J(\theta')}\pi(\theta)J(\theta')\, d\theta\, d\theta' + \iint\limits_{\{a(\theta,\theta')>1\}} \pi(\theta)J(\theta')\, d\theta\, d\theta' \\
&= \iint\limits_{\{a(\theta',\theta)\geq 1\}} \pi(\theta')J(\theta)\, d\theta\, d\theta' + \iint\limits_{\{a(\theta,\theta')>1\}} \pi(\theta)J(\theta')\, d\theta\, d\theta' \\
&= 2\mathrm{E}[\mathbb{1}\{a(\theta,\theta') > 1\}] + \mathrm{E}[\mathbb{1}\{a(\theta,\theta') = 1\}] \\
&= 2P\left[\log\pi(\theta') - \log J(\theta') > \log\pi(\theta) - \log J(\theta)\right] \\
&\quad + P\left[\log\pi(\theta') - \log J(\theta') = \log\pi(\theta) - \log J(\theta)\right].
\end{aligned}$$

Now suppose that $J(\theta)$ is the Laplace approximation to $\pi(\theta)$ so that defining $\ell(\theta) = \log\pi(\theta)$ and $p(\theta) = N(\theta^*, H^{-1})$ where θ^* is the mode of $\pi(\theta)$ and H is the negative Hessian evaluated at the mode:

$$H = -\left.\frac{\partial^2 \ell}{\partial\theta\partial\theta'}\right|_{\theta=\theta^*}.$$

Then we have

$$\begin{aligned}
\ell(\theta) &= \ell(\theta^*) + (\theta-\theta^*)'\left.\frac{\partial \ell}{\partial\theta}\right|_{\theta=\theta^*} + \frac{1}{2}(\theta-\theta^*)'\left.\frac{\partial^2 \ell}{\partial\theta\partial\theta'}\right|_{\theta=\theta^*}(\theta-\theta^*) + R(\theta) \\
&= \ell(\theta^*) - \frac{1}{2}(\theta-\theta^*)'H(\theta-\theta^*) + R(\theta) \\
&= C + \log J(\theta) + R(\theta)
\end{aligned}$$

where $R(\theta)$ is the remainder term of the Taylor approximation and C is a constant. Then we have

$$\mathrm{E}[a^*(\theta,\theta')] = 2P\left[R(\theta') - R(\theta) > 0\right] + P\left[R(\theta') = R(\theta)\right].$$

where $\theta' \sim \pi \perp\!\!\!\perp \theta \sim J$.

[1] See optimal scaling of random walk metropolis paper for a theorem similar to this with more fancy details.

2 Characterizing the remainder term for exponential dispersion families

Supoose we have a latent Gaussian process $x_i \overset{ind}{\sim} N(w_i\beta, \phi^{-1})$ for $i = 1, 2, \ldots, N$ with $\phi > 0$ the precision of x_i's distribution. Further suppose that conditional on $x_{1:N}$ we have $y_i \overset{ind}{\sim} \pi(y_i|x_i, \lambda)$, an exponential dispersion family with link function $\eta(x)$:

$$\pi(y|x, \lambda) = \exp\left[\lambda(y\eta(x) - \kappa(x)) - c(y, \lambda)\right]$$

where $\lambda > 0$ is the dispersion parameter and both $\kappa(.)$ and $c(., .)$ are known functions. Suppose that λ and ϕ are both known and that $\beta \sim N(\bar{\beta}, \Omega^{-1})$ with mean $\bar{\beta}$ and precision matrix Ω known. Then we can write the log posterior of $x \equiv x_{1:N}$ and β as

$$\log \pi(x, \beta|y) \equiv \ell(x, \beta) = C + \lambda \sum_{i=1}^{N} [y_i\eta(x_i) - \kappa(x_i)] - \frac{\phi}{2} \sum_{i=1}^{N} (x_i - w_i\beta)^2 - \frac{1}{2}(\beta - \bar{\beta})'\Omega(\beta - \bar{\beta})$$

where C is an arbitrary constant. Then the negative Hessian evaluated at the posterior mode (x^*, β^*) is

$$H = - \left.\frac{\partial^2 \ell}{\partial(x, \beta)\partial(x', \beta')}\right|_{x=x^*, \beta=\beta^*} = \begin{bmatrix} D(x^*) & -\phi W \\ -\phi W' & \phi W'W + \Omega \end{bmatrix}$$

where $D(x^*)$ is an $N \times N$ diagonal matrix with $D_{ii} = \lambda[\kappa''(x_i^*) - \eta''(x_i^*)y_i] + \phi$, and W is an $N \times p$ matrix with rows w_i, and p is the dimension of β. So the proposal distribution is Gaussian with mean (x^*, β^*) and precision matrix H.

We can now characterize the remainder term $R(x, \beta) = \log \pi(x, \beta|y) - \log J(x, \beta)$:

$$\begin{aligned}
R(x, \beta) = C &+ \lambda \sum_{i=1}^{N} [y_i\eta(x_i) - \kappa(x_i)] - \frac{\phi}{2} \sum_{i=1}^{N} (x_i - w_i\beta)^2 - \frac{1}{2}(\beta - \bar{\beta})'\Omega(\beta - \bar{\beta}) \\
&+ \frac{1}{2} \sum_{i=1}^{N} (x_i - x_i^*)^2 (\lambda[\kappa''(x_i^*) - \eta''(x_i^*)y_i] + \phi) + \phi \sum_{i=1}^{N} (x_i - x_i^*)w_i(\beta - \beta^*) \\
&- \frac{1}{2}(\beta - \beta^*)'[\phi W'W + \Omega](\beta - \beta^*) \\
= C &+ \lambda \sum_{i=1}^{N} [y_i\eta(x_i) - \kappa(x_i)] + \frac{\lambda}{2} \sum_{i=1}^{N} (x_i - x_i^*)^2 [\kappa''(x_i^*) - \eta''(x_i^*)y_i] \\
&+ \beta'\Omega\bar{\beta} + \phi \sum_{i=1}^{N} x_i w_i \beta^* + \phi x_i^* w_i \beta - \beta'[\phi W'W + \Omega]\beta^* \\
= C &+ \lambda \sum_{i=1}^{N} [y_i\eta(x_i) - \kappa(x_i)] + \frac{\lambda}{2} \sum_{i=1}^{N} (x_i - x_i^*)^2 [\kappa''(x_i^*) - \eta''(x_i^*)y_i] \\
&+ \left[\Omega(\bar{\beta} - \beta^*) + \phi \sum_{i=1}^{N} x_i^* w_i - \phi\beta^* W'W\right] \beta + \phi \sum_{i=1}^{N} w_i \beta^* x_i.
\end{aligned}$$

Then

$$\begin{aligned}
R(x', \beta') - R(x, \beta) = &\left[\Omega(\bar{\beta} - \beta^*) + \phi \sum_{i=1}^{N} x_i^* w_i - \phi\beta^* W'W\right] (\beta' - \beta) + \phi \sum_{i=1}^{N} w_i \beta^* (x_i' - x_i) \\
+ \lambda \sum_{i=1}^{N} &\left[y_i[\eta(x_i') - \eta(x_i)] - \kappa(x_i') + \kappa(x_i)\right] + \frac{\lambda}{2} \sum_{i=1}^{N} \left[(x_i' - x_i^*)^2 - (x_i - x_i^*)^2\right] [\kappa''(x_i^*) - \eta''(x_i^*)y_i]
\end{aligned}$$

Thoughts

- So outside of the term with $\eta(x_i)$ and $\kappa(x_i)$ this look a lot like the probability that 1) a weighted difference between β' and β plus 2) a weighted average of $x_i' - x_i$ plus 3) the difference in weighted sample variances is ¿ 0.
- Might be able to characterize this for different link functions ($\eta(.)$). Maybe pick a particular distribution and see what pops out.
- Can instead characterize the remainder term using taylor's formula — and note that because of the exponential family, we can get a better grip on what integral looks like!

3 Generic framework and examples

Suppose we have data $z_i \overset{ind}{\sim} \pi(z_i|y_i, \phi)$ where $y_i = x_i'\beta$ with $\beta \sim N(\bar{\beta}, B^{-1})$ with ϕ, $\bar{\beta}$, and B known. $x_i'\beta$ could be a combination of fixed and random effects as long as the random effects are normally distributed with a known covariance matrix.

3.1 Gamma data model

In this case the log posterior is

$$\ell(\beta) = C + \sum_{i=1}^{N} \left[\phi e^{y_i} \log(\phi z_i) - z_i \phi - \log \Gamma(\phi e^{y_i})\right] - \frac{1}{2}(\beta - \bar{\beta})' B (\beta - \bar{\beta}).$$

Then the proposal distribution is $\beta \sim N(\beta^*, \Omega^{-1})$ where $\beta^* = \arg\max \ell(\beta)$, $\Omega = - \left.\frac{\partial^2 \ell}{\partial\beta\partial\beta'}\right|_{\beta=\beta^*}$ and

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial\beta\partial\beta'} =& \sum_{i=1}^{N} x_i x_i' \left[\phi e^{x_i'\beta} \log(\phi z_i) - \Psi(\phi e^{x_i'\beta})\phi e^{x_i'\beta} - \Psi'(\phi e^{x_i'\beta})\phi^2 e^{2x_i'\beta}\right] - B \\
=& - X'DX - B
\end{aligned}$$

where $\Psi(x) = d \log \Gamma(x)/dx$, $X = (x_1, x_2, \ldots, x_n)'$, and D is an $n \times n$ diagonal matrix with diagonal entries

$$D_{ii}(\beta) = -\phi e^{x_i'\beta} \log(\phi z_i) + \Psi(\phi e^{x_i'\beta})\phi e^{x_i'\beta} + \Psi'(\phi e^{x_i'\beta})\phi^2 e^{2x_i'\beta}.$$

So $\Omega = X'D^*X + B$ where $D^* = D(\beta^*)$.

Then the remainder term is

$$\begin{aligned}
R(\beta) =& C + \sum_{i=1}^{N} \left[\phi e^{x_i'\beta} \log(\phi z_i) - \log \Gamma(\phi e^{x_i'\beta})\right] - \frac{1}{2}(\beta - \bar{\beta})' B (\beta - \bar{\beta}) + \frac{1}{2}(\beta - \beta^*)'(X'D^*X + B)(\beta - \beta^*) \\
=& C + \sum_{i=1}^{N} \left[\phi e^{x_i'\beta} \log(\phi z_i) - \log \Gamma(\phi e^{x_i'\beta})\right] - \frac{1}{2}\beta' X' D^* X \beta - \left[(\beta^* - \bar{\beta})' B + (\beta^*)' X' D^* X\right] \beta.
\end{aligned}$$