Bayesian Inference: Tutorial 2

1. Consider binary data where success is denoted with $X = 1$, and failure with $X = 0$. Suppose we observe x successes from n trials, so that $x_1 = 0, x_2 = 1, ..., x_n = 1$, $\sum_{i=1}^{n} x_i = x$. Given p, the observations are i.i.d. with, $P(X_i = 1|p) = p$.

    a) When the exact 0 or 1 observation for each one of the variables is modelled, the likelihood of the data is,

    $$P(\text{data}|p) = p(x_1, ..., x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^x(1-p)^{n-x}.$$

    Assuming a Beta prior for p, $Beta(a, b)$, derive the marginal distribution for $\boldsymbol{x}$, $f(\boldsymbol{x})$. Check that this is a valid probability mass function for $n = 1$.

    b) When the total number of successes is modelled, the likelihood of the data is the Binomial distribution,

    $$P(\text{data}|p) = p(x_1, ..., x_n|p) = \binom{n}{x} p^x(1-p)^{n-x}.$$

    Assuming the same Beta prior for p, $Beta(a, b)$, derive the marginal distribution for $\boldsymbol{x}$, $f(\boldsymbol{x})$. Check that this is a valid probability mass function for $n = 1$. Are the two marginal densities the same?

    Hint: You could use directly Bayes Theorem, with the posterior distribution derived in the lecture notes.

2. Suppose that we specify a prior on the parameter θ such that,

$$\begin{aligned} \theta|\lambda &\sim Po(\lambda) \\ \lambda &\sim \Gamma(\alpha, \beta), \end{aligned}$$

where α and β are known values. By integrating out λ (i.e. by using $p(\theta) = \int_{-\infty}^{\infty} p(\theta, \lambda)d\lambda = \int_{-\infty}^{\infty} p(\theta|\lambda)p(\lambda)d\lambda$), show that this is equivalent to the prior,

$$\theta \sim Neg - Bin(\alpha, \beta).$$

where α and β are the negative binomial shape and inverse scale (sometimes called rate) parameters respectively. (Note that distributional information is given in one of the Appendices in the Lecture Notes section in Moodle.)

Now suppose that we observe data x, such that, given θ and p,

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

Derive an expression for the corresponding posterior distribution for θ. (Note - the distribution is not of a standard form)

3. (From December 2012 exam - number in brackets correspond to number of marks - the total exam is out of 50 marks.) A biologist designs an experiment in order to investigate the variability in the recorded observations from some laboratory equipment. It is assumed that the observations, denoted $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, independently follow a normal distribution $N(1, \sigma^2)$, where σ^2 is unknown.

a) Consider an inverse Gamma $\Gamma^{-1}(\alpha, \beta)$ prior for σ^2. Show that the posterior distribution of σ^2 given the data is given by,

$$\sigma^2|\boldsymbol{x} \sim \Gamma^{-1}\left(\frac{n}{2} + \alpha, 0.5\sum_{i=1}^{n}(x_i - 1)^2 + \beta\right).[4]$$

b) Consider the two following priors: (i) $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$; and (ii) $\sigma^2 \sim U(0, K)$, where α, β, K are known. Comment on the differences between the two priors and any impact they may have on the posterior distribution of σ^2. [2]

c) Consider data $x_1 = 5, x_2 = 7, x_3 = 7, x_4 = 9$ and the prior $\sigma^2 \sim \Gamma^{-1}(1, 2)$. State (i) the posterior distribution for σ^2 and (ii) the posterior mean for $\frac{1}{\sigma^2}$. [2]

d) For the same data as in (c) above, consider the prior $\sigma^2 \sim U(0, 100)$. Calculate the posterior distribution for σ^2. [4]

4. (From December 2012 exam - number in brackets correspond to number of marks - the total exam is out of 50 marks.)
Consider a sequence of independent Bernoulli trials with probability of success p. A Bayesian experiment is designed in order to obtain inference on p where the data correspond to the number of failures, y, before a pre-determined total number of m successes. Conditional on p, Y has a negative Binomial distribution with probability mass function,

$$p(y|p, m) = \binom{m + y - 1}{y} p^m (1 - p)^y,$$

for $y = 0, 1, \ldots, \infty$.

a) Calculate Jeffreys' prior for p. [*Note:* $E_Y(Y) = m(1 - p)/p$]. [4]

b) When the observed data are the total number of successes after a set number of independent Bernoulli trials, Jeffreys' prior for p is $Beta(0.5, 0.5)$. Discuss the implication of the given result in relation to your calculation in (a) and the Likelihood Principle. [2]

5. We observe data $\boldsymbol{x}$ from a Poisson distribution, with unknown mean μ. Calculate the Jeffreys' prior for μ. Then, what is the corresponding posterior distribution for μ, after observing data $\boldsymbol{x}$? Suppose that we observe data:

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

Calculate the posterior mean for μ.

6. (From 2009 exam - number in brackets correspond to number of marks - the total exam is out of 50 marks.) Let $X_1, \ldots, X_n$ be independent and identically distributed $Po(\lambda)$ random variables.

(a) Suggest **two** uninformative priors that could be specified for the parameter λ, explicitly giving the form of the prior probability density functions. Comment on the advantages and disadvantages of the two suggested priors. [6]