Stat 544 – Spring 2005
Homework Assignment 2 - Answer key

Problem 1

a - Jeffrey's prior of $\theta = (\mu, \sigma^2)$

$$I(\theta) = \left[\begin{array}{cc} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{array} \right] \qquad \Rightarrow \qquad |I(\theta)|^{1/2} = \frac{n}{\sqrt{2}\sigma^3}$$

Thus, $p(\mu, \sigma^2) \propto \sigma^{-3}$

b - For X = # of failures before r successes are observed

$$p(X|\theta) \propto \theta^r(1-\theta)^x \Rightarrow \log(p(\theta|x) \propto r\log(\theta) + x\log(1-\theta)$$

$$\Rightarrow \frac{\partial}{\partial\theta}\log p(x|\theta) = \frac{r}{\theta} - \frac{x}{1-\theta} \Rightarrow \frac{\partial^2}{\partial^2\theta}\log p(x|\theta) = -\frac{r}{\theta^2} - \frac{x}{(1-\theta)^2}$$

$$I(\theta) = \frac{1}{\theta^2(1-\theta)} \Rightarrow p(\theta) \propto \frac{1}{\theta\sqrt{1-\theta}}$$

c - Recall that the *likelihood principle* states that for a given sample of data, any two probability models $p(\theta|y)$ that have the same likelihood function yield the same inference for θ.

Now, assume that we have observed n trials, on y of those n trials we observed "success", i.e. we observed $x = n - y$ failures. Further, assume that we do not have information as to whether the experiment followed design A or design B, where the designs are as follows

A = to observe n trials and then count on how many of those trials we observed "success".

B = to observe y"successes" and then count how many trials were needed to achieve those y successes.

Let θ be the probability of observing a success. After we have our sample (we know that we observed n trials and y of those where successes) the information regarding which design we followed is superfluous given that the likelihood under design A is proportional to the likelihood under design B. Thus, any posterior inference about θ should be the same independently of the design used to collect the sample. This is not the case here since

- Design A

$$p(\theta|y) \propto \theta^{y-.5}(1-\theta)^{n-y-.5} \Rightarrow \theta|y \sim \text{Beta}(y+.5, n-y+.5)$$

- Design B

$$p(\theta|y) \propto \theta^{y-1}(1-\theta)^{n-y-.5} \Rightarrow \theta|y \sim \text{Beta}(y, n-y+.5)$$

Problem 2

The "Speed of light" code can be use with minimal modifications to obtain the numerical values to solve this problem. All theoretical results can be found on the textbook. To obtain the exact results for part (e) you can use the following R code:

```
y<-scan("c:/hw/scores.dat")
y.mean <- mean(y)
y.sd <- sd(y)
n <- length(y)

prob.mu.gt.10 <- pt((10-y.mean)*sqrt(n)/y.sd,n-1,lower.tail = FALSE)
prob.mu.gt.6 <- pt((6-y.mean)*sqrt(n)/y.sd,n-1,lower.tail = FALSE)
```

$$P(\mu \geq 6|y) = 0.01401350 \qquad P(\mu \geq 10|y) = 1.089205e-09$$

Problem 3

a - The constraint $\theta_1 > \theta_2$ can be included on the prior distribution of (θ_1, θ_2), i.e.

$$p(\theta_1, \theta_2) \propto I(\theta_1 > \theta_2)$$

where $I(A)$ is the indicator function of the set A. Thus,

$$p(\theta_1, \theta_2|y_1, y_2) = \frac{(2\pi)^{-1}\exp\{-.5\sum_{i=1}^2(y_i-\theta_1)^2\}I[\theta_1 > \theta_2]}{\int_{R^2}(2\pi)^{-1}\exp\{-.5\sum_{i=1}^2(y_i-\theta_1)^2\}I[\theta_1 > \theta_2]d\theta_1 \times \theta_2}$$

For a fixed (y_1, y_2) the denominator is equal to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}\exp\left\{-\sum_{i=1}^2\frac{(y_i-\theta_i)^2}{2}\right\}I[\theta_1 > \theta_2]\,d\theta_2\,d\theta_1$$

$$= \int_{-\infty}^{\infty}\underbrace{\frac{\exp(-.5(y_1-\theta_1)^2)}{\sqrt{2\pi}}}_{\phi(\theta_1-y_1)}\underbrace{\int_{-\infty}^{\theta_1}\frac{\exp(-.5(y_2-\theta_2)^2)}{\sqrt{2\pi}}d\theta_2}_{\Phi(\theta_1-y_2)}\,d\theta_1$$

$$= \int_{-\infty}^{\infty} \phi(\theta_1 - y_1)\Phi(\theta_1 - y_2)d\theta_1 \qquad (1)$$

Note that for a fixed (y_1, y_2), equation (1) is the probability that $\theta_2 < \theta_1$ when $(\theta_1, \theta_2)'$ are distributed bivariate normal with vector of means equal to (y_1, y_2) and variance-covariance matrix equal to the 2 by 2 identity. Hence, the posterior distribution of $(\theta_1, \theta_2 | y_1, y_2)$ is a truncated normal.

b- This is a straightforward consequence of part a.

c-

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

Then with

$$A := \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

We have the well known result

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = A\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(A\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, A'\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}A\right)$$

Thus, with $\mu_1 := \theta_1 - \theta_2$ and $\mu_2 := \theta_1 + \theta_2$, and by noting that $\theta_1 > \theta_2 \Rightarrow \mu_1 > 0$ we have

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\right)$$

subject to the constraint $\mu_1 > 0$, i.e.

$$p(x_1, x_2 | \mu_1, \mu_2) = \frac{1}{4\pi}\exp\left\{-\frac{1}{2}\sum_{i=1}^{2}\frac{(x_i - \mu_i)^2}{2}\right\} I[\mu_1 > 0]$$

Further, note that $p(\theta_1, \theta_2) \propto 1 \Rightarrow p(\mu_1, \mu_2) \propto 1$, we get

$$p(\mu_1, \mu_2 | x_1, x_2) = \frac{1}{4\pi}\exp\left\{-\frac{1}{2}\sum_{i=1}^{2}\frac{(x_i - \mu_i)^2}{2}\right\} I[\mu_1 > 0][A(x_1, x_2)]^{-1}$$

where

$$\begin{aligned}
A(x_1, x_2) &= \textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2 | \mu_1, \mu_2)p(\mu_1, \mu_2)d\mu_1\, d\mu_2 \\
&= \textstyle\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{4\pi}\exp\left\{-\frac{1}{2}\sum_{i=1}^{2}\frac{(x_i-\mu_i)^2}{2}\right\} d\mu_1\, d\mu_2 \\
&= \textstyle\int_{0}^{\infty} \frac{1}{2\sqrt{\pi}}\exp\left\{-\frac{1}{2}\frac{(x_1-\mu_1)^2}{2}\right\}\, d\mu_1 \\
&= \textstyle\int_{0}^{\infty} \frac{1}{2\sqrt{\pi}}\exp\left\{-\frac{1}{2}\frac{(\mu_1-x_1)^2}{2}\right\}\, d\mu_1 = \left(\lambda = \frac{\mu_1-x_1}{\sqrt{2}}\right) \\
&= \textstyle\int_{-\frac{x_1}{\sqrt{2}}}^{\infty} \frac{\exp(-.5\lambda^2)}{\sqrt{2\pi}}d\lambda = 1 - \Phi\left(\frac{-x_1}{\sqrt{2}}\right) \\
&= \Phi\left(\frac{x_1}{\sqrt{2}}\right)
\end{aligned}$$

Whence

$$p(\mu_1, \mu_2 | x_1, x_2) = \frac{1}{4\pi\Phi\left(\frac{x_1}{\sqrt{2}}\right)}\exp\left\{-\frac{1}{2}\sum_{i=1}^{2}\frac{(\mu_i - x_i)^2}{2}\right\} I[\mu_1 > 0]$$

We want to calculate $\mathrm{E}(\theta_1 | y_1, y_2)$, but since A is of full rank ($|A| = 2$), we know that (θ_1, θ_2) and (y_1, y_2) are uniquely determined by (μ_1, μ_2) and (x_1, x_2) respectively. So it is enough to calculate

$$\mathrm{E}\left(\frac{\mu_1 + \mu_2}{2}|x_1, x_2\right) = \frac{1}{2}(\mathrm{E}(\mu_1 | x_1, x_2) + \mathrm{E}(\mu_2 | x_1, x_2))$$

Further, note the following:

$$p(\mu_1, \mu_2 | x_1, x_2) = c(x_1)c(x_2)g(\mu_1, x_1)g(\mu_2, x_2)$$

where

$$g(\mu_i, x_i) = \exp\left\{-\frac{1}{2}\frac{(x_i - \mu_i)^2}{2}\right\} \text{ for } i = 1, 2$$

$$c(x_1) = \frac{1}{2\sqrt{\pi}\Phi\left(\frac{x_1}{\sqrt{2}}\right)}I[\mu_1 > 0] \qquad c(x_2) = \frac{1}{2\sqrt{\pi}}$$

Then,

$$\mathrm{E}(\mu_1 | x_1, x_2) + \mathrm{E}(\mu_2 | x_1, x_2) = \mathrm{E}(\mu_1 | x_1) + \mathrm{E}(\mu_2 | x_2)$$

Further, for $i = 1, 2$

$$p(\mu_i | x_i) = c(x_i)g(\mu_i, x_i)$$

Thus,

$$\mu_2 | x_2 \sim \mathrm{N}(x_2, 2) \Rightarrow \mathrm{E}(\mu_2 | x_2) = x_2$$

To calculate $\mathrm{E}(\mu_1 | x_1)$ we can rewrite it as

$$\mathrm{E}(\mu_1 | x_1) = \mathrm{E}((\mu_1 - x_1 + x_1) | x_1) = \mathrm{E}(\mu_1 - x_1 | x_1) + \mathrm{E}(x_1 | x_1)$$

1- $\mathrm{E}(x_1 | x_1) = x_1$

2- $\mathrm{E}(\mu_1 - x_1 | x_1) =$

$$\int_0^{\infty}(\mu_1 - x_1)p(\mu_1 | x_1)d\mu_1 = \left[\Phi\left(\frac{x_1}{\sqrt{2}}\right)\right]^{-1}\underbrace{\int_0^{\infty}\frac{(\mu_1 - x_1)}{2\sqrt{\pi}}\exp\left\{-\frac{1}{2}\frac{(x_i - \mu_i)^2}{2}\right\} d\mu_1}_{:=A(x_1)}$$

To calculate $A(x_1)$ we will

- use the following mathematical relationship
$$\frac{\partial}{\partial x_1} \exp\left\{-\frac{1}{2}\frac{(\mu_1 - x_1)^2}{2}\right\} = \frac{(\mu_1 - x_1)}{2} \exp\left\{-\frac{1}{2}\frac{(\mu_1 - x_1)^2}{2}\right\}$$
- note that it is possible to interchange the integral and derivative signs since all needed conditions hold.

$$\begin{aligned} A(x_1) &= \int_0^\infty \frac{1}{\sqrt{\pi}}\frac{\partial}{\partial x_1} \exp\left\{-\frac{1}{2}\frac{(x_i - \mu_i)^2}{2}\right\} d\mu_1 \\ &= 2\frac{\partial}{\partial x_1}\int_0^\infty \frac{1}{2\sqrt{\pi}} \exp\left\{-\frac{1}{2}\frac{(x_i - \mu_i)^2}{2}\right\} d\mu_1 \\ &= 2\frac{\partial}{\partial x_1}\Phi\left(\frac{x_1}{\sqrt{2}}\right) = \frac{2}{\sqrt{2}}\phi\left(\frac{x_1}{\sqrt{2}}\right) \end{aligned}$$

Hence,
$$\mathrm{E}(\mu_1|x_1) = x_1 + \frac{2}{\sqrt{2}}\phi\left(\frac{x_1}{\sqrt{2}}\right)\left[\Phi\left(\frac{x_1}{\sqrt{2}}\right)\right]^{-1}$$

Putting all pieces together
$$\begin{aligned} \mathrm{E}\left(\frac{\mu_1 + \mu_2}{2}\,|x_1, x_2\right) &= \frac{1}{2}\left(x_1 + \frac{2}{\sqrt{2}}\phi\left(\frac{x_1}{\sqrt{2}}\right)\left[\Phi\left(\frac{x_1}{\sqrt{2}}\right)\right]^{-1} + x_2\right) \\ &= \frac{x_1 + x_2}{2} + \frac{1}{\sqrt{2}}\phi\left(\frac{x_1}{\sqrt{2}}\right)\left[\Phi\left(\frac{x_1}{\sqrt{2}}\right)\right]^{-1} \end{aligned}$$

This gives us the desired result since
$$y_1 = \frac{x_1 + x_2}{2} \quad \text{and} \quad \frac{x_1}{\sqrt{2}} = \frac{y_1 - y_2}{\sqrt{2}} = d$$

Problem 4

A (one-dimensional) *scale density* is a density of the form
$$\frac{1}{\sigma}f\left(\frac{x}{\sigma}\right)$$
where $\sigma > 0$. The parameter σ is called a *scale parameter*. To derive a noninformative prior for this situation, imagine that, instead of observing X, we observe the random variable $Y = cX$ where $c > 0$. Defining $\eta = c\sigma$, and easy calculation shows that the density of Y is $\eta^{-1}f(y/\eta)$. If now the sample space $\mathcal{X} = \mathbb{R}$ or $\mathcal{X} = (0, \infty)$, then the sample and parameter spaces for the (X, σ) problem are the same as those for the (Y, η) problem. The two problems are thus identical in structure, which again indicates that they should have the same noninformative prior. Letting π and π^* denote the priors in (X, σ) and (Y, η) problems, respectively. With P^b representing the posterior distribution with respect to the prior b, this means that the equality
$$P^\pi(\sigma \in A) = P^{\pi^*}(\eta \in A)$$
should hold for all $A \subset (0, \infty)$. Since $\eta = c\sigma$, it should also be true that
$$P^{\pi^*}(\eta \in A) = P^\pi(\sigma \in c^{-1}A)$$
where $c^{-1}A = \{c^{-1}z | z \in A\}$. Putting these together, it follows that π should satisfy
$$P^\pi(\sigma \in A) = P^\pi(\sigma \in c^{-1}A) \tag{1}$$
This should hold for all $c > 0$, and any distribution π for which this is true is called *scale invariant*.

Assuming densities, equation (1) states that
$$\int_A \pi(\sigma)\,d\sigma = \int_{c^{-1}A} \pi(\sigma)\,d\sigma = \int_A c^{-1}\pi(c^{-1}\sigma)\,d\sigma$$
and conclude that, for this to hold for all A, it must be true that
$$\pi(\sigma) = c^{-1}\pi(c^{-1}\sigma)$$
for all σ. Choosing $\sigma = c$, it follows that
$$\pi(c) = c^{-1}\pi(1)$$
Since $\pi(1) = 1$ for convenience, and noting that the above equality must hold for all $c > 0$, it follows that a reasonable *noninformative prior for a scale parameter* is $\pi(\sigma) = \sigma^{-1}$. Observe that this is an improper prior since $\int_0^\infty \sigma^{-1}\,d\sigma = \infty$.

Problem 5

Recall that the participant has chosen box 1 and box 3 has been opened and it is empty. The following table shows the elements of the sample space and its corresponding (conditional) probabilities:

PEE2	PEE3	EPE2	EPE3	EEP2	EEP3
$p_{12}/3$	$p_{13}/3$	$p_{22}/3$	$p_{23}/3$	$p_{32}/3$	$p_{33}/3$

Note that $p_{i2} + p_{i3} = 1$ for $i = 1, 2, 3$

Let W denote the event "Winning" and S the event "To Switch". Thus:
$$P(W|S) = \frac{p_{23}}{p_{13} + p_{23}}$$

a- $p_{13} = p_{23} = p_{33} = 1 - p = q$ and $p_{12} = p_{22} = p_{32} = p$.

$$P(W|S) = \frac{q}{q+q} = \frac{1}{2}$$

b- $p_{13} = p_{23} = p32 = 1$, all others are equal to zero.

$$P(W|S) = \frac{1}{1+1} = \frac{1}{2}$$

c- $p_{12} = p$, $p_{13} = q$, $p_{23} = p_{32} = 1$, and $p_{22} = p_{33} = 0$.

$$P(W|S) = \frac{1}{1+q} > \frac{1}{2} \quad \text{since } 0 < q < 1$$

Thus, if the host follows strategies (a) or (b), it does not matter whether the participant switch or not. If the host follows strategy (c), the participant should switch to improve her probability of winning.

Problem 6

A reasonable sampling model could be:

$$y_i|\theta \sim \text{Poisson}(\lambda_i = \theta\ E_i)$$

where E_i is the exposure (i.e. population per 10,000 individuals) for region i . Assuming this sampling model, the following program could be used to get the results from WinBUGS.

```
model{ for (i in 1:15) { y[i] ~ dpois(lambda[i])
                         lambda[i] <- theta*pop[i]
                        }
       theta ~ dgamma(0.01,0.01)
     }

list(y=c(16,9,12,5,11,20,5,8,12,15,14,6,2,7,7),
     pop=c(13.5,16,24,5.5,22,40,10,16.5,21.5,23.5,25,8.5,3,8,11.5))
```

Note that the data could also have been declared using the following format (note that any string of characters in a line after the symbol # is treated as a comment)

```
y[] pop[] # highlight this row to load data
16 13.5
 9 16
12 24
 5  5.5
11 22
20 40
 5 10
 8 16.5
12 21.5
15 23.5
14 25
 6  8.5
 2  3
 7  8
 7 11.5
.END # this is the end of the data
```