MA322: Advanced Statistical Methods – Solutions Work Sheet 1

1. Consider a disease that is thought to occur in 1% of the population. Using a particular blood test a physician observes that out of the patients with disease 98% possess a particular symptom. Also assume that $100\,a\%$ of the population without the disease have the same symptom. A randomly chosen person from the population is blood tested and is shown to have the symptom. What is the conditional probability that the person has the disease? Plot this probability as a function of a where $0.001 < a < 0.5$. Use 50 equally spaced points.

   Solution:

   Here $k = 2$ in the Bayes theorem and let B_1 be the event that a randomly chosen person has the disease and B_2 is the complement of B_1. Let A be the event that a randomly chosen person has the symptom. The problem is to determine $Pr\{B_1|A\}$.

   We have $Pr(B_1) = 0.01$ since 1% of the population has the disease, and $Pr(A|B_1) = 0.98$. Also $Pr(B_2) = 0.99$ and $Pr(A|B_2) = a$. Now

$$\begin{aligned} Pr(\text{disease} \mid \text{symptom}) = Pr(B_1|A) &= \frac{Pr(A|B_1)\,Pr(B_1)}{Pr(A|B_1)\,Pr(B_1)+Pr(A|B_2)\,Pr(B_2)} \\ &= \frac{0.98\times 0.01}{0.98\times 0.01 + a\times 0.99} \end{aligned}$$

   We issue the S-Plus commands:

```
a <- seq(from=0.0001, to=0.5, length=50)
p <- 0.98 * 0.01 /(0.98 * 0.01 + a * 0.99)
plot(a, p)
```

□

2. Suppose that the number of defects on a roll of magnetic recording tape has a Poisson distribution for which the mean θ is unknown and that the prior distribution of θ is a gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. When five rolls of this tape are selected at random and inspected, the number of defects found on the rolls are 2, 2, 6, 0, and 3. Derive the posterior distribution. If the squared error loss function is used what is the Bayes estimate of θ? What is the Bayes estimate under the absolute error loss? What is the posterior probability of $\theta < 2$?

   Solution: We have $X_1, \ldots, X_n$ follows the Poisson distribution with mean θ. Therefore we have the likelihood:

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} \\ &\propto e^{-n\theta}\,\theta^{\sum_{i=1}^{n} x_i}. \end{aligned}$$

   The prior distribution is

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}\,e^{-\beta\theta}, \quad \theta > 0.$$

The posterior distribution is proportional to likelihood times the prior. Thus

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto e^{-n\theta}\,\theta^{\sum_{i=1}^n x_i}\,\theta^{\alpha-1}e^{-\beta\theta}\\ &= \theta^{\sum_{i=1}^n x_i+\alpha-1}\,e^{-(n+\beta)\theta} \quad \theta > 0.\end{aligned}$$

Thus the posterior distribution is seen to be a gamma distribution with parameters $\sum_{i=1}^n x_i+\alpha$ and $n+\beta$. For the given values the parameters are 16 and 6, respectively.

We issue the following two S-Plus commands.

```
qgamma(0.5, 16, 6)
pgamma(2, 16, 6)
```

□

3. Suppose that a child is given an IQ test. Assume that the test result X is $N(\theta, 100)$ where θ is the true IQ level of the child, as measured by the test. Assume also that, in the population as a whole, θ is distributed according to a $N(100, 225)$ distribution. Show that the posterior distribution is normal with mean $(400+9x)/13$ and variance 69.23. Thus if a child scores 115 on the test, show that his true IQ θ has a $N(110.39, 69.23)$ distribution. What is the posterior probability that $\theta < 120$? Obtain the 95% highest posterior density credible interval for θ.

   Solution:

   We know (from notes) that the posterior distribution is normal with mean

$$\frac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2},$$

   and variance

$$\frac{1}{n/\sigma^2 + 1/\tau^2}.$$

   Here $n = 1, \sigma^2 = 100, \mu = 100$ and $\tau^2 = 225$. Just putting these numbers back into the above two formulae yield that the posterior is normal with mean $(400 + 9x)/13$ and variance 69.23.

   To calcluate the posterior probability of $\theta < 120$ and obtain the highest posterior density credible interval we issue the S-Plus commands:

```
pnorm(120, 110.39, sqrt(69.23) )
c(110.39 - sqrt(69.23) * 1.96, 110.39 +sqrt(69.23) * 1.96)
```

□

4. An accountant makes $x = 3$ errors out of $n = 100$ entries and wishes to draw an inference regarding her underlying error rate θ when expressed as a proportion. Assume the beta prior distribution with parameters α and β. Derive the posterior distribution. Draw the posterior density on the same graphsheet for $\alpha = \beta = 1/2$ and $\alpha = \beta = 1$. Although θ ranges from 0 to

1, the posterior is effectively contained in the interval from 0 to 0.15. Find equal tailed 95% Bayesian credible sets for θ under the above two prior distributions.

Solution:

Here X follows the binomial distribution with parameters n and θ. So the likelihood is:

$$f(x|\theta) \propto \theta^x (1-\theta)^{n-x}.$$

We have the prior distribution

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}, 0 < \theta < 1.$$

So the posterior is

$$\pi(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}, 0 < \theta < 1..$$

which is beta distribution with parameters $x+\alpha$ and $n-x+\beta$.

Now we use the following S-Plus function to solve the problem.

```
prob1.4 <- function(tmax=0.15) {
  theta <- seq(from=0.0, to=tmax, length=100)
  y1 <- dbeta(theta, 3+0.5, 100-3+0.5)
  y2 <- dbeta(theta, 3+1.0, 100-3+1.0)

  yr <- range(c(y1, y2))
  plot(theta, y1, ylim=yr, xlab="theta", ylab="posterior density", type="l")
  lines(theta, y2, lty=2)

  con1 <- c(qbeta(0.025, 3+0.5, 100-3+0.5),
                qbeta(0.975, 3+0.5, 100-3+0.5) )
  con2 <- c(qbeta(0.025, 3+1.0, 100-3+1.0),
                qbeta(0.975, 3+1.0, 100-3+1.0) )

  round(cbind(con1, con2), 4)

}
```

□