## 3.4 Bayesian Theory Workshop 2

(17 marks total[10])

1. (2 marks) A motorist travels regularly from Edinburgh to St Andrews. On each occasion they choose a route, $r$, at random with equal probability from four possible routes. The probabilities of completing the journey, $j$, in under 1.5 hours via these routes are given in the table below. Given that on a certain occasion they complete the journey in under 1.5 hours, calculate the probability that they travelled on each of the possible routes.

| Route | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $Pr(j = 1.5 \| r)$ | 0.2 | 0.5 | 0.8 | 0.9 |

2. (4 marks) Assume that a Lateral Flow Test (LFT) for COVID19 has a sensitivity (i.e. probability of testing positive given that the person is infected $Pr(T+|I)$) estimated at approximately 0.75 and a specificity estimated at approximately 0.99 (i.e. probability of testing negative given that the person is not infected $Pr(T-|S)$), where $S$ denotes Susceptible but not Infected.

   (a) The "pre-test" or prior probability of infection of a random individual is given by the prevalence of that disease (number of infected individuals per capita) in their population. Let's suppose that in Edinburgh we currently have a prevalence of $2.22 * 10^{-2}$. What is the probability that a random person in Edinburgh who tested positive with an LFT has COVID19?

   (b) If we have more information we can group individuals according to risk and use the prevalence in that group instead. Let's suppose that based on the symptoms they are presenting and their contact history patients have been placed in the following risk groups.

| Risk | Moderate | Medium | High |
|---|---|---|---|
| $Pr(I_{risk})$ | 0.1 | 0.3 | 0.7 |

   Use Bayes Theorem to calculate the probability that a person with *medium* risk (e.g. more than one symptom but no known contact) has COVID19 if they tested positive.

   (c) Repeat for *moderate* and *high* risk.

   (d) Discuss with your peers the implications of these results.

3. (3 marks) An ecologist is surveying the abundance of ladybirds in a field by laying out a random sample of n grids of fixed size across the field and counting their number in each grid. You have been asked to examine this data. You assume that the number of ladybirds, Y, follows a Poisson($\mu$) distribution as is standard for biological counts, and use the Jeffrey's prior ($1/\sqrt{\mu}$) for $\mu$.

   (a) What is the corresponding posterior distribution for $\mu$, after observing data $y_1, y_2, \cdots, y_n$?

   (b) The observed counts from the random sample of $n = 10$ grids are shown below.

$$5, 6, 5, 6, 7, 5, 4, 5, 3, 6.$$

   What is the posterior distribution, mean, and standard deviation for $\mu$?

4. (3 marks) A discrete random variable Y can take on values 0, 1, or 2 and has the following probability mass function.

[10] This is not to turn in—you can self-mark.

| Y | 0 | 1 | 2 |
|---|---|---|---|
| $Pr(Y\|\theta)$ | $\theta$ | $\theta^2$ | $1-\theta-\theta^2$ |

where $0 < \theta < 1$. A discrete prior distribution is specified on $\theta$:

| $\theta$ | 0.2 | 0.4 | 0.6 |
|---|---|---|---|
| $\pi(\theta)$ | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

Show that the prior predictive distribution for Y is the following:

| Y | 0 | 1 | 2 |
|---|---|---|---|
| m(Y) | 0.35 | 0.15 | 0.5 |

5. (5 marks) Suppose that we specify a prior on the parameter $\theta$ such that,

$$\theta|\lambda \sim \text{Poisson}(\lambda)$$

$$\lambda \sim \text{Gamma}(\alpha, \beta).$$

The motivation: $\theta$ is a parameter that must take on non-negative integer values. While using a Poisson distribution ensures that, we don't think that the Poisson distribution alone will give a prior that is "wide enough" and so we are introducing some *extra-Poisson* variation.

(a) Show that the (marginal) distribution for $\theta$ is Negative Binomial by integrating out $\lambda$ :

$$p(\theta) = \int_{-\infty}^{\infty} p(\theta, \lambda) d\lambda = \int_{-\infty}^{\infty} p(\theta|\lambda) p(\lambda) d\lambda.$$

*Note*: The pmf for the Negative Binomial is:

$$Pr(\theta = k|\alpha, \beta) = \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)\Gamma(k+1)} \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left(\frac{1}{\beta+1}\right)^{k},$$

where k = 0, 1, 2,... . It may be useful to recall that for an integer $x$, $x! = \Gamma(x+1)$.

(b) Now suppose that we observe data x, such that,

$$X \sim Bin(\theta, p).$$

Thus $\theta$ corresponds to the number of Bernoulli trials.
Derive an expression for the corresponding posterior distribution for $\theta$ using the distribution from 5a as the prior. *Note*: the distribution is not of a standard form.