LTCC Applied Bayesian Methods
Exam Questions 2010

All parts of all questions should be attempted.

Question I

A political party proposes a new tax on the bonuses paid to investment bankers. According to this proposal, individuals will pay 20% of their bonus in tax if and only if their bonus is greater than k (let's call such people 'large-bonus' individuals). Let Y denote the size of the bonus of a random high-bonus individual, and assume that the distribution of Y can be described by a Pareto distribution:

$$p(y \mid \theta, k) = \theta k^{\theta} y^{-(\theta+1)}, \;\; y > k,$$

where θ is an unknown parameter ($\theta > 0$).

1. Show that $X = \log(Y/k)$ has an Exponential(θ) distribution.

2. Briefly explain what is meant by a conjugate prior and comment on the main advantages and disadvantages of choosing such a prior.

3. In order to estimate θ, a sample of n large-bonus individuals are interviewed, their incomes y_i ($i = 1, ..., n$) are recorded, and these are transformed to yield data $\mathbf{x} = (x_1, ..., x_n)$, where $x_i = \log(y_i/k)$ ($i = 1, ..., n$). Assuming a conjugate Gamma(a, b) prior for θ, derive the posterior distribution of θ.

4. There are 100,000 high-bonus individuals in the country. Show that the expected total revenue from imposing this new tax, conditional on θ, is $20000\, \theta\, k/(\theta - 1)$ when $\theta > 1$, and derive its value when $\theta \leq 1$.

5. Suppose that $a = 6$, $b = 3$, $n = 4$, $y_1 = 50400$, $y_2 = 79100$, $y_3 = 73000$ and $y_4 = 109900$. Using your answer to Part 4, calculate the posterior expected total revenue from imposing this new tax.

Question II

Two random variables, X and Y, are measured on a sample of n individuals. Y is treated as an outcome and is regressed on X. Denote the values of X and Y for individual i ($i = 1, \ldots, n$) as X_i and Y_i respectively.

1. Write down equations for a Bayesian model for the linear regression of Y on X, remembering to specify all necessary priors. You should use suitable non-informative priors.

2. Draw a directed acyclic graph for this model, defining all notation you use.

3. Now suppose that the value of X is unobserved for some of the n individuals in the sample. When X is observed for all individuals in the sample, there is no need to specify a prior for it, but now that X is sometimes missing, a prior distribution is needed for it. Suppose we assume that $X_1, \ldots, X_n$ are independently and identically normally distributed with unknown mean μ_X and unknown variance ϕ_X.

   Draw a directed acyclic graph for this new, expanded model.

4. Use the global Markov property to determine, for each of the random variables in the model, whether or not it is conditionally independent of μ_X given $Y_1, \ldots, Y_n$.

Question III

A company employs ten salespeople to sell an undesirable product. The data in the table below are the numbers of successful sales y_i made by saleperson i $(i = 1, ..., 10)$ during t_i months spent working for the company.

	Salesperson, i									
	1	2	3	4	5	6	7	8	9	10
y_i	5	1	5	14	3	19	1	1	4	22
t_i	94.3	15.7	62.9	125.8	5.2	31.4	1.0	1.0	2.1	10.5

Conditional on the number of months spent working, $t_1, \ldots, t_{10}$, and 'true' success rates, $\theta_1, \ldots, \theta_{10}$, the numbers of successful sales by each salesperson i may be assumed to be independently distributed as

$$Y_i \sim \text{Poisson}(\theta_i t_i).$$

1. Assuming independent Gamma$(0.001, 0.001)$ prior distributions for each θ_i, calculate the posterior expected values of the 'true' success rates for salespersons 1, 5, 6 and 9.

2. An alternative model is to specify a hierarchical prior distribution for the 'true' success rates, as follows:

$$\begin{aligned} \theta_i &\sim \text{Gamma}(\alpha, \beta) \quad (i = 1, \ldots, 10) \\ \alpha &\sim \text{Exponential}(1) \\ \beta &\sim \text{Gamma}(0.1, 1) \end{aligned}$$

   Draw a directed acyclic graph (DAG) representing the relationships between all the variables in the model.

3. Derive the full conditional distributions for θ_i, α and β and show that those for θ_i and β are both Gamma distributions.

4. Explain in detail how these full conditional distributions are used in the Gibbs sampler to obtain a sample from the joint posterior distribution $p(\theta_1, \ldots, \theta_{10}, \alpha, \beta \mid \mathbf{y})$, where $\mathbf{y} = (y_1, ..., y_{10})$, and to estimate the posterior means of $\theta_1, ..., \theta_{10}$. Do not forget to explain how you would decide how long to run the Gibbs sampler.

5. Selected posterior summaries from running a Gibbs sampler for this model for 1000 'burn-in' iterations followed by a further 10,000 iterations are shown in the following table

parameter	posterior mean	posterior sd	2.5%	median	97.5%
α/β	0.9448	0.6477	0.3636	0.8175	2.2410
θ_1	0.0598	0.0254	0.0213	0.0563	0.1195
θ_5	0.6056	0.3150	0.1529	0.5529	1.3590
θ_6	0.6105	0.1393	0.3668	0.5996	0.9096
θ_9	1.5930	0.7728	0.4762	1.4601	3.4310

(i) Compare these results for θ_1, θ_5, θ_6 and θ_9 to your answers from Part 1, and comment on any differences.

(ii) According to the hierarchical model results shown above, the posterior means of θ_5 and θ_6 are very similar, but the latter has a much narrower 95% posterior credible interval. Comment on the reason for this.