# MAS3301 Bayesian Statistics 

M. Farrow<br>School of Mathematics and Statistics<br>Newcastle University

Semester 2, 2008-9

## 17 Large Samples: Examples

1. The numbers of telephone calls arriving at a busy exchange in 200 short intervals are observed. Suppose that these are $X_{1}, \ldots, X_{n}$, where $n=200$, that $X_{i} \mid \lambda \sim \operatorname{Poisson}(\lambda)$ and that $X_{i}$ and $X_{j}$ are independent, given $\lambda$, when $i \neq j$. The likelihood is therefore

$$
L=\prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_{i}}}{x_{i}!} \propto e^{-n \lambda} \lambda^{s}
$$

where $s=\sum_{i=1}^{n} x_{i}$.

Suppose that the prior for $\lambda$ is a gamma $(a, b)$ distribution. The posterior is then gamma $(s+$ $a, n+b)$. This has mean $(s+a) /(n+b)$ and variance $(s+a) /(n+b)^{2}$ and we can approximate this with a normal distribution

$$
N\left(\frac{s+a}{n+b}, \frac{s+a}{(n+b)^{2}}\right) .
$$

Suppose that $a=2, b=0.2, s=1723$. Then the posterior distribution is approximately $N\left(8.6164,0.2075^{2}\right)$ and an approximate $95 \%$ interval for $\lambda$ is $8.6164 \pm 1.96 \times 0.2075$. This gives

$$
8.210<\lambda<9.023
$$

Alternatively we can use the posterior mode to find the mean of the normal distribution. The posterior density is proportional to

$$
\lambda^{s+a-1} e^{-(n+b) \lambda}
$$

The $\log$ of this is

$$
g(\lambda)=(s+a-1) \log (\lambda)-(n+b) \lambda .
$$

Differentiating wrt $\lambda$ we obtain

$$
g^{\prime}(\lambda)=\frac{s+a-1}{\lambda}-(n+b) .
$$

This gives the posterior mode at $\lambda_{m}=(s+a-1) /(n+b)=8.6114$. Differentiating again we obtain

$$
g^{\prime \prime}(\lambda)=-\frac{s+a-1}{\lambda^{2}}
$$

This gives an approximate variance of

$$
\frac{\lambda_{m}^{2}}{s+a-1}=\frac{s+a-1}{(n+b)^{2}}=0.2075^{2}
$$

So our $95 \%$ interval becomes $8.6114 \pm 1.96 \times 0.2075$. This gives

$$
8.205<\lambda<9.018
$$

Calculating the hpd interval without using the normal approximation (using hpdgamma) gives

$$
8.211<\lambda<9.024
$$

which is very close to the first of our approximations.
2. We observe the result of a sample survey in which respondents are asked a simple "Yes/No" question. The number answering "Yes" is $x$. The sample size $n$ is large and we assume that the effect of the prior distribution is negligible. We also assume that $x$ is an observation from a binomial $(n, \theta)$ distribution where

$$
\theta=\frac{e^{\mu}}{1+e^{\mu}}
$$

That is

$$
\mu=\log \left(\frac{\theta}{1-\theta}\right)
$$

We are interested in the parameter $\mu$. Find an approximation to its posterior distribution.

The likelihood is

$$
L(\mu)=\binom{n}{x} \theta^{x}(1-\theta)^{n-x}
$$

The log likelihood is

$$
l(\mu)=\text { constant }+x \log (\theta)+(n-x) \log (1-\theta)
$$

We need to differentiate this with respect to $\mu$. We note that

$$
\frac{d l}{d \mu}=\frac{d l}{d \theta} \frac{d \theta}{d \mu}
$$

and that

$$
\frac{d \theta}{d \mu}=\frac{\left(1+e^{\mu}\right) e^{\mu}-e^{\mu} e^{\mu}}{\left(1+e^{\mu}\right)^{2}}=\frac{e^{\mu}}{\left(1+e^{\mu}\right)^{2}}=\theta(1-\theta)
$$

So

$$
\begin{aligned}
\frac{d l}{d \theta}=\frac{x}{\theta}-\frac{n-x}{1-\theta} & =\frac{x-\theta x-n \theta+\theta x}{\theta(1-\theta)} \\
& =\frac{x-n \theta}{\theta(1-\theta)} \\
\frac{d l}{d \mu}=\frac{d l}{d \theta} \frac{d \theta}{d \mu} & =x-n \theta \\
\frac{d}{d \theta}\left(\frac{d l}{d \mu}\right) & =-n \\
\frac{d^{2} l}{d \mu^{2}}=\frac{d}{d \theta}\left(\frac{d l}{d \mu}\right) \frac{d \theta}{d \mu} & =-n \theta(1-\theta)
\end{aligned}
$$

Solving $d l / d \mu=0$ we obtain

$$
x=n \hat{\theta} \quad \text { and } \quad \hat{\theta}=\frac{x}{n}
$$

and therefore

$$
\hat{\mu}=\log \left(\frac{x / n}{1-x / n}\right)=\log \left(\frac{x}{n-x}\right)
$$

Substituting this into the second derivative we obtain, at $\theta=\hat{\theta}$,

$$
\frac{d^{2} l}{d \mu^{2}}=-n \hat{\theta}(1-\hat{\theta})<0
$$

Therefore $\hat{\mu}$ is a maximum and is approximately the posterior mean and the posterior variance is approximately

$$
V=[n \hat{\theta}(1-\hat{\theta})]^{-1}
$$

The posterior distribution of $\mu$ is approximately $N(\hat{\mu}, V)$.
Suppose that $n=2000$ and $x=1456$. Then $\hat{\theta}=\overline{x / n=0.728}$ and

$$
\hat{\mu}=\log \left(\frac{1456}{2000-1456}\right)=0.9845
$$

The approximate variance is

$$
V=[2000 \times 0.728(1-0.728)]^{-1}=[396.032]^{-1}=0.002525
$$

giving a standard deviation of 0.05025 . An approximate $95 \%$ posterior hpd interval for $\mu$ is therefore $0.9845 \pm 1.96 \times 0.05025$. that is

$$
0.8860<\mu<1.0830
$$

## 18 Predictive Distributions

### 18.1 Introduction

In Lecture 3 we introduced parameters as a means of structuring our beliefs and of transferring information from observations to beliefs about future observations. Since then we have spent some time looking at how information from observations affects our beliefs about parameters. Information goes from observations to our beliefs about the values of parameters via the likelihood. In this lecture we will return to thinking about our beliefs for the values of future observations.

Suppose that our beliefs about a (possibly vector) parameter $\theta$ are represented by the distribution with density function $f_{\theta}(\theta)$. (We assume here that $\theta$ is continuous. It could be discrete. In that case we replace the density function by a probability function and the integration by a summation). The distribution of a future observation $Y$ given $\theta$ has density function, if $Y$ is continuous, or probability function, if $Y$ is discrete, $f_{Y \mid \theta}(y \mid \theta)$. For convenience we will assume for now that $Y$ is continuous.

The joint density of $\theta$ and $Y$ is therefore

$$
f_{\theta}(\theta) f_{Y \mid \theta}(y \mid \theta)
$$

To find the marginal density of $Y$ we simply integrate out $\theta$.

$$
f_{Y}(y)=\int_{-\infty}^{\infty} f_{\theta}(\theta) f_{Y \mid \theta}(y \mid \theta) d \theta
$$

This marginal distribution is called a predictive distribution of $Y$. Its mean is the predictive mean and so on.

Very often, of course, we are interested in the case where we observe some data and then make predictions about future observations which we have not yet seen. In this case, before we have seen the data, we can evaluate the prior predictive distribution:

$$
f_{Y}^{(0)}(y)=\int_{-\infty}^{\infty} f_{\theta}^{(0)}(\theta) f_{Y \mid \theta}(y \mid \theta) d \theta
$$

where $f_{\theta}^{(0)}(\theta)$ is the prior density of $\theta$. After we have seen data $D$, we can evaluate the posterior predictive distribution:

$$
f_{Y}^{(1)}(y \mid D)=\int_{-\infty}^{\infty} f_{\theta \mid D}^{(1)}(\theta \mid D) f_{Y \mid \theta}(y \mid \theta) d \theta
$$

where $f_{\theta \mid D}^{(1)}(\theta \mid D)$ is the posterior density of $\theta$ given the data $D$.
Predictive distributions reflect both the aleatory uncertainty in the future observations and the epistemic uncertainty in the parameters. All (remaining) uncertainty is properly reflected. This is an important difference from non-Bayesian statistics. The prior predictive distribution describes our beliefs before we have seen data and the posterior predictive distribution describes our beliefs afterwards.

Predictive distributions are often used in model checking (or model criticism) where we examine whether there is evidence that we made invalid assumptions by comparing observations with their predictive distributions.

### 18.2 Example: Chester Road

In section 4.4 we looked at the rate $\lambda$ of vehicles arriving along a road (in vehicles per second). Our posterior distribution for $\lambda$ was a gamma $(119,1220)$ with pdf

$$
f_{\lambda}^{(1)}(\lambda)=\frac{1220^{119}}{118!} \lambda^{118} e^{-1220 \lambda}
$$

1. What is the posterior probability now of observing $j$ vehicles in the next $t$ seconds?

Given the value of $\lambda$, the distribution of the number of vehicles in the next $t$ seconds would be Poisson $(\lambda t)$.
The joint probability (density) of $\lambda$ and $j$ is

$$
\frac{1220^{119}}{118!} \lambda^{118} e^{-1220 \lambda} \frac{\lambda^{j} t^{j} e^{-\lambda t}}{j!}=\frac{1220^{119}}{118!} \frac{t^{j}}{j!} \lambda^{118+j} e^{-(1220+t) \lambda}
$$

We integrate out $\lambda$ to get the marginal probability of $j$ vehicles. By comparing the function with a gamma p.d.f. we see that

$$
\begin{aligned}
\int_{0}^{\infty} \lambda^{118+j} e^{-(1220+t) \lambda} \cdot d \lambda & =\frac{\Gamma(119+j)}{(1220+t)^{119+j}} \\
& =\frac{(118+j)!}{(1220+t)^{119+j}}
\end{aligned}
$$

Hence the probability of observing $j$ vehicles in the next $t$ seconds is

$$
\frac{(118+j)!}{118!j!} \frac{1220^{119} t^{j}}{(1220+t)^{119+j}}=\binom{118+j}{j}\left(\frac{1220}{1220+t}\right)^{119}\left(\frac{t}{1220+t}\right)^{j}
$$

This represents a negative binomial distribution. For large $n$ and $\tau$ this would be approximately a Poisson distribution with mean equal to $t$ times the posterior mean for $\lambda$.
2. What is the posterior p.d.f. now for the time $T$ to the next arrival?

Given the value of $\lambda$, the waiting time has an exponential $(\lambda)$ distribution.
The joint p.d.f. of $\lambda$ and $T$ is

$$
\frac{1220^{119}}{118!} \lambda^{118} e^{-1220 \lambda} \lambda e^{-\lambda t}=\frac{1220^{119}}{118!} \lambda^{119} e^{-(1220+t) \lambda}
$$

We integrate out $\lambda$ to get the marginal p.d.f. for $T$. By analogy with the gamma p.d.f.,

$$
\int_{0}^{\infty} \lambda^{119} e^{-(1220+t) \lambda} \cdot d \lambda=\frac{119!}{(1220+t)^{120}}
$$

Hence the predictive p.d.f. for the waiting time is

$$
\frac{119}{1220}\left(\frac{1220}{1220+t}\right)^{120}
$$

which would be approximately a negative exponential distribution for large $n$.

### 18.3 Example: beta-binomial

This is an example where the observation model is a discrete distribution.

- Observation model: $X \mid \theta \sim \operatorname{bin}(n, \theta)$.
- Beliefs about $\theta: \theta \sim \operatorname{beta}(a, b)$.
- Predictive distribution: $\operatorname{Pr}(X=j)$ for $j=0,1, \ldots, n$.

$$
\begin{aligned}
\operatorname{Pr}(X=j)= & \int_{0}^{1} \frac{\Gamma(a+b)}{\Gamma(a) \Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}\binom{n}{j} \theta^{j}(1-\theta)^{n-j} d \theta \\
= & \int_{0}^{1} \frac{\Gamma(a+b)}{\Gamma(a) \Gamma(b)}\binom{n}{j} \theta^{a+j-1}(1-\theta)^{b+n-j-1} d \theta \\
= & \frac{\Gamma(a+b)}{\Gamma(a) \Gamma(b)}\binom{n}{j} \frac{\Gamma(a+j) \Gamma(b+n-j)}{\Gamma(a+b+n)} \\
& \times \int_{0}^{1} \frac{\Gamma(a+b+n)}{\Gamma(a+j) \Gamma(b+n-j)} \theta^{a+j-1}(1-\theta)^{b+n-j-1} d \theta \\
= & \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+n-j)}{\Gamma(b)} \frac{\Gamma(a+b)}{\Gamma(a+b+n)}\binom{n}{j}
\end{aligned}
$$

E.g. $a=2, b=2, n=7, j=3$.

$$
\begin{aligned}
\operatorname{Pr}(X=3) & =\frac{(5-1)!}{(2-1)!} \frac{(6-1)!}{(2-1)!} \frac{(4-1)!}{(11-1)!} \frac{7!}{3!4!} \\
& =\frac{4!5!3!7!}{1!1!10!3!4!}=\frac{5!7!}{10!}=0.16667
\end{aligned}
$$

### 18.4 Example: normal, known precision

- Observation model: $Y \mid \mu \sim N\left(\mu, \tau^{-1}\right)$.
- Beliefs about $\mu$ :

$$
\mu \sim N\left(m, P^{-1}\right)
$$

Let $c=P / \tau$ so $P=c \tau$. Then

$$
\mu \sim N\left(m,[c \tau]^{-1}\right)
$$

- Predictive distribution has density $f_{Y}(y)$.

The joint density of $\mu, Y$ is proportional to

$$
\exp \left\{-\frac{c \tau}{2}(\mu-m)^{2}\right\} \tau^{1 / 2} \exp \left\{-\frac{\tau}{2}(y-\mu)^{2}\right\} .
$$

We can write $Y=\mu+\varepsilon$ where $\varepsilon \sim N\left(0, \tau^{-1}\right)$ independently of $\mu$. So

$$
Y \sim N\left(m,[c \tau]^{-1}+\tau^{-1}\right) .
$$

That is

$$
Y \sim N\left(m,\left[c_{p} \tau\right]^{-1}\right)
$$

where

$$
c_{p}=\frac{c}{c+1}
$$

and the variance is

$$
\left[c_{p} \tau\right]^{-1}=\frac{c+1}{c \tau}=\tau^{-1}+[c \tau]^{-1} .
$$

E.g. $m=23.0, \tau=0.05, P=0.25$ so $c=5$,

$$
c_{p}=\frac{5}{6}, c_{p} \tau=\frac{0.25}{6}, \operatorname{var}(Y)=\frac{6}{0.25}=24.0 .
$$

Hence

$$
\operatorname{Pr}(Y<20)=\Phi\left(\frac{20-23}{\sqrt{24}}\right)=\Phi(-0.612)=0.270
$$

### 18.5 Example: normal, conjugate prior

- Observation model: $Y \mid \mu, \tau \sim N\left(\mu, \tau^{-1}\right)$.
- Beliefs about $\mu, \tau$ :

$$
\begin{aligned}
\tau & \sim \operatorname{gamma}(d / 2, d v / 2) \\
\mu \mid \tau & \sim N\left(m,(c \tau)^{-1}\right)
\end{aligned}
$$

- Predictive distribution has density $f_{Y}(y)$.

The joint density of $\tau, \mu, Y$ is proportional to

$$
\tau^{d / 2-1} e^{-d v \tau / 2}(c \tau)^{1 / 2} \exp \left\{-\frac{c \tau}{2}(\mu-m)^{2}\right\} \tau^{1 / 2} \exp \left\{-\frac{\tau}{2}(y-\mu)^{2}\right\}
$$

Conditional on $\tau$, we can write $Y=\mu+\varepsilon$ where $\varepsilon \sim N\left(0, \tau^{-1}\right)$ independently of $\mu$. So

$$
Y \mid \tau \sim N\left(m,[c \tau]^{-1}+\tau^{-1}\right)
$$

That is

$$
Y \mid \tau \sim N\left(m,\left[c_{p} \tau\right]^{-1}\right)
$$

where

$$
c_{p}=\frac{c}{c+1}
$$

and the variance is

$$
\left[c_{p} \tau\right]^{-1}=\frac{c+1}{c \tau}=\tau^{-1}+[c \tau]^{-1}
$$

So,

$$
\frac{Y-m}{\sqrt{v / c_{p}}} \sim t_{d}
$$

E.g. $d=4, v=0.25, m=15, c=0.5$ so

$$
\begin{gathered}
c_{p}=\frac{0.5}{1.5}=\frac{1}{3}, \quad \frac{v}{c_{p}}=0.75, \\
\frac{Y-15}{\sqrt{0.75}} \sim t_{4} .
\end{gathered}
$$

$95 \%$ predictive (hpd) interval:

$$
15 \pm 2.776 \sqrt{0.75}
$$

That is

$$
12.596<Y<17.404
$$

### 18.6 Problems 5

1. (Some of this question is also in Problems 4). I recorded the attendance of students at tutorials for a module. Suppose that we can, in some sense, regard the students as a sample from some population of students so that, for example, we can learn about the likely behaviour of next year's students by observing this year's. At the time I recorded the data we had had tutorials in Week 2 and Week 4. Let the probability that a student attends in both weeks be $\theta_{11}$, the probability that a student attends in week 2 but not Week 4 be $\theta_{10}$ and so on. The data are as follows.

| Attendance | Probability | Observed frequency |
| :--- | :---: | ---: |
| Week 2 and Week 4 | $\theta_{11}$ | $n_{11}=25$ |
| Week 2 but not Week 4 | $\theta_{10}$ | $n_{10}=7$ |
| Week 4 but not Week 2 | $\theta_{01}$ | $n_{01}=6$ |
| Neither week | $\theta_{00}$ | $n_{00}=13$ |

Suppose that the prior distribution for $\left(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}\right)$ is a Dirichlet distribution with density proportional to

$$
\theta_{11}^{3} \theta_{10} \theta_{01} \theta_{00}^{2} .
$$

(a) Find the prior means and prior variances of $\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}$.
(b) Find the posterior distribution.
(c) Find the posterior means and posterior variances of $\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}$.
(d) Using the R function hpdbeta which may be obtained from the Web page (or otherwise), find a $95 \%$ posterior hpd interval, based on the exact posterior distribution, for $\theta_{00}$.
(e) Find an approximate $95 \%$ hpd interval for $\theta_{00}$ using a normal approximation based on the posterior mode and the partial second derivatives of the log posterior density.
Compare this with the exact hpd interval.
Hint: To find the posterior mode you will need to introduce a Lagrange multiplier.
(f) The population mean number of attendances out of two is $\mu=2 \theta_{11}+\theta_{10}+\theta_{01}$. Find the posterior mean of $\mu$ and an approximation to the posterior standard deviation of $\mu$.
2. Samples are taken from twenty wagonloads of an industrial mineral and analysed. The amounts in ppm (parts per million) of an impurity are found to be as follows.
$44.350 .251 .749 .450 .655 .053 .548 .648 .8 \quad 53.3$
59.451 .452 .051 .951 .648 .349 .354 .152 .453 .1

We regard these as independent samples from a normal distribution with mean $\mu$ and variance $\sigma^{2}=\tau^{-1}$.
Find a $95 \%$ posterior hpd interval for $\mu$ under each of the following two conditions.
(a) The value of $\tau$ is known to be 0.1 and our prior distribution for $\mu$ is normal with mean 60.0 and standard deviation 20.0.
(b) The value of $\tau$ is unknown. Our prior distribution for $\tau$ is a gamma distribution with mean 0.1 and standard deviation 0.05 . Our conditional prior distribution for $\mu$ given $\tau$ is normal with mean 60.0 and precision $0.025 \tau$ (that is, standard deviation $\sqrt{40} \tau^{-1 / 2}$ ).
3. We observe a sample of 30 observations from a normal distribution with mean $\mu$ and precision $\tau$. The data, $y_{1}, \ldots, y_{30}$, are such that

$$
\sum_{i=1}^{30} y_{i}=672 \quad \text { and } \quad \sum_{i=1}^{30} y_{i}^{2}=16193
$$

(a) Suppose that the value of $\tau$ is known to be 0.04 and that our prior distribution for $\mu$ is normal with mean 20 and variance 100 . Find the posterior distribution of $\mu$ and evaluate a posterior $95 \% \mathrm{hpd}$ interval for $\mu$.
(b) Suppose that we have a gamma $(1,10)$ prior distribution for $\tau$ and our conditional prior distribution for $\mu$ given $\tau$ is normal with mean 20 and variance $(0.1 \tau)^{-1}$. Find the marginal posterior distribution for $\tau$, the marginal posterior distribution for $\mu$ and the marginal posterior $95 \% \mathrm{hpd}$ interval for $\mu$.
4. The following data come from the experiment reported by MacGregor et al. (1979). They give the supine systolic blood pressures ( mm Hg ) for fifteen patients with moderate essential hypertension. The measurements were taken immediately before and two hours after taking a drug.

| Patient | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Before | 210 | 169 | 187 | 160 | 167 | 176 | 185 | 206 |
| After | 201 | 165 | 166 | 157 | 147 | 145 | 168 | 180 |
| Patient | 9 | 10 | 11 | 12 | 13 | 14 | 15 |  |
| Before | 173 | 146 | 174 | 201 | 198 | 148 | 154 |  |
| After | 147 | 136 | 151 | 168 | 179 | 129 | 131 |  |

We are interested in the effect of the drug on blood pressure. We assume that, given parameters $\mu, \tau$, the changes in blood pressure, from before to after, in the $n$ patients are independent and normally distributed with unknown mean $\mu$ and unknown precision $\tau$. The fifteen differences are as follows.

$$
\begin{array}{lllllllllllllll}
-9 & -4 & -21 & -3 & -20 & -31 & -17 & -26 & -26 & -10 & -23 & -33 & -19 & -19 & -23
\end{array}
$$

Our prior distribution for $\tau$ is a gamma $(0.35,1.01)$ distribution. Our conditional prior distribution for $\mu$ given $\tau$ is a normal $N\left(0,[0.003 \tau]^{-1}\right)$ distribution.
(a) Find the marginal posterior distribution of $\tau$.
(b) Find the marginal posterior distribution of $\mu$.
(c) Find the marginal posterior $95 \% \mathrm{hpd}$ interval for $\mu$.
(d) Comment on what you can conclude about the effect of the drug.
5. The lifetimes of certain components are supposed to follow a Weibull distribution with known shape parameter $\alpha=2$. The probability density function of the lifetime distribution is

$$
f(t)=\alpha \rho^{2} t \exp \left[-(\rho t)^{2}\right]
$$

for $0<t<\infty$.
We will observe a sample of $n$ such lifetimes where $n$ is large.
(a) Assuming that the prior density is nonzero and reasonably flat so that it may be disregarded, find an approximation to the posterior distribution of $\rho$. Find an approximate $95 \%$ hpd interval for $\rho$ when $n=300, \sum \log (t)=1305.165$ and $\sum t^{2}=3161776$.
(b) Assuming that the prior distribution is a gamma $(a, b)$ distribution, find an approximate $95 \% \mathrm{hpd}$ interval for $\rho$, taking into account this prior, when $a=2, b=100, n=300$, $\sum \log (t)=1305.165$ and $\sum t^{2}=3161776$.
6. Given the value of $\lambda$, the number $X_{i}$ of transactions made by customer $i$ at an online store in a year has a Poisson $(\lambda)$ distribution, with $X_{i}$ independent of $X_{j}$ for $i \neq j$. The value of $\lambda$ is unknown. Our prior distribution for $\lambda$ is a gamma $(5,1)$ distribution.

We observe the numbers of transactions in a year for 45 customers and

$$
\sum_{i=1}^{45} x_{i}=182
$$

(a) Using a $\chi^{2}$ table (i.e. without a computer) find the lower $2.5 \%$ point and the upper $2.5 \%$ point of the prior distribution of $\lambda$.
(These bound a $95 \%$ symmetric prior credible interval).
(b) Find the posterior distribution of $\lambda$.
(c) Using a normal approximation to the posterior distribution, based on the posterior mean and variance, find a $95 \%$ symmetric posterior credible interval for $\lambda$.
(d) Find an expression for the posterior predictive probability that a customer makes $m$ transactions in a year.
(e) As well as these "ordinary customers," we believe that there is a second group of individuals. The number of transactions in a year for a member of this second group has, given $\theta$, a Poisson $(\theta)$ distribution and our beliefs about the value of $\theta$ are represented by a gamma $(1,0.05)$ distribution.
A new individual is observed who makes 10 transactions in a year. Given that our prior probability that this is an ordinary customer is 0.9 , find our posterior probability that this is an ordinary customer.
Hint: You may find it best to calculate the logarithms of the predictive probabilities before exponentiating these. For this you might find the R function lgamma useful. It calculates the log of the gamma function. Alternatively it is possible to do the calculation using the R function dnbinom.
(N.B. In reality a slightly more complicated model is used in this type of application).
7. The following data give the heights in cm of 25 ten-year-old children. We assume that, given the values of $\mu$ and $\tau$, these are independent observations from a normal distribution with mean $\mu$ and variance $\tau^{-1}$.

| 66 | 66 | 69 | 61 | 58 | 53 | 78 | 71 | 49 | 57 | 54 | 61 | 49 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 64 | 63 | 60 | 53 | 51 | 65 | 70 | 55 | 55 | 74 | 70 | 42 |  |

(a) Assuming that the value of $\tau^{-1}$ is known to be 64 and our prior distribution for $\mu$ is normal with mean 55 and standard deviation 5 , find a $95 \% \mathrm{hpd}$ interval for the height in cm of another ten-year-old child drawn from the same population.
(b) Assume now that the value of $\tau$ is unknown but we have a prior distribution for it which is a gamma $(2,128)$ distribution and our conditional prior distribution for $\mu$ given $\tau$ is normal with mean 55 and variance $(2.56 \tau)^{-1}$. Find a $95 \%$ hpd interval for the height in cm of another ten-year-old child drawn from the same population.
8. A random sample of $n=1000$ people was chosen from a large population. Each person was asked whether they approved of a proposed new law. The number answering "Yes" was $x=372$. (For the purpose of this exercise all other responses and non-responses are teated
as simply "Not Yes"). Assume that $x$ is an observation from the $\operatorname{binomial}(n, p)$ distribution where $p$ is the unknown proportion of people in the population who would answer "Yes."
Our prior distribution for $p$ is a uniform distribution on $(0,1)$.
Let $p=\Phi(\theta)$ so $\theta=\Phi^{-1}(p)$ where $\Phi(y)$ is the standard normal distribution function and $\Phi^{-1}(z)$ is its inverse.
(a) Find the maximum likelihood estimate of $p$ and hence find the maximum likelihood estimate of $\theta$.
(b) Disregarding the prior distribution, find a large-sample approximation to the posterior distribution of $\theta$.
(c) Using your approximate posterior distribution for $\theta$, find an approximate $95 \% \mathrm{hpd}$ interval for $\theta$.
(d) Use the exact posterior distribution for $p$ to find the actual posterior probability that $\theta$ is inside your approximate hpd interval.

Notes: - The standard normal distribution function $\Phi(x)=\int_{-\infty}^{x} \phi(u) d u$ where $\phi(u)=$ $(2 \pi)^{-1 / 2} \exp \left\{-u^{2} / 2\right\}$.

- Let $l$ be the log-likelihood. Then

$$
\frac{d l}{d \theta}=\frac{d l}{d p} \frac{d p}{d \theta}
$$

and

$$
\begin{aligned}
\frac{d^{2} l}{d \theta^{2}}=\frac{d}{d \theta}\left\{\frac{d l}{d p} \frac{d p}{d \theta}\right\} & =\frac{d}{d \theta}\left\{\frac{d l}{d p}\right\} \frac{d p}{d \theta}+\frac{d l}{d p} \frac{d^{2} p}{d \theta^{2}} \\
& =\frac{d^{2} l}{d p^{2}}\left(\frac{d p}{d \theta}\right)^{2}+\frac{d l}{d p} \frac{d^{2} p}{d \theta^{2}}
\end{aligned}
$$

- Derivatives of $p$ :

$$
\begin{aligned}
\frac{d p}{d \theta} & =\phi(\theta) \\
\frac{d^{2} p}{d \theta^{2}} & =-\theta \phi(\theta)
\end{aligned}
$$

- You can evaluate $\Phi^{-1}(u)$ using R with
qnorm ( $u, 0,1$ )
and $\phi(u)$ is given by
dnorm (u, 0,1 )

9. The amounts of rice, by weight, in 20 nominally 500 g packets are determined. The weights, in g , are as follows.

| 496 | 506 | 495 | 491 | 488 | 492 | 482 | 495 | 493 | 496 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 487 | 490 | 493 | 495 | 492 | 498 | 491 | 493 | 495 | 489 |

Assume that, given the values of parameters $\mu, \tau$, the weights are independent and each has a normal $N(\mu, \tau)$ distribution.
The values of $\mu$ and $\tau$ are unknown. Our prior distribution is as follows. We have a gamma $(2,9)$ prior distribution for $\tau$ and a $N\left(500,(0.005 \tau)^{-1}\right)$ conditional prior distribution for $\mu$ given $\tau$.
(a) Find the posterior probability that $\mu<495$.
(b) Find the posterior predictive probability that a new packet of rice will contain less than 500 g of rice
10. A machine which is used in a manufacturing process jams from time to time. It is thought that the frequency of jams might change over time as the machine becomes older. Once every three months the number of jams in a day is counted. The results are as follows.

| Observation $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
| Age of machine $t_{i}$ (months) | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 |
| Number of jams $y_{i}$ | 10 | 13 | 24 | 17 | 20 | 22 | 20 | 23 |

Our model is as follows. Given the values of two parameters $\alpha, \beta$, the number of jams $y_{i}$ on a dat when the machine has age $t_{i}$ months has a Poisson distribution

$$
y_{i} \sim \operatorname{Poisson}\left(\lambda_{i}\right)
$$

where

$$
\log _{e}\left(\lambda_{i}\right)=\alpha+\beta t_{i}
$$

Assume that the effect of our prior distribution on the posterior distribution is negligible and that large-sample approximations may be used.
(a) Let the values of $\alpha$ and $\beta$ which maximise the likelihood be $\hat{\alpha}$ and $\hat{\beta}$. Assuming that the likelihood is differentiable at its maximum, show that these satisfy the following two equations

$$
\begin{gathered}
\sum_{i=1}^{8}\left(\hat{\lambda}_{i}-y_{i}\right)=0 \\
\sum_{i=1}^{8} t_{i}\left(\hat{\lambda}_{i}-y_{i}\right)=0
\end{gathered}
$$

where

$$
\log _{e}\left(\hat{\lambda}_{i}\right)=\hat{\alpha}+\hat{\beta} t_{i}
$$

and show that these equations are satisfied (to a good approximation) by

$$
\hat{\alpha}=2.552 \quad \text { and } \quad \hat{\beta}=0.02638
$$

(You may use R to help with the calculations, but show your commands).
You may assume from now on that these values maximise the likelihood.
(b) Find an approximate symmetric $95 \%$ posterior interval for $\alpha+24 \beta$.
(c) Find an approximate symmetric $95 \%$ posterior interval for $\exp (\alpha+24 \beta)$, the mean jam-rate per day at age 24 months.
(You may use R to help with the calculations, but show your commands).

## Homework 5

Solutions to Questions 9 and 10 of Problems 5 are to be submitted in the Homework Letterbox no later than 4.00pm on Tuesday May 5th.

## Reference

MacGregor, G.A., Markandu, N.D., Roulston, J.E. and Jones, J.C., 1979. Essential hypertension: effect of an oral inhibitor of angiotensin-converting enzyme. British Medical Journal, 2, 1106-1109.

