Homework 7

Math 315, Fall 2019

Due 11 October by 4:00 p.m.

Instructions: Complete the following problems and submit them by 4 p.m. on the due date. Please make sure that your solution is neatly written, clearly organized, and stapled (if there are multiple pages). You may complete derivations by hand, but any R work should be completed using R Markdown to render the final write up. You may hide the plotting code chunks, but please do not hide the code chunks where you derive your posteriors or simulate from distributions, since we may need to "dig into" these to point out minor errors.

1.

Suppose that $y_1, \ldots, y_n$ are a random sample from the Poisson/gamma density

$$f(y|\alpha, \beta) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!} \cdot \frac{\beta^\alpha}{(\beta+1)^{y+\alpha}},$$

where $\alpha > 0$ and $\beta > 0$. This density is an appropriate model for observed counts that show more dispersion (i.e. variability) than predicted under a Poisson model. Suppose that (α, β) are assigned the noninformative prior proportional to $1/(\alpha\beta)^2$. If we transform to the real-valued parameters $\theta_1 = log(\alpha)$ and $\theta_2 = log(\beta)$, the posterior density is proportional to

$$p(\theta_1, \theta_2 | y_1, \ldots, y_n) \propto \frac{1}{(\alpha\beta)^2} \prod_{i=1}^{n} \frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)y_i!} \cdot \frac{\beta^\alpha}{(\beta+1)^{y_i+\alpha}},$$

where $\alpha = e^{\theta_1}$ and $\beta = e^{\theta_2}$. Use this framework to model data collected by Gilchrist (1984), in which a series of 33 insect traps were set across sand dunes and the numbers of different insects caught over a fixed time were recorded. The number of insects of the taxa Staphylinoidea caught in the traps are given in the `insects.csv` data file and can be loaded using the below command:

```
insects <- read.csv("http://aloy.rbind.io/data/insects.csv")
```

(a) Use grid approximation to approximate the posterior density, and use Monte Carlo sampling to simulate 1000 draws from the joint posterior density of (θ_1, θ_2).

(b) From your Monte Carlo sample, calculate 90% interval estimates for the parameters α and β.

2.

The following table gives the records of accidents in 1998 compiled by the Department of Highway Safety and Motor Vehicles in Florida.

	Fatal	Nonfatal
None	1601	162527
Seat belt	510	412368

Denote the number of accidents and fatalities when no safety equipment was in use by n_N and y_N , respectively. Similarly, let n_S and y_S denote the number of accidents and fatalities when a seat belt was in use. Assume that y_N and y_S are independent with $y_N \sim \text{Binomial}(n_N, p_N)$ and $y_S \sim \text{Binomial}(n_S, p_S)$. Assume a uniform prior is placed on the vector of probabilities (p_N, p_S). In this problem, treat n_S and n_N as fixed and known.

(a) Derive the joint posterior distribution of (p_N, p_S).

(b) Show that p_N and p_S have independent beta posterior distributions.
(c) Use the function `rbeta()` to simulate 1000 values from the joint posterior distribution of (p_N, p_S).
(d) Using your sample, construct a histogram of the relative risk p_N/p_S. Calculate a 95% interval estimate of this relative risk.
(e) Construct a histogram of the difference in risks $p_N - p_S$.
(f) Compute the posterior probability that the difference in risks exceeds 0.