Bayesian Statistics: Final Exam (2020)

1. Suppose that θ has a Gamma(α, α) distribution; i.e., the scale parameter is the reciprocal of the shape so that the mean is 1. That is, the pdf of $Gamma(a, b)$ given by

$$\frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}.$$

Suppose that, conditional on θ, y has a Poisson distribution with mean $\mu\theta$.

   (a) Determine the marginal mean and variance of y.

   (b) Show that the marginal distribution of y has the following form:

$$p(y) = \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)} \left(\frac{\mu}{\mu+\alpha}\right)^y \left(\frac{\alpha}{\mu+\alpha}\right)^\alpha, \quad y = 0, 1, 2, \cdots$$

2. Assume that you want to investigate the proportion (θ) of defective items manufactured at a production line. Your colleague takes a random sample of 30 items. Three were defective in the sample.

   (a) Assume a uniform prior for θ. Compute the posterior distribution of θ.

   (b) Repeat (a), but this time using Jeffreys' prior.

   (c) Do a Bayesian test of the hypothesis: $H_0 : \theta = 0.05$. Assume a beta($\alpha = 1$, $\beta = 1$) prior. Calculate Bayes factor for the hypothesis testing.

3. Let $y_1, \ldots, y_n \overset{iid}{\sim} Gamma(a, a\theta)$ where a is known and $\theta > 0$. Note that the gamma pdf of $Gamma(\alpha, \beta)$ is given by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

   (a) Compute the Jeffreys' prior for θ. Then find the posterior distribution of θ.

   (b) Identify a conjugate prior and resulting posterior for θ.

   (c) Now assume that a is unknown, but it is to be fixed using an empirical Bayes approach. Find a value for a in terms of the sample mean and variance of $y_1, \ldots, y_n$, denoted by $\bar{y}$ and s^2, respectively. (Hint: Use the method of moment estimation.)

4. There is need to calculate $\int_\alpha^{\alpha+1} f(x)dx$, where $f(x)$ has the Pareto distribution $f(x) = \frac{2\alpha^2}{x^3}$, $\alpha < x < \infty$. A student suggests using a Metropolis-Hastings algorithm with candidate (jumping) density e^{-x}, $0 < x < \infty$.

   (a) Write out this algorithm in detail, showing exactly how to calculate the integral.

   (b) Another student suggests using the Cauchy as a candidate (jumping) density:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

   Briefly discuss some pros and cons of this.

(c) Can you suggest a way of calculating the integral that is better than those in parts (a) and (b). You do not have to provide anything, but you should define your choice.

5. The table below shows data on the numbers of work-related accidents $A_1, \ldots, A_8$ occurring within a sample of similar-sized companies in the same industry, recorded over a year, as part of a study about accident rates in the industry as a whole, which is made up of a much larger number of companies.

Index i	1	2	3	4	5	6	7	8
A_i	54	19	44	60	49	51	20	70

The WinBUGS code below implements a model intended to help with the interpretation of these data.

```
model{
  for(j in 1:N){
    L[j] ~ dnorm(M,P)
    R[j] <- exp(L[j])
    A[j] ~ dpois(R[j])
  }
  M ~ dnorm(3,0.25)
  P ~ dgamma(1,4)
  V <- 1/P
  S <- sqrt(V)
}

list(N=8,A=c(54,19,44,60,49,51,20,70))
```

The table below shows statistical summaries (in WinBUGS) of some of the output from running the model

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
L[1]	3.974	0.1363	0.001321	3.699	3.975	4.229	1001	10000
L[2]	2.96	0.2265	0.002139	2.495	2.967	3.376	1001	10000
L[3]	3.773	0.1504	0.001427	3.465	3.776	4.054	1001	10000
L[4]	4.08	0.1306	0.001223	3.815	4.083	4.327	1001	10000
L[5]	3.878	0.1439	0.001263	3.588	3.882	4.157	1001	10000
L[6]	3.919	0.1392	0.001389	3.637	3.923	4.188	1001	10000
L[7]	3.005	0.2189	0.00203	2.553	3.013	3.414	1001	10000
L[8]	4.232	0.1198	0.001179	3.991	4.233	4.463	1001	10000
M	3.705	0.4147	0.004599	2.831	3.708	4.524	1001	10000
R[1]	53.69	7.289	0.07072	40.42	53.25	68.67	1001	10000
R[2]	19.79	4.42	0.04223	12.13	19.44	29.26	1001	10000
R[3]	43.99	6.578	0.06294	31.97	43.66	57.65	1001	10000
R[4]	59.66	7.763	0.07137	45.39	59.34	75.75	1001	10000

R[5]	48.85	6.995	0.06093	36.16	48.51	63.88	1001	10000
R[6]	50.85	7.048	0.07051	37.96	50.54	65.91	1001	10000
R[7]	20.66	4.465	0.04041	12.85	20.36	30.39	1001	10000
R[8]	69.33	8.275	0.08152	54.12	68.94	86.7	1001	10000
S	1.135	0.3796	0.005883	0.5375	1.086	1.993	1001	10000

(a) Write down the model in mathematical terms.

(b) Based on the table, give 95% central posterior intervals (HPD) for the annual accident rate for company 1, and for the underlying average accident rate for the industry.

(c) What can you say about the variability of accident rates between companies? (using posterior mean and standard deviation)