MA 578 — Bayesian Statistics

Homework 5

(Due: Tuesday, 11/12/19)

1. BDA problem 11.2.

2. BDA problem 11.3.

3. The table below shows the survival times y_i of patients in a study on multiple myeloma[1].

```
y <- c(13, 52,  6, 40, 10,  7, 66, 10, 10, 14, 16,  4,
       65,  5, 11, 10, 15,  5, 76, 56, 88, 24, 51,  4,
       40,  8, 18,  5, 16, 50, 40,  1, 36,  5, 10, 91,
       18,  1, 18,  6,  1, 23, 15, 18, 12, 12, 17,  3)
```

Our goal is to conduct a Bayesian analysis of these survival times. You start by noticing that y is positive and right-skewed, and so it might benefit from a power transformation. In particular, the *Box-Cox* transformation of a dataset aims to stabilize its variance and bring its distribution closer to a normal distribution, and so it might be useful here. The Box-Cox transformation w_i with (power) parameter λ of observation y_i is given by

$$w_i = \begin{cases} \dfrac{{y_i}^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \log y_i & \text{if } \lambda = 0. \end{cases}$$

Assume that $w_i \,|\, \mu, \sigma^2 \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ and that the joint prior on λ, μ, and σ^2 is non-informative: $\mathbb{P}(\lambda, \mu, \sigma^2) \propto 1/\sigma$.

   (a) Show that the posterior is given by

$$\mathbb{P}(\lambda, \mu, \sigma^2 \,|\, y) \propto \frac{1}{\sigma} \prod_{i=1}^{n} N\big(w_i(\lambda) \,|\, \mu, \sigma^2\big) {y_i}^{\lambda - 1}.$$

   (b) Implement a Metropolis-within-Gibbs hybrid scheme to sample from this posterior. You should be able to sample directly from the conditional posteriors on μ and σ^2, but use a *random walk* Metropolis-Hastings step to sample from the conditional posterior on λ.

   (c) Provide trace plots, auto-correlation plots, and convergence diagnostics based on two parallel chains. Now give 95% posterior intervals for all parameters. Which transformation seems more likely: a log transformation or a square root transformation?

   (d) A new patient is about to enter the study. Obtain posterior samples for $\widetilde{y}$, its predictive survival time, and give a 95% posterior predictive interval for $\widetilde{y}$. Plot two boxplots, one for the predictive survival time and other for the data y, and compare them as a rough posterior predictive check; do you see any significant disagreements?

[1] This problem is adapted from Exercise 6.10 in "Bayesian Computation with R", by Jim Albert.