MATH 640: Bayesian Statistics

Homework 6, due Sunday, April 14

Please submit a PDF or .doc version of your homework to Blackboard by 11:59pm on the due date. Please type *all* responses. You are encouraged to use `R` for all calculations.

Theoretical Exercises

1. Assume $y_i \sim L(\mu, \sigma)$. We can represent this model as a mixture of a normal likelihood and inverse-gamma priors. Thus if we let $y_i \sim N\left(\mu, \frac{4\sigma^2}{\alpha_i}\right)$ where $\sigma^2 \sim IG(a, b)$ and $\alpha_i \overset{iid}{\sim} IG(1, 1/2)$ for fixed hyper-parameters a and b. Using this specification and a flat prior on μ, $\pi(\mu) \propto 1$, state the full posterior and determine the conditional posterior distributions for all model parameters, μ, σ^2, and $\alpha_1, \ldots, \alpha_n$. Write out the steps of a Gibbs Sampler you could use to draw posterior samples. (Hint: the conditional posterior for α_i, with some manipulation, should be recognizable as an inverse-Gaussian.)

2. Suppose we wish to build a more general Bayesian model for a binomial sample. Let $X \sim Binom(N, p)$. Further, let $p \sim Beta(\alpha, \beta)$ where $\alpha \sim Gamma(a_1, b_1)$ and $\beta \sim Gamma(a_2, b_2)$. Find the likelihood, posterior, and the full conditionals. If a full conditional is recognizable, state its name. If they are not recognizable, suggest a potential proposal distribution.

Computing Exercises

For each of the following exercises, please provide your code as part of your solution.

1. The Kumaraswamy distribution is a distribution that, like the Beta, can be used to model probabilities. It has as its pdf the following:

$$p(\theta) = ab\theta^{a-1}(1-\theta^a)^{b-1},\ \theta \in (0,1).$$

   We wish to draw samples from $p(\theta)$ when $a = 2$ and $b = 2$ using a Metropolis-Hastings algorithm. Compare the following proposal densities to each other based on acceptance rate, ACF, and the resulting sampled density: $Beta(1,1)$, $Beta(2,1)$, $Beta(2,2)$, and $Beta(3,2)$. Select the proposal density you think is best out of these four and provide the criteria by which you made your selection. Set the seed to 1218 and take $B = 20000$ samples. Discard the burn-in before examining ACF and the sampled density. Without thinning, do you notice any differences between proposals?

Analysis Exercises

1. Returning to the binomial model from the second theoretical exercise, write a sampler to generate posterior estimates fro p, α, and β. Note that you may need to use a metropolis-hastings step for some of the model components. Test your sampler by setting the seed to 1789 and generating 100 Bernoulli random variables with true probability of success 0.27, i.e. run `x <- rbinom(100, 1, 0.27)`. Discuss any acceptance rates that are needed as well as model convergence. For your test run, generate 20000 samples and set $a_1 = b_1 = a_2 = b_2 = 1$. Describe how well your model predicts the truth. (Hint: you may need the function `gamma()` in `R`.)

2. We are interested in estimating the country-specific variability of the risk of coup during the 1980s. Coup risk is calculated in the Rulers, Elections, and Irregular Governance or REIGN dataset[1]. We are interested in

[1]Bell, Curtis. 2016. The Rulers, Elections, and Irregular Governance Dataset (REIGN). Broomfield, CO: OEF Research. Full data available at oefresearch.org.

modeling each nation's standard deviation in log transformed risk. The summarized data can be found in the file `coupsd.txt`. A plausible model for standard deviations is the inverse-Gaussian which has a pdf of

$$p(s_i) = \left(\frac{\lambda}{2\pi s_i^3}\right)^{1/2} \exp\left[-\frac{\lambda(s_i-\mu)^2}{2\mu^2 s_i}\right].$$

Using a joint prior of $\pi(\lambda, \mu) \propto \lambda^{-1}$, determine the posterior and find the full conditionals. The full conditional for λ should be immediately recognizable. For the full conditional of μ, implement a Metropolis-Hastings step using a normal proposal density with mean equal to the sample mean of the data and standard deviation tuned to achieve an acceptance rate between 0.23 and 0.3. Set the starting value of $\mu^{(1)}$ equal to the sample mean the starting value of $\lambda^{(1)}$ equal to the inverse of the sample standard deviation. Using a seed of 1980, take $B = 20000$ posterior samples for both parameters. After dropping the burn-in, prepare ACF plots for the remaining samples. Next, thin those remaining samples by 5 (i.e. keep every fifth sample) and generate the ACF plot. Then thin by 10 and generate the ACF plot. Finally, thin by 20 and generate the ACF plot. What do you notice each time you increase your thinning? What level of thinning attains the most desirable ACF plot? How many samples are left?