Bayesian Inference and Computational Methods 2009

Homework 2

1. *Inference for a binomial parameter p.* Recall that the class of *beta* distributions form the natural *conjugate class* for the *binomial/Bernoulli* family, and that if y *successes* are observed in n *trials*, a $\beta(a, b)$ *prior distribution* is transformed to a $\beta(a + y, b + n - y)$ *posterior distribution.*

   Suppose now that a coin is to be tossed so as to make inference for the probability p of a head.

   (a) It is natural to choose a symmetric prior distribution $\beta(a, a)$ for some $a > 0$. Suppose it is believed that the coin is so badly made that there is a 20% chance that the value of p lies outside the range $[0.4, 0.6]$. Find, to the nearest integer, the value of a appropriate to the belief.

   (b) Suppose now that the coin is tossed 10 times, and that 1 head is observed. Using the prior distribution identified above, find the posterior distribution for p given these observations. Identify its mean and standard deviation, and use **R** to sketch it. Suppose further that the coin is to be tossed a further 5 times. Compute the predictive distribution of the number of heads to be obtained, the predictive probability that at most 2 heads will be obtained, and the predictive mean of the number of heads to be obtained.

   (c) Suppose instead that, in the original 10 tosses of the coin, it is only observed that *at most* 1 head is obtained. Find the new posterior distribution for p, and compare it graphically with that obtained previously.

2. *Inference for a geometric parameter p.* Let $\boldsymbol{y} = (y_1, \dots, y_n)$ be a vector of independent observations from a Geo(p) distribution (i.e. with probability function $p(y) = p(1-p)^{y-1}$, $y = 1, 2, \dots$). Write down the associated likelihood function. Deduce that Bayesian inference for the parameter p is exactly the same as if p were a binomial parameter and that in $\sum_{i=1}^n y_i$ trials a total of n successes were observed. (*Understand why this is!*) Deduce that the *beta* distributions again form the natural *conjugate class* for the *geometric* family. For the $\beta(a, b)$ prior distribution, what is the posterior distribution induced by the above data.

3. *Inference for a Poisson parameter λ.* Recall that the class of *gamma* distributions form the natural *conjugate class* for the *Poisson* family, and that for Poisson observations $\boldsymbol{y} = (y_1, \dots, y_n)$, a $\Gamma(a, b)$ *prior distribution* is transformed to a $\Gamma(a + \sum_{i=1}^n y_i, b + n)$ *posterior distribution.*

   Suppose now that, within a given portfolio, motorcycle insurance claims are assumed to arise as a Poisson process with rate λ per week. The insurance company intends a Bayesian estimation of λ.

   (a) Previous experience suggests the use of a prior distribution whose 0.1-quantile and 0.9-quantile are approximately 2.9 and 5.2 respectively (e.g., the probability that λ should be less than 2.9 is approximately 0.1). Show that if the prior distribution is taken to be $\Gamma(a, b)$ where, for simplicity, a and b are taken to be integer valued, then the best choices of these are $a = 20$ and $b = 5$.

   (b) Over a 45-week period the total number of claims is observed to be 280. Calculate the posterior distribution of λ and sketch its density along with that of the prior distribution. Comment.

   (c) Use this posterior distribution to estimate the predictive distribution of the number n of claims occurring in the next two weeks. Find the value of n which is exceeded with probability 0.05.

(d) Suppose instead that the company had had little previous experience of the claim rate, and had therefore decided to use an Exp(0.1) prior distribution. (*Why might this have been sensible?*) Given the above data, what would the posterior distribution of λ then have been, and how different would it have been from that found previously?

4. *Inference for a Poisson parameter λ—continued.* Show that the Jeffreys' prior distribution for a Poisson parameter λ is given by the *improper* prior distribution with density

$$\pi(\lambda) \propto \lambda^{-1/2}.$$

What is the corresponding posterior distribution, given Poisson observations $\boldsymbol{y} = (y_1, \ldots, y_n)$? Interpret this.

5. *Inference for an exponential parameter λ.*

(a) Show that the class of *gamma* distributions also form the natural *conjugate class* for the *exponential* family. Given a $\Gamma(a, b)$ *prior distribution*, and for exponential observations $\boldsymbol{y} = (y_1, \ldots, y_n)$, what is the corresponding *posterior distribution*? Compare this situation with that for inference for a Poisson parameter λ and comment intelligently. (*Hint:* recall the close connection between the Poisson and exponential distributions via the Poisson process.)

(b) Suppose that, in a certain slightly strange country (where the sole significant cause of death is accidents), insured lives have durations in years which are independently Exp(λ)-distributed. As usual a Bayesian estimation of λ is required. The prior distribution of λ is taken to be Exp(10) (to what sort of belief about likely lifetimes does this correspond?). A total of 40 new policyholders are followed for a 10 year period. At the end of this time it is observed that 18 of these have died, with the following insured lifetimes (years).

0.54	0.63	0.93	1.92	2.10	2.13	2.73	2.81	2.82
2.86	2.97	3.30	3.77	5.47	7.22	7.54	8.22	9.68

The remaining policyholders are still alive.

i. Use (e.g.) a Q-Q plot to investigate the assumption that lifetimes are exponentially distributed, and to make an informal estimate of λ.

ii. Obtain the posterior distribution of λ, and compare it with the prior.

iii. Given this posterior distribution, find the predictive distribution of the insured lifetime of a further policyholder. In particular identify the predictive probability that a further policyholder survives at least 10 years.

iv. Similarly, given this posterior distribution, find the predictive probability that two further policyholders both survive at least 10 years. Why is this not the square of the probability found for a single policyholder?

6. *[1997 Statistical Inference Exam, Q4 (modified)]* In a raid on a coffee shop, Bayesian trading inspectors take a random sample of n packets of coffee, each of nominal weight 125 g. They model these data as independent observations $Y_1, \ldots, Y_n$ from a Normal $\mathrm{N}\left(\mu, \sigma^2\right)$ distribution. They take σ^2 to be known, while for μ they assume a prior distribution of $\mathrm{N}\left(\mu_0, \sigma_0^2\right)$, where μ_0 and σ_0^2 are specified values.

   The data they obtain are (weights in grams):

   ```
   105.3 113.3 114.5 121.2 122.9 123.7 124.0 124.6 124.9 124.9
   124.9 125.1 125.5 125.9 126.8 127.7 128.2 128.3 128.5 130.2
   ```

   ($n = 20$, $\sum y_i = 2470.4$, $\sum y_i^2 = 305828.98$).

   The parameter values they assume are $\mu_0 = 126$, $\sigma_0^2 = 1$, $\sigma^2 = 4$.

   The inspectors can impose a fine if their 95% credible interval for μ falls wholly below the claimed value of $\mu = 125$ g.

   (a) Recall the result of the lecture notes for the posterior distribution for the $\mathrm{N}\left(\mu, \sigma^2\right)$ model (with σ^2 known) with $\mathrm{N}\left(\mu_0, \sigma_0^2\right)$ prior distribution for μ. Show that the inspectors' 95% credible interval for μ for these data does lie wholly below 125 g; they therefore impose a fine on the owners of the coffee shop.

   (b) Use a normal Q-Q plot (or other appropriate graphical technique) to investigate the validity of the assumed normal distribution for the data.

   (c) Comment briefly as to whether the inspectors are justified in imposing a fine on the basis of this sample.

   (d) Using the posterior distribution found above, calculate the predictive probabilities that

      i. a single further randomly chosen packet of coffee weighs at least 125 g;

      ii. two further randomly chosen packets of coffee have a mean weight of at least 125 g.

7. Suppose that $y_1, y_2, \ldots, y_n$ form a random sample of observations from an $\mathrm{Exp}(\lambda)$ distribution where λ is unknown.

   (a) Show that the Jeffreys' prior for λ in this case is given by $\pi(\lambda) \propto \lambda^{-1}$.

   (b) For $n = 5$ and $\sum_{i=1}^{5} y_i = 10$ calculate a 95% equal-tailed credible interval for λ using the Jeffreys prior.

Bayesian Inference and Computational Methods 2009

Homework 2 – Solutions

1. (a) It follows from the symmetry of the $\beta(a, a)$ distribution that we need to choose a such that the 0.1 quantile of this distribution is 0.4. This is true, to a very good approximation, if we take $a = 20$.

   (b) The posterior distribution is $\beta(21, 29)$. It has mean 0.420 and standard deviation 0.069. Suitable **R** code to sketch the density of this distribution is given by

```
p = seq(0,1,0.001)
plot(p,dbeta(p,21,29),type='l')
```

   (c) Let Z be the number of heads to be obtained in a further 5 tosses of the coin. Given p, we have $Z \sim \text{Bin}(5, p)$ and $\mathbf{E}Z = 5p$. Arguing exactly as in the example of the lecture notes, we have that the *predictive probability function* of Z, given this posterior distribution, is given by

$$f(z \mid \boldsymbol{y}) = \frac{49!5!}{20!28!54!} \frac{(20+z)!(33-z)!}{z!(5-z)!}, \qquad 0 \le z \le 5.$$

   The predictive probability that at most 2 heads will be obtained is 0.642 and the predictive mean of the number of heads to be obtained is 2.1. Note that the predictive mean may also be calculated directly as the expectation of the conditional expectation $5p$ with respect to the posterior distribution, i.e. the predictive mean is

$$\mathbf{E}_{\text{pos}}(5p) = 5\mathbf{E}_{\text{pos}}(p) = 2.1,$$

   where $\mathbf{E}_{\text{pos}}$ denotes expectation with respect to the posterior distribution.

   (d) In the case where, in the original 10 tosses of the coin, it is only observed that *at most* 1 head is obtained, the likelihood function becomes

$$L(p, \boldsymbol{y}) = (1-p)^9(1+9p),$$

   and so the *posterior density*, given these data, is given by

$$kp^{19}(1-p)^{28}(1+9p),$$

   where the normalising constant

$$k = \frac{1}{229}\frac{49!}{19!28!} = 7.162 \times 10^{13}.$$

2. The likelihood function is

$$L(p; \boldsymbol{y}) = \prod_{i=1}^{n} p(1-p)^{y_i - 1} = p^n(1-p)^{\sum_{i=1}^n y_i - n}.$$

This is the same as the likelihood function in the case where p is a binomial parameter and, in $\sum_{i=1}^n y_i$ trials, a total of n successes are observed. It follows that Bayesian inference for p is exactly the same in the two cases. For a $\beta(a, b)$ prior distribution, the posterior distribution induced by the above data is $\beta(a + n, b + \sum_{i=1}^n y_i - n)$.

3. (a) It is simply necessary to verify that, for the suggested values of a and b, the 0.1-quantile and 0.9-quantile are as suggested.

   (b) The posterior distribution of λ is $\Gamma(300, 50)$ (with mean 6.000 and standard deviation 0.346). Note that suitable **R** code to plot the density of this distribution, and then to add that of the prior distribution is given by

```
lambda = seq(1,20,.01)
plot(lambda, dgamma(lambda,300,50), type='l')
lines(lambda, dgamma(lambda,20,5))
```

   Note also that most of the contribution to the posterior distribution comes from the data.

   (c) Given λ, the number of claims occurring in the next two weeks has a Poisson distribution with parameter (mean) 2λ. Hence, given the above posterior distribution for λ (with density $\pi(\lambda \,|\, \boldsymbol{y})$ say), the predictive distribution of the number n of claims occurring in the next two weeks has probability function

$$
\begin{aligned}
f(n \,|\, \boldsymbol{y}) &= \int_0^\infty e^{-2\lambda} \frac{(2\lambda)^n}{n!} \pi(\lambda \,|\, \boldsymbol{y}) \, d\lambda \\
&= \int_0^\infty e^{-2\lambda} \frac{(2\lambda)^n}{n!} \frac{50^{300}}{299!} \lambda^{299} e^{-50\lambda} \, d\lambda \\
&= \frac{50^{300} 2^n}{299! n!} \int_0^\infty \lambda^{299+n} e^{-52\lambda} \, d\lambda \\
&= \frac{50^{300} 2^n (299+n)!}{299! n! 52^{300+n}} \\
&= \left(\frac{50}{52}\right)^{300} \binom{299+n}{n} 52^{-n}.
\end{aligned}
$$

   The smallest value of n which is exceeded with *predictive probability* at most 0.05 is 18. (The *predictive probability* of obtaining a value *greater* than 18 is 0.041. Note also that, if instead we simply replaced the posterior distribution of λ by its mean of 6, so that the number of claims occurring in the next two weeks was modelled as having a Pois(12) distribution, then, under this distribution, the probability that the number of claims would exceed 17 is 0.037. The reason why the two probabilities are relatively similar is that, as the earlier density plot shows, the posterior distribution of λ is here fairly closely concentrated around its mean.)

   (d) The prior distribution is now $\Gamma(1, 0.1)$, reflecting little prior knowledge about the claim rate. Hence, given the above data, the posterior distribution of λ is $\Gamma(281, 45.1)$. In particular this has mean 6.23 and standard deviation 0.37, and so is not too different from the earlier posterior distribution, again reflecting the fact that most of the weight of the inference is being carried by the data.

4. Given any vector of n observations $\boldsymbol{y}$, the log-likelihood function is given by

$$
l(\lambda; \boldsymbol{y}) = -n\lambda + \sum_{i=1}^n y_i \ln \lambda + k,
$$

for some k which does not depend on λ. Hence

$$
-l''(\lambda; \boldsymbol{y}) = \frac{\sum_{i=1}^n y_i}{\lambda^2} \quad \text{and so} \quad -\mathbf{E}(l''(\lambda; \boldsymbol{y})) = \frac{n}{\lambda}.
$$

Thus the Jeffreys' prior distribution for λ is given by the *improper* prior distribution with density $\pi(\lambda) \propto \lambda^{-1/2}$ as required.

The corresponding posterior distribution is $\Gamma(\frac{1}{2} + \sum_{i=1}^n y_i, n)$, placing nearly all the weight of inference on the data.

5. (a) For a vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ of independent observations from the Exp(λ) distribution, we have the likelihood function

$$l(\lambda; \boldsymbol{y}) = \lambda^n e^{-\lambda \sum_{i=1}^n y_i},$$

and so it is easy to check that a $\Gamma(a, b)$ *prior distribution* is transformed by the observations $\boldsymbol{y}$ to a $\Gamma(a + n, b + \sum_{i=1}^n y_i)$ *posterior distribution*. In particular, the class of *gamma* distributions form the natural *conjugate class* for the *exponential* family.

(b) Note that the Exp(10) prior distribution for λ has a median of 0.069. If this were the true value of λ then the mean lifetime $1/\lambda$ would be 14.4 years. However the Exp(10) prior distribution allows for a very wide spread of values of λ (very little prior information). [Indeed, under the *prior* distribution, the *predictive* mean lifetime $10 \int_0^\infty \lambda^{-1} e^{-10\lambda}\, d\lambda = \infty$.]

i. Suitable **R** code for a Q-Q plot to investigate the assumption that lifetimes are exponentially distributed is given by

```
plot(qexp(ppoints(40)[1:18]),sort(lifetimes))
```

where `lifetimes` is a vector of the 18 non-truncated lifetimes. An informal estimate of λ is given by the reciprocal of the slope of this plot, suggesting a value of λ of around 0.06.

ii. Arguing as in the example of the lecture notes, the posterior distribution of λ is $\Gamma(19, 10 + \sum_{i=1}^{18} y_i + 220) = \Gamma(19, 307.64)$. This has mean 0.0618 and standard deviation 0.0142 and is very much more concentrated than the prior distribution.

iii. For any given $\Gamma(a, b)$ posterior distribution, a further insured lifetime has predictive distribution function F with tail (survival function) given by

$$1 - F(t) = \frac{b^a}{\Gamma(a)} \int_0^\infty e^{-\lambda t} \lambda^{a-1} e^{-b\lambda}\, d\lambda = \left(\frac{b}{b+t}\right)^a.$$

The required predictive probability is given by $1 - F(10)$ with $a = 19$, $b = 307.64$, and hence is 0.545.

iv. For any given $\Gamma(a, b)$ posterior distribution, the predictive probability that *two* further policyholders both survive at least t years is

$$\frac{b^a}{\Gamma(a)} \int_0^\infty e^{-2\lambda t} \lambda^{a-1} e^{-b\lambda}\, d\lambda = \left(\frac{b}{b+2t}\right)^a.$$

For $a = 19$, $b = 307.64$ and $t = 10$ this is 0.302.

6. (a) For the given prior distribution and observed data, the posterior distribution of μ is $\mathrm{N}(123.93, 0.1667)$, and hence the inspectors' 95% credible interval for μ is (123.1124.7).

   (b) A normal Q-Q plot (`qqnorm`), or indeed any rough plot, of the data should indicate that the assumption of normality is suspect, with the three lowest observations giving a pronounced left tail to the data. The median is 124.9, much closer to the desired 125 g. than the mean. Finally, the sample variance is 36.1, much greater than the inspectors' assumed value of 4 (if we omit the 3 lowest values, this reduces to 5.3).

   (c) The coffee packets vary in weight much more than the inspectors assumed, perhaps representable as a normal distribution contaminated with occasional outliers. The substandard mean weight can be accounted for by these outliers (3 in this sample). While the inspectors' modelling is revealed as less than ideal for their job, they are probably correct in fining a shop with such poor quality control.

   (d) i. Since $\sigma^2 = 4$, it follows that, *given* μ, the probability that a single further randomly chosen packet of coffee weighs at least 125 g is $\Phi((\mu - 125)/2)$ where Φ is the (cumulative) distribution function of the $\mathrm{N}(0, 1)$ distribution. Hence, given the *posterior distribution* of μ, the *predictive* probability that a single further randomly chosen packet of coffee weighs at least 125 g is the expectation of $\Phi((\mu - 125)/2)$ with respect to this posterior distribution. This can be evaluated as an integral, or by simulation. More simply, we may note that $\Phi((\mu - 125)/2)$ is also the probability, given μ, that μ plus an $\mathrm{N}\left(0, 2^2\right)$ random variable exceeds 125, so that the required predictive probability is the probability that an $\mathrm{N}(123.93, 4.1667)$ random variable exceeds 125. This is 0.301.

   ii. Given μ, the probability that two further randomly chosen packets of coffee have a mean weight of at least 125 is $\Phi((\mu - 125)/2^{1/2})$. Arguing as above, the required predictive probability is the probability that an $\mathrm{N}(123.93, 2.1667)$ random variable exceeds 125. This is 0.234.

7. (a) The log-likelihood function is given by $l(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} y_i$. Hence $l''(\lambda) = -n/\lambda^2$ and so $\mathbf{E}(-l''(\lambda)) = n/\lambda^2$, giving the required result.

   (b) For the Jeffreys prior and these data, the posterior distribution of λ is given by $\pi(\lambda \,|\, \boldsymbol{y}) \propto \lambda^4 e^{-10\lambda}$, i.e. is $\Gamma(5, 10)$. Hence a 95% equal-tailed credible interval for λ is $(0.162, 1.024)$.