Metropolis-Hastings Sampling

- When the full conditionals for each parameter cannot be obtained easily, another option for sampling from the posterior is the Metropolis-Hastings (M-H) algorithm.
- The M-H algorithm also produces a **Markov chain** whose values approximate a sample from the posterior distribution.
- For this algorithm, we need the form (except for a normalizing constant) of the posterior $\pi(\cdot)$ for $\boldsymbol{\theta}$, the parameter(s) of interest.
- We also need a **proposal** (or **instrumental**) distribution $q(\cdot|\cdot)$ that is easy to sample from.

Metropolis-Hastings Sampling

- The M-H algorithm first specifies an initial value for $\boldsymbol{\theta}$, say $\boldsymbol{\theta}^{[0]}$. Then:
- After iteration t, suppose the most recently drawn value is $\boldsymbol{\theta}^{[t]}$.
- Then sample a candidate value $\boldsymbol{\theta}^*$ from the proposal density.
- Let the $(t+1)$-st value in the chain be

$$\boldsymbol{\theta}^{[t+1]} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \min\{a(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{[t]}), 1\} \\ \boldsymbol{\theta}^{[t]} & \text{with probability } 1 - \min\{a(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{[t]}), 1\} \end{cases}$$

where

$$a(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{[t]}) = \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{[t]})} \frac{q(\boldsymbol{\theta}^{[t]} | \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{[t]})}$$

is the “acceptance ratio.”

Metropolis-Hastings Sampling

- In practice we would accomplish this by sampling $U^{[t]} \sim U(0,1)$ and choosing $\boldsymbol{\theta}^{[t+1]} = \boldsymbol{\theta}^*$ if $a(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{[t]}) > u^{[t]}$; otherwise choose $\boldsymbol{\theta}^{[t+1]} = \boldsymbol{\theta}^{[t]}$.
- Note that if the proposal density $q(\cdot|\cdot)$ is **symmetric** such that $q(\boldsymbol{\theta}^{[t]}|\boldsymbol{\theta}^*) = q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[t]})$, then the acceptance ratio is simply

$$\frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{[t]})}.$$

Metropolis-Hastings Example

Example 5 (Sparrow data): We gather data on a sample of 52 sparrows:

$$X_i = \text{age of sparrow (to nearest year)}$$
$$Y_i = \text{Number of offspring that season}$$

- We expect that the offspring number rises and then falls with age, so we assume a quadratic trend.
- We model the number of offspring at a given age x as Poisson:

$$Y|x \sim \text{Pois}(\mu_x)$$

Metropolis-Hastings Example

- Since we know μ_x must be positive, we use the model:

$$E[Y|x] = e^{\beta_0+\beta_1 x+\beta_2 x^2}$$

- This Poisson regression model is a **generalized linear model** (GLM).
- Our parameter of interest is $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$.
- But note that conjugate priors do not exist for **non-normal** GLMs.
- We will use the M-H algorithm to sample from our posterior.

Metropolis-Hastings Example

- Let the prior on $\boldsymbol{\beta}$ be multivariate normal with **independent** components:

$$\boldsymbol{\beta} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}), \text{ where } \boldsymbol{\Sigma} = 100 \times \mathbf{I}_3$$

- We will choose our **proposal** density to be multivariate normal with mean vector $\boldsymbol{\beta}^{[t]}$ (the current value).
- The covariance matrix of the proposal density is sort of a tuning parameter. We will choose

$$\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \text{ where } \hat{\sigma}^2 = \text{var}\{\ln(y_1 + 0.5), \ldots, \ln(y_n + 0.5)\}.$$

- We can adjust this if our **acceptance rate** is too high or too low.
- Usually we like an acceptance rate between 20% and 50%.

Metropolis-Hastings Example

- Since our proposal density is symmetric, our acceptance ratio is simply

$$\frac{\pi(\beta^*)}{\pi(\beta^{[t]})} = \frac{L(\beta^*|\mathbf{X},\mathbf{y})p(\beta^*)}{L(\beta^{[t]}|\mathbf{X},\mathbf{y})p(\beta^{[t]})}$$

$$= \frac{\prod_{i=1}^{n} \text{dpois}(y_i, \exp[\mathbf{x}_i^T \beta^*]) \prod_{j=1}^{3} \text{dnorm}(\beta_j^*, 0, 10)}{\prod_{i=1}^{n} \text{dpois}(y_i, \exp[\mathbf{x}_i^T \beta^{[t]}]) \prod_{j=1}^{3} \text{dnorm}(\beta_j^{[t]}, 0, 10)}$$

  where the Poisson density `dpois` and the normal density `dnorm` can be found easily in `R`.
- See `R` example with real sparrow data.

Other Metropolis-Hastings Issues

- In practice, it is recommended to check the acceptance rate (the proportion of proposed $\boldsymbol{\beta}^*$ values that are "accepted").
- We also check the serial correlation of the $\left\{\beta_j^{[t]}\right\}$ values using a plot of the **autocorrelation function**.
- If the values do not "appear" independent, we can alleviate this by choosing every k th value in the chain as our posterior sample (**thinning**).