Stat 580 (Spring 2015): Homework 5

Due, in class, March 26, 2015

Please submit your homework with codes (hard copy) in class and upload the corresponding codes to the Blackboard. Problems marked with * will be graded in detail and they are worth 50% of the total score. Remaining problems, worth the remaining 50% of the total score, will be given full mark if reasonable amount of work is shown.

For this homework, use R for programming parts unless otherwise specified.

1. * In the *zero-inflated Poisson* (ZIP) model, random data $X_1, \ldots, X_n$ are assumed to be of the form $X_i = R_i Y_i$, where the Y_i's have a Poisson(λ) distribution and the R_i's have a Bernoulli(p) distribution, all independent of each other. Given an outcome $\boldsymbol{x} = (x_1, \ldots, x_n)$, the objective is to estimate both λ and p. Consider the following hierarchical Bayes model:

   - $p \sim \text{Uniform}(0, 1)$ (prior for p),
   - $(\lambda | p) \sim \text{Gamma}(a, b)$ (prior for λ),
   - $(r_i | p, \lambda) \sim \text{Bernoulli}(p)$ independently (from the model above),
   - $(x_i | \boldsymbol{r}, \lambda, p) \sim \text{Poisson}(\lambda r_i)$ independently (from the model above),

   where a and b are known parameters, and $\boldsymbol{r} = (r_1, \ldots, r_n)$. It follows that

   $$f(\boldsymbol{x}, \boldsymbol{r}, \lambda, p) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \prod_{i=1}^{n} \frac{e^{-\lambda r_i} (\lambda r_i)^{x_i}}{x_i!} p^{r_i} (1-p)^{1-r_i}.$$

   We wish to sample from the posterior pdf $f(\lambda, p, \boldsymbol{r} | \boldsymbol{x})$ using the Gibbs sampler.

   (a) Generate a random sample of size $n = 100$ for the ZIP model using parameters $p = 0.3$ and $\lambda = 2$.

   (b) Show that

   i. $(\lambda | p, \boldsymbol{r}, \boldsymbol{x}) \sim \text{Gamma}(a + \sum_i x_i, b + \sum_i r_i)$,

   ii. $(p | \lambda, \boldsymbol{r}, \boldsymbol{x}) \sim \text{Beta}(1 + \sum_i r_i, n + 1 - \sum_i r_i)$,

   iii. $(r_i | \lambda, p, \boldsymbol{x}) \sim \text{Bernoulli}\left(\frac{pe^{-\lambda}}{pe^{-\lambda} + (1-p) I_{\{x_i = 0\}}}\right)$.

   (c) Implement the Gibbs sampler: generate a large (dependent) sample from the posterior distribution and use this to construct 95% Bayesian confidence intervals for p and λ using the data in (a). Compare these with the true values. Try different values of a and b, but you can start with $a = b = 1$.

2. *Independence-Metropolis-Hastings Algorithm* is an importance-sampling version of MCMC. We draw the proposal from a fixed distribution g. Generally, g is chosen to be an approximation to f. The acceptance probability becomes

   $$r(x, y) = \min\left\{\frac{f(y)}{f(x)} \frac{g(x)}{g(y)}, 1\right\}.$$

   A random variable Z has a inverse Gaussian distribution if it has density

   $$f(z) \propto z^{-3/2} \exp\left\{-\theta_1 z - \frac{\theta_2}{z} + 2\sqrt{\theta_1 \theta_2} + \log \sqrt{2\theta_2}\right\}, z > 0,$$

   where $\theta_1 > 0$ and $\theta_2 > 0$ are parameters. It can be shown that

   $$E(Z) = \sqrt{\frac{\theta_2}{\theta_1}} \quad \text{and} \quad E\left(\frac{1}{Z}\right) = \sqrt{\frac{\theta_1}{\theta_2}} + \frac{1}{2\theta_2}.$$

   Let $\theta_1 = 1.5$ and $\theta_2 = 2$. Draw a sample of size 1,000 using the independence-Metropolis-Hastings algorithm. Use a Gamma distribution as the proposal density. To assess the accuracy, compare the mean of Z and $1/Z$ from the sample to the theoretical means. Try different Gamma distributions to see if you can get an accurate sample.

3. (optional) Consider a common application in statistics: three different treatments are to be compared by applying them to randomly selected experimental units. This, of course, usually leads us to "analysis of variance" using a model such as $y_{ij} = \mu + \alpha_i + e_{ij}$ with standard meanings of these symbols and the usual assumptions about the random component e_{ij} in the model. Suppose that instead of the usual assumptions, we assume that the e_{ij} have independent and identical double exponential distributions centered on zero.

   (a) Describe how you would perform a Monte Carlo test instead of the usual ANOVA test. Be clear in stating the alternative hypothesis.

   (b) Describe some other computer-intensive test that you could use even if you make no assumptions about the distribution of e_{ij}.