Some Examples

May 29, 2016

Example 1

- Suppose we want to generate n random points in the circle of radius 1 centered at the origin, conditional on the event that no two points are within a distance d of each other,where $\beta = P(\text{no two points are within d of each other})$ is assumed to be a small positive number.
- We generate the points with Gibbs sampling.

Example 1

Algorithm:

- Generate n points in the circle randomly.
- Generate a random number U, and let $I = int(nU) + 1$,
- Generate a random point in the circle.
- If this point is not within d of any other $n-1$ points excluding x_I, then replace x_I by this point; otherwise, generate a new point and repeat the process.

Example 2

- Let $X_i, i = 1, 2, 3, 4, 5$, be independent exponential random variables, with X_i having mean i, and suppose we are interested in using simulation to estimate

$$\beta = P\{\Pi_{i=1}^5 X_i > 120 | \sum_{i=1}^{5} X_i = 15\}$$

Example 2

- Suppose X and Y are independent exponentials with respective rates λ and μ, where $\mu < \lambda$. The conditional distribution of X given $X + Y = a$ is:

$$
\begin{aligned}
f_{X|X+Y}(x|a) &= C_1 f_{X,Y}(x, a - x), \\
&= C_2 e^{-\lambda x} e^{-\mu(a-x)}, \\
&= C_3 e^{-(\lambda-\mu)x}, 0 < x < a
\end{aligned}
$$

Example 2

Algorithm:

- Start with $(x_1, x_2, x_3, x_4, x_5)$ with $x_1 + x_2 + x_3 + x_4 + x_5 = 15$.
- Randomly choose two elements from $\{1, 2, 3, 4, 5\}$, say $I = 2$, $J = 5$.
- Generate the random variables x_2, x_5 with mean 2 and 5, given their sum $x_2 + x_5 = 15 - x_1 - x_3 - x_4$. This is generated by generating the value of an exponential with rate $1/2 - 1/5 = 3/10$, and set x_2 equal to that value and reset x_5 to make $x_1 + x_2 + x_3 + x_4 + x_5 = 15$.
- Repeat the process and the proportion of $\Pi_{i=1}^{5} X_i > 120$ is the estimate of β.

Bayesian Model Selection

	*Covariate*1	*Covariate*2	⋯	*Covariate*100	⋯
*A*1	2	2	⋯	1	⋯
*A*2	1	1	⋯	2	⋯
*A*3	1	2	⋯	1	⋯
*A*4	3	1	⋯	2	⋯
*B*1	2	2	⋯	2	⋯
*B*2	3	1	⋯	1	⋯
*B*3	1	2	⋯	2	⋯
*B*4	3	1	⋯	1	⋯

Define indicators $I = (I_1, I_2, \cdots, I_{100})$.
I_i means covariate i from H_1, $I_i = 2$ means covariate i from H_2.

$$P(data|I) = \Pi_{i=1}^{N_{covariate}} P(covariate_i|I_i)$$

$$P(covariate_i|I_i = 1) = P(covariate_i|H_1)$$

$$P(covariate_i|I_i = 2) = P(covariate_i|H_2)$$

$$P(I) = \Pi_{i=1}^{N_{covariate}} P(I_i)$$

$$P(I|data) \propto P(I)P(data|I) = \Pi_{i=1}^{N_{covariate}} P(I_i)P(covariate_i|I_i)$$

- H_1: A and B come from two different distributions.
- H_2: A and B are from the same distribution.
- H_3: Two covariates are dependent, but A's and B's are from different distributions.
- H_4: Two covariates are dependent, but A's and B's are from the same distributions.

Algorithm:

- Randomly assign I a starting value x_0.
- Propose: randomly choose one I_i, and change it to other values with equal probabilities, the new I is y.
- Evaluate: $q(x_t, y) = q(y, x_t) = 1/3$. $\alpha = \min\{1, \frac{P(I_i=y|data)}{P(I_i=x_t|data)}\}$
- Generate $U \sim U[0, 1]$, if $\alpha < U$, accept y, otherwise, reject y.
- If $t \geq N$, stop.

Linkage problem

- 197 animals are distributed into 4 categories $Y = (Y_1, Y_2, Y_3, Y_4)$ according to the genetic Linkage model:
$p_1 = (2+\phi)/4, p_2 = (1-\phi)/4, p_3 = (1-\phi)/4, p_4 = \phi/4$
- $(Y_1, Y_2, Y_3, Y_4) \sim multinomial(n = 197, p_1, p_2, p_3, p_4)$
- Under the prior:$\phi \sim beta(a, b)$, find the posterior mode.

EM method

- Multinomial distribution:

$$P(X_1 = n_1, \cdots, X_k = n_k) = \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}, \quad \sum_{i=1}^k n_i = n$$

- The observed data vector of frequencies:

$$y = (y_1, y_2, y_3, y_4)^T.$$

- Suppose the data are generated from a multinomial distribution with probabilities:

$$\frac{1}{2} + \frac{1}{4}\phi, \frac{1}{4}(1-\phi), \frac{1}{4}(1-\phi), \frac{1}{4}\phi$$

- Use MLE to estimate ϕ.

- The probability function is:

$$L(\phi; y) = \frac{n!}{y_1!y_2!y_3!y_4!}(\frac{1}{2} + \frac{1}{4}\phi)^{y_1}(\frac{1}{4}(1-\phi))^{y_2}(\frac{1}{4}(1-\phi))^{y_3}(\frac{1}{4}\phi)^{y_4}$$

- The log likelihood function apart from an additive term not involving ϕ is:

$$\log L(\phi) = y_1 \log(2+\phi) + (y_2 + y_3)\log(1-\phi) + y_4 \log\phi$$

- $$\frac{\partial \log L(\phi)}{\partial\phi} = \frac{y_1}{2+\phi} - \frac{y_2+y_3}{1-\phi} + \frac{y_4}{\phi}$$

$$I(\phi; y) = -\partial^2 \log L(\phi)/\partial\phi^2 = \frac{y_1}{(2+\phi)^2} + \frac{y_2+y_3}{(1-\phi)^2} + \frac{y_4}{\phi^2}$$

- Suppose $y_1 = y_{11} + y_{12}$, where y_{11} and y_{12} have probabilities $\frac{1}{2}$ and $\frac{1}{4}\phi$.
- Suppose y_{11}, y_{12} are unobservable, we only observe their sum y_1. Then the observed vector of frequencies y is viewed as being incomplete and the complete-data vector is taken to be

$$x = (y_{11}, y_{12}, y_2, y_3, y_4)^T.$$

They are assumed to arise from a multinomial distribution with probabilities

$$\frac{1}{2}, \frac{1}{4}\phi, \frac{1}{4}(1-\phi), \frac{1}{4}(1-\phi), \frac{1}{4}\phi$$

The log likelihood for the complete-data is:

$$\log L_c(\phi) = (y_{12} + y_4)\log\phi + (y_2 + y_3)\log(1-\phi)$$

- Use MLE, we will get $\phi = \frac{y_{12}+y_4}{y_{12}+y_2+y_3+y_4}$.
- Since the frequency y_{12} is unobservable, we are unable to estimate ϕ.
- We can use an iterative method to estimate ϕ.
- For unobserved data, we fill in by averaging the complete-data log likelihood over its conditional distribution given the observed data y.

- Given a specified value of ϕ^0, the conditional expectation of $\log L_c(\phi)$ can be written as:

$$Q(\phi; \phi^0) = E_{\phi^0}\{\log L_c(\phi)|y\}$$

- As $\log L_c(\phi)$ is a linear function of y_{11}, y_{12}, we can replace y_{12} by its current conditional expectation given the observed data y.
- The random variable Y_{11} corresponding to y_{11} has a binomial distribution with sample size y_1 and probability parameter $\frac{1}{2}/(\frac{1}{2} + \frac{1}{2}\phi^0)$.

$$E_{\phi^0}(Y_{11}|y_1) = y_{11}^0 = \frac{1}{2}y_1/(\frac{1}{2} + \frac{1}{4}\phi^0)$$

$$y_{12}^0 = y_1 - y_{11}^0 = \frac{1}{4}y_1\phi^0/(\frac{1}{2} + \frac{1}{4}\phi^0)$$

M-step: Maximization Q, we get

$$\phi^1 = \frac{y_{12}^0 + y_4}{y_{12}^0 + y_2 + y_3 + y_4}$$

Iteration steps:

$$\phi^{k+1} = (y_{12}^k + y_4)/(n - y_{11}^k)$$

where

$$y_{11}^k = \frac{1}{2}y_1/(\frac{1}{2} + \frac{1}{2}\phi^k)$$

$$y_{12}^k = y_1 - y_{11}^k$$