AUSTRALIAN NATIONAL UNIVERSITY

RESEARCH SCHOOL OF FINANCE ACTUARIAL STUDIES, AND APPLIED STATISTICS

INTRODUCTION TO BAYESIAN DATA ANALYSIS (STAT3016/4116/7016)
SEMESTER 2 2017

ASSIGNMENT 2 - SOLUTIONS

Problem 1

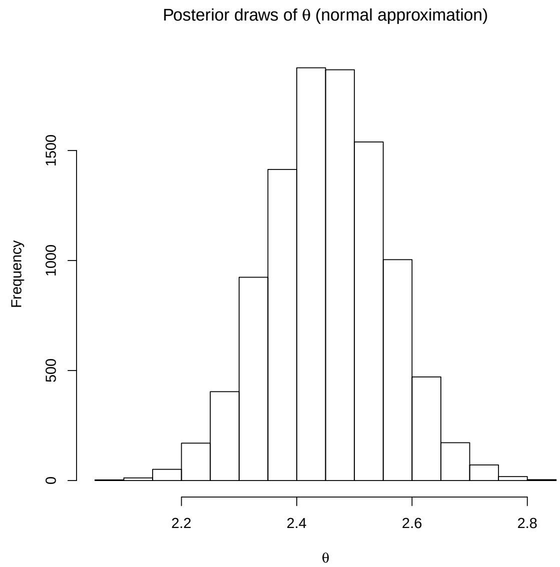
(a) The posterior density of θ is given by

$$g(\theta|y) \propto \left(\frac{\exp(y\theta)}{(1+\exp(\theta))^n}\right)\exp\left[\frac{-(\theta-\mu)^2}{2\sigma^2}\right]$$

(b) For the normal approximation to the posterior, let's assume σ is known . The prior for θ has parameters $\mu_0 = 0$ and $\tau_0^2 = 0.25$. If we apply the normal approximation to the binomial data $y \overset{\approx}{\sim} N(n\theta, n\theta(1-\theta))$, and so $E[y|\theta] = n\theta$ and $Var[y|\theta] = n\theta(1-\theta)$. Using the results of a conjugate prior for a normal model, conditional on the variance, we have $\theta|y \sim N(\mu_n, \tau_n^2)$ where $\tau_n^2 = \frac{1}{\frac{1}{0.25^2}+\frac{1}{0.25^2}}$ (where we have set $\sigma^2 = \tau_0^2$) and $\mu_n = \frac{\frac{1}{0.25^2}\times 0+\frac{1}{0.25^2}\times\bar{y}}{\frac{1}{0.25^2}+\frac{1}{0.25^2}}$

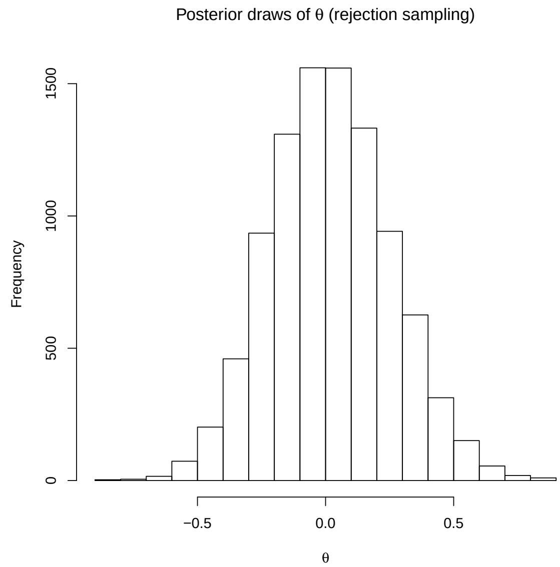
The data n=5 and y=5 suggests that $\hat{p}_j = 1$ for $j = 1, ..., 5$. As logit(1) is not defined, approximate $\bar{y}$ with $logit(0.95) = \ln(19)$.

A histogram of posterior draws from the normal approximation to the posterior is shown below based on 10000 posterior draws. All of the 10000 draws exceeded zero. Hence, the probability that the coin is biased is estimated to be 1.



(c) The R code to implement the rejection sampling algorithm is below. The value of M which produces a high acceptance rate (aroudn 95%) was set to 0.01.

```
y<-5
n<-5
theta.prop<-rnorm(S,mu0,sqrt(tau02))
target<-function(theta) exp(y*theta)/(1+exp(theta))^n*exp(-(theta-mu0)^2/(2*tau02))
M<-0.01
prop<-function(theta) dnorm(theta,mu0,sqrt(tau02))
alpha<-target(theta.prop)/(M*prop(theta.prop))
U<-runif(S)
mean(alpha>=U)
[1] 0.957
theta.post<-theta.prop[alpha>=U]
> mean(theta.post>=0)
[1] 0.5231975
```

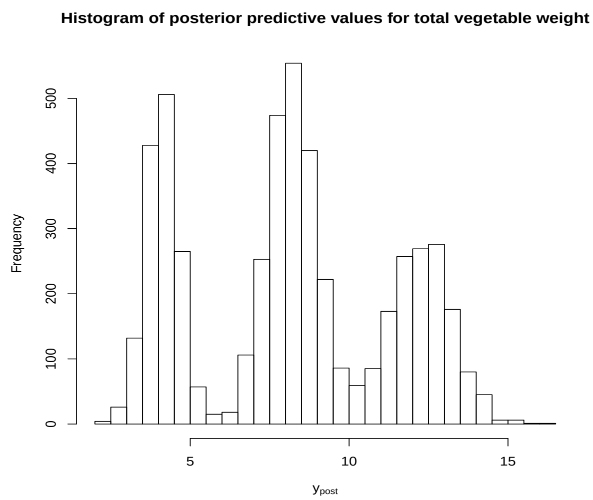


A histogram of posterior draws from the rejection sampling algorithm is shown below based on 10000 posterior draws. 52% of the 10000 draws exceeded zero. Hence, the probability that the coin is biased is estimated to be 0.52. This is a more realistic answer than our estimate of the probability in part (b), because with only 5 draws, we cannot guarantee that all future draws will be heads and conclude for certain that the coin is biased.

Problem 2

(a) Using the posterior distributions of the parameters (μ, σ^2), we generate 5000 posterior predictive values of y as follows:

```
sigma2.post<-1/rgamma(5000,10,2.5)
theta.post<-rnorm(5000,4.1,sqrt(sigma2.post/20))
s<-sample(c(1,2,3),5000,prob=c(0.31,0.46,0.31),replace=TRUE)
y.post<-rnorm(5000,s*theta.post,sqrt(s*sigma2.post))
```

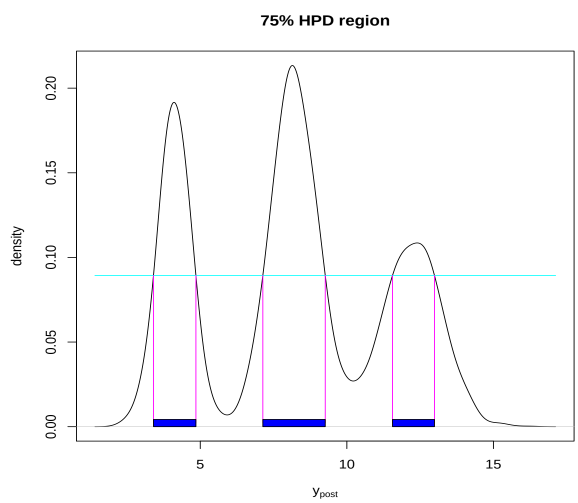


The histogram of posterior predictive values shows three modes corresponding to the three subpopulations.

(b)
```
> quantile(y.post,probs=c(0.5-0.375,0.5+0.375))
    12.5%     87.5%
 4.028819 12.440252
```

A 75% quantile-based interval for a new value of Y is (4.03, 12.44).

(c)
```
x<-density(y.post)$x
d<-density(y.post)$y
d<-d/sum(d)
dd<-cbind(x,d)

dd<-dd[order(dd[,2],decreasing=TRUE),]
i<-1
while(sum(dd[1:i,2])<0.75) {
i<-i+1
}
HPD<-dd[1:i,1]
sort(HPD)
```



The HPD region is given by the intervals: (3.25,4.99) ; (6.94,9.48); (11.32,13.17) nb: you can also use the `'hdr'` function in the library `hdrcde` and you will get similar values for the HPD region (highlighted in blue in the plot above).

(d) A mixture sampling distribution might be appropriate if it is believed that the population of vegetable weights can be divided into three subpopulations, each defined by a different mean and or variance parameter.

(e) To be done

Problem 3

(a) For a noninformative prior distribution, use $p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$. Then the posterior distribution is

$$
\begin{aligned}
p(\mu, \sigma^2 | y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\right) \\
&= \sigma^{-7} \exp\left(-\frac{1}{2\sigma^2}[4(0.5) + 5(10 - \mu)^2]\right)
\end{aligned}
$$

equivalent to $\mu | \sigma^2, y \sim N(10, \sigma^2/5)$ and $\sigma^2 | y \sim \text{Inv-Gamma}(\frac{4}{2}, \frac{4(0.5)}{2})$

(b) Now treat the data as rounded. Let z be the unrounded data points so that $z_i \sim N(\mu, \sigma^2)$, $\epsilon_i \sim \text{Unif}(-0.5, 0.5)$, and $y_i = z_i - \epsilon_i$. Then

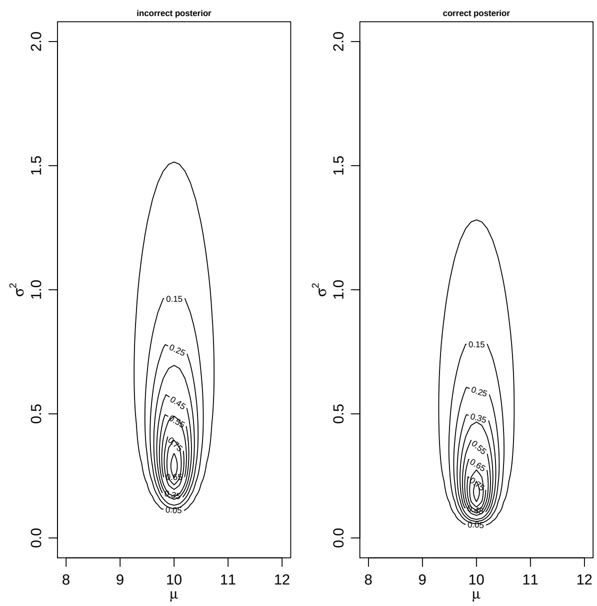
$$
p(\mu, \sigma^2 | y, \epsilon) \propto \sigma^{-2} \prod_{i=1}^{5} \frac{1}{\sigma} \exp\left(-\frac{(y_i + \epsilon_i - \mu)^2}{2\sigma^2}\right)
$$

and

$$
\begin{aligned}
p(\mu, \sigma^2 | y) &= \int p(\mu, \sigma^2 | y, \epsilon) p(\epsilon) d\epsilon \\
&\propto \sigma^{-2} \prod_{i=1}^{5} \frac{1}{\sigma} \int_{-0.5}^{0.5} \exp\left(-\frac{(y_i + \epsilon_i - \mu)^2}{2\sigma^2}\right) d\epsilon_i \\
&\propto \sigma^{-2} \prod_{i=1}^{5} \left[\Phi\left(\frac{y_i + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{y_i - 0.5 - \mu}{\sigma}\right)\right]
\end{aligned}
$$

where $\Phi(z)$ is the standard normal CDF.

(c) We can approximate the incorrect and correct posterior densities by evaluating them over a 200×200 grid from 0 to 20 in μ and -4 to 1 in $\log \sigma^2$.



The contour plot shows slightly less posterior variation in σ^2 under the correct distribution

μ	posterior mean	posterior variance
Correct	10.00	0.09
Incorrect	10.00	0.11

$\boldsymbol{\sigma^2}$	posterior mean	posterior variance
Correct	0.33	0.09
Incorrect	0.47	0.12

The posterior summaries for μ are approximately similarly. The expected value of σ^2 (=0.33) using the correct distribution is lower than the expected value of σ^2 (=0.47) using the incorrect distribution. This tells us that using the incorrect distribution can lead us to overestimate the variance of the population. This is possible, if we consider for example, two actual values of 10.1 and 10.6, which would be rounded to 10 and 11 respectively, and thus overstating the range between the two values. Of course, the opposite situation could occur, where, for example the two actual values are 10.8 and 11.2, but both are rounded to 11, and if we use the incorrect distribution, we would be underestimating the true variance. So whether the true variance is underestimated or overestimated depends on the relative direction of the rounding.

(d) We need to know the posterior distribution $p(z_i|y_i = 10, y)$. We can draw values of $(\mu^{(i)}, \sigma^{2(i)})$ from the posterior distribution of (μ, σ^2), draw $z^{(i)} \sim N(\mu^{(i)}, \sigma^{2(i)})$, and keep only those values for which `round`$(z^{(i)}) = 10$. We repeat this process and keep only those values for which `round`$(z^{(i)}) = 11$. We then sample from these values and save the draws in the vectors z_1 and z_3. This results in a posterior mean for $(z_1 - z_3)^2$ of 0.76.

Problem 4

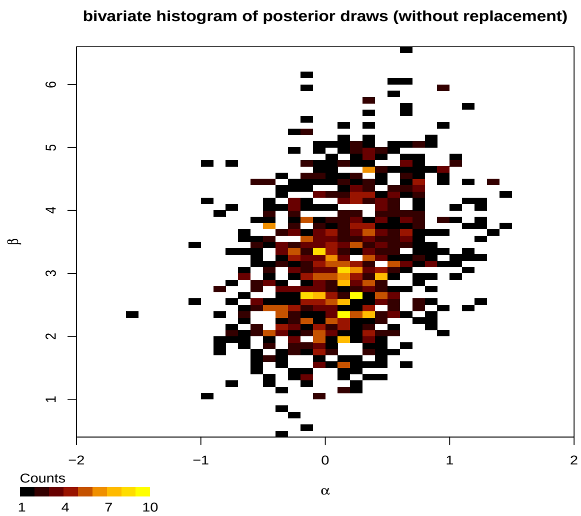
For the normal approximation, let $g(\alpha, \beta)$ be a multivariate normal density with mean vector $(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$ and variance-covariance matrix $\hat{V}_{MLE}$, where $\hat{\alpha}_{MLE}, \hat{\beta}_{MLE}$ and $\hat{V}_{MLE}$ are estimated from the logistic regression of θ_i on x_i.
We obtain,

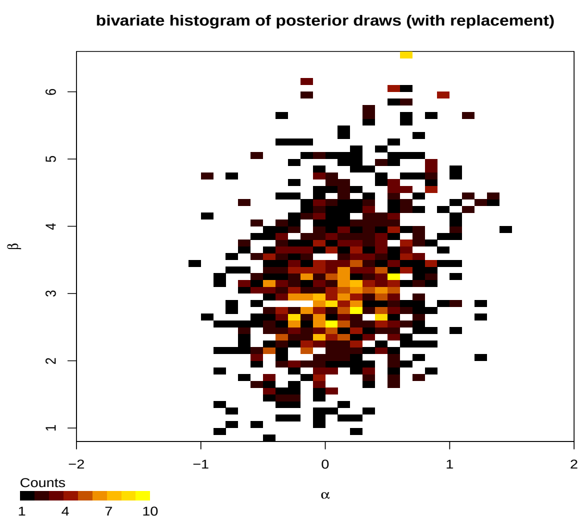
$$(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE}) = (0.1074,\ 2.8414) \text{ and } \hat{V}_{MLE} = \begin{pmatrix} 0.1697 & 0.1066 \\ 0.1066 & 0.8464 \end{pmatrix}$$

The true density is proportional to $\prod_{i=1}^{4} \binom{n_i}{y_i} \text{invlogit}(\alpha + \beta x_i)^{y_i} (1 - \text{invlogit}(\alpha + \beta x_i))^{n_i - y_i}$
The R-code to implement the importance resampling with and without replacement is:

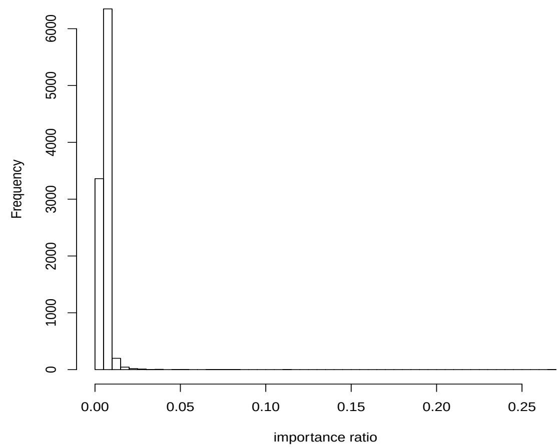
```
library("mvtnorm")
#data
x<-c(-0.86,-0.30,-0.05,0.73)
n<-c(10,10,10,10)
y<-c(0,3,7,8)



#parameters  of normal approximation
model<-glm(cbind(y,n-y)~x,family=binomial)
S<-10000
#10000 samples from normal approximation
g.draw<-rmvnorm(S,mean=model$coef,sigma=summary(model)$cov.unscaled)

invlogit<-function(x) exp(x)/(1+exp(x))
#true posterior density
lik<-function(a) prod(choose(n,y)*invlogit(a[1]+a[2]*x)^y*(1-invlogit(a[1]+a[2]*x))^(n-y))

k<-1000

#resample without replacement
imp.draw.id<-sample(seq(1,S,1),size=k,prob=apply(g.draw,1,lik)/
  apply(g.draw,1,function(d) dmvnorm(d,mean=model$coef,sigma=summary(model)$cov.unscaled)))
imp.draw<-g.draw[imp.draw.id,]
```

```
#resample with replacement
imp.draw.id2<-sample(seq(1,S,1),size=k,replace=TRUE, prob=apply(g.draw,1,lik)/
  apply(g.draw,1,function(d) dmvnorm(d,mean=model$coef,sigma=summary(model)$cov.unscaled)))

imp.draw2<-g.draw[imp.draw.id2,]
```





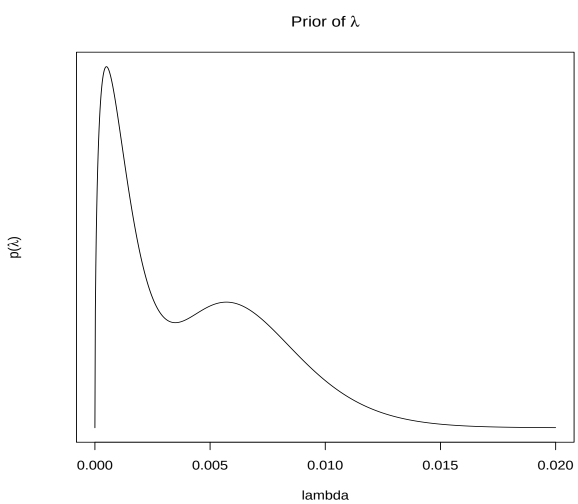
The frequency of draws in the upper region of the plot is higher with replacement. Below is a histogram of the importance weights.



Most of the ratios are small, but there are a few large weights. In this case, sampling with replacement will pick the same few values of (α, β) repeatedly. Sampling without replacement is preferred in this case. If the importance weights are moderate, sampling with and without replacement give similar results.

Problem 5

(a) The prior distribution is $p(\lambda) = 0.5p_1(\lambda)+0.5p_2(\lambda)$ where $p_1(\lambda)$ is a gamma(1.5,1000) distribution and $p_2(\lambda)$ is a gamma(7,1000) distribution. A sketch of the prior distribution is shown in the graph below.

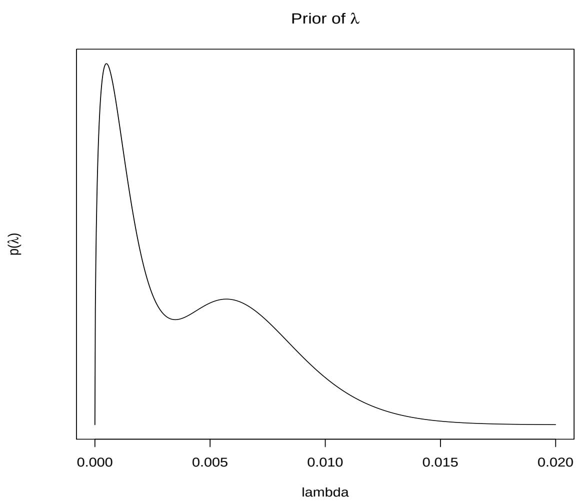


(b) $y = 4$; x=exposure=1767

$$\begin{aligned} p(\lambda|Y) &\propto p(y|\lambda)p(\lambda) \\ &= (0.5p_1(\lambda) + 0.5p_2(\lambda)) \exp(-\lambda x)(x\lambda)^y \\ &\propto \lambda^{1.5-1} \exp(-1000\lambda) \exp(-1767\lambda)(\lambda)^4 + \lambda^{7-1} \exp(-1000\lambda) \exp(-1767\lambda)(\lambda)^4 \\ &= \lambda^{1.5+4-1} \exp(-(1000 + 1767)\lambda) + \lambda^{7+4-1} \exp(-(1000 + 1767)\lambda) \\ &= w \times gamma(5.5, 8767) + (1 - w) \times gamma(11, 8767) \end{aligned}$$

where $w = \frac{\Gamma(5.5)/8767^{5.5}}{\Gamma(5.5)/8767^{5.5} + \Gamma(11)/8767^{11}}$

(c) The prior and posterior density of λ are shown in the graph below. The prior is bimodal but the posterior is unimodal.



(d) Using Monte Carlo simulation, we estimate that the probability that the mortality rate exceeds 0.005 is 0.1%.

```
> lambda.sample<-sample(lambda,1000,prob=p.post,replace=TRUE)
> mean(lambda.sample>0.005)
[1] 0.001
```

(e) The posterior weight for the first expert is $w = 1$ which tells us that the data were more consistent with the first expert's opinion.