Poisson Regression

Model:

$$y_i \sim \text{Pois}(\lambda_i), \quad i = 1, \ldots, n$$

$$\log \lambda_i = \underline{x}_i' \underline{\beta}$$

$$\underline{\beta} \sim N(\underline{\mu}_\beta, \Sigma_\beta)$$

Posterior:

$$[\underline{\beta} \mid \underline{y}] \propto [\underline{y} \mid \underline{\beta}][\underline{\beta}]$$

$$\propto \left(\prod_{i=1}^{n} \frac{(e^{\underline{x}_i' \underline{\beta}})^{y_i} e^{-e^{\underline{x}_i'\underline{\beta}}}}{y_i!} \right) \exp\left\{ -\frac{1}{2} (\underline{\beta} - \underline{\mu}_\beta)' \Sigma_\beta^{-1} (\underline{\beta} - \underline{\mu}_\beta) \right\}$$

Use M-H w/ block random walk proposals for $\underline{\beta}$.

MCMC Algorithm:

1.) Set $\underline{\beta}^{(0)}$, $k = 0$

2.) $k = k+1$

3.) Sample $\underline{\beta}^{(*)} \sim N(\underline{\beta}^{(k-1)}, \sigma^2_{tune} I)$

4.) Let $\underline{\beta}^{(k)} = \underline{\beta}^{(*)}$ w.p. $\min\left(\frac{[\underline{y} \mid \underline{\beta}^{(*)}][\underline{\beta}^{*}]}{[\underline{y} \mid \underline{\beta}^{(k-1)}][\underline{\beta}^{(k-1)}]}, 1 \right)$

and $\underline{\beta}^{(k)} = \underline{\beta}^{(k-1)}$ otherwise

5.) Goto 2, until $k = K$.

Model Checking:

We can use posterior prediction to check if model is capable of generating observed data.

Posterior Predictive P-value:

prediction

characteristic to check

$$P(f(\tilde{y}) \geq f(y) \mid y) =$$

posterior pred. dist'n

$$= \int 1_{\{f(\tilde{y}) \geq f(y)\}} [\tilde{y} \mid y] \, d\tilde{y}$$

pred. full-cond.

posterior

$$= \iint 1_{\{f(\tilde{y}) \geq f(y)\}} [\tilde{y} \mid \underline{\theta}, y][\underline{\theta} \mid y] \, d\underline{\theta} \, d\tilde{y}$$

note: extreme p-values suggest data are unlikely to arise from model.

With MCMC:

1.) Sample $\underline{\theta}^{(k)} \sim [\underline{\theta} \mid y]$ as usual w/ MCMC.

2.) Sample $\tilde{y}^{(k)} \sim [\tilde{y} \mid \underline{\theta}^{(k)}, y]$

3.) Compute $f(\tilde{y}^{(k)})$

4.) Compute $1_{\{f(\tilde{y}^{(k)}) \geq f(y)\}}$

5.) $P(f(\tilde{y}) \geq f(y) \mid y) \approx \dfrac{\sum_{k=1}^{K} 1_{\{f(\tilde{y}^{(k)}) \geq f(y)\}}}{K}$

Example: $f(y) = \dfrac{\sum_{i=1}^{n} (y_i - E(y_i \mid \underline{\theta}))^2}{n}$ (MSE)

Negative Binomial Regression:

$$y_i \sim NB(\mu_i, N) = \frac{\Gamma(y_i + N)}{\Gamma(N)\, y_i!} \left(\frac{N}{N+\mu_i}\right)^N \left(1 - \frac{N}{N+\mu_i}\right)^{y_i}$$

$$\log(\mu_i) = \underline{x}_i' \underline{\beta}$$

$$\underline{\beta} \sim N(\underline{\mu}_\beta, \Sigma_\beta)$$

$$\log(N) \sim N(\mu_N, \sigma_N^2)$$

Posterior:

$$[\underline{\beta}, \log(N) \mid \underline{y}] \propto [\underline{y} \mid \underline{\beta}, \log N][\underline{\beta}][\log N]$$

Full-conditional Distns:

$$[\underline{\beta} \mid \cdot] \propto \left(\prod_{i=1}^{n} [y_i \mid \underline{\beta}, \log N]\right)[\underline{\beta}]$$

$$[\log N \mid \cdot] \propto \left(\prod_{i=1}^{n} [y_i \mid \underline{\beta}, \log N]\right)[\log N]$$

MCMC Algorithm:

1.) Set $\underline{\beta}^{(0)}$, $\log N^{(0)}$, $k=0$
2.) $k = k+1$
3.) Use M-H to update $\underline{\beta}$ w/ $\underline{\beta}^{(*)} \sim N(\underline{\beta}^{(k-1)}, \sigma^2_{\beta\,\text{tune}} I)$
4.) use M-H to update $\log N$ w/ $\log N^{(*)} \sim N(\log N^{(k-1)}, \sigma^2_{N\text{tune}})$
5.) Goto 2, until $k = K$.