AUSTRALIAN NATIONAL UNIVERSITY

RESEARCH SCHOOL OF FINANCE ACTUARIAL STUDIES, AND APPLIED STATISTICS

INTRODUCTION TO BAYESIAN DATA ANALYSIS (STAT3016/4116/7016)

SEMESTER 2 2017

ASSIGNMENT 1 - SOLUTIONS

(see `"assign1.R"` for R code).

Problem 1
The R code for the Monte Carlo function plus sample output is below:

```
m<-1000
n<-c(10,25,100)
p<-c(0.05,0.25,0.50)

binom.conf.interval<-function(y,n){
  z<-qnorm(0.975)
  phat<-y/n
  se<-sqrt(phat*(1-phat)/n)
  return(c(phat-z*se,phat+z*se))
}

CI_CR<-function(){
CR<-NULL
for (i in 1:3){
  for (j in 1:3){
    COV<-0
  y<-rbinom(m,n[i],p[j])
  for(k in 1:m){
    CI<-binom.conf.interval(y[k],n[i])
    COV<-COV+1*(p[j]>CI[1]& p[j]<CI[2])
  }
  CR<-c(CR,COV/m)
```

```
}
}
names(CR)<-c("n=10; p=0.05","n=10; p=0.25","n=10; p=0.50","n=25; p=0.05","n=25; p=0.25",
      "n=25; p=0.25", "n=100; p=0.05","n=100; p=0.25","n=100; p=0.50")
return(CR)
}

> CI_CR()
 n=10; p=0.05  n=10; p=0.25  n=10; p=0.50  n=25; p=0.05  n=25; p=0.25  n=25; p=0.50
        0.387         0.905         0.894         0.713         0.898         0.952

n=100; p=0.05 n=100; p=0.25 n=100; p=0.50
        0.874         0.934         0.945
```

From the sample output of our Monte Carlo study, we see that coverage rates of the traditional interval increase and approach 95% as the sample size increases and the true proportion increases to 0.5 (subject to simulation error). This makes sense because the normal approximation to the sampling distribution of $\hat{p}$ requires $np > 10$ and $n(1-p) > 10$.

Problem 2

First note that θ can take on the values in the interval [-1,1]

(a) $\theta \sim$ Unif(-1,1). This is a flat prior on the allowable range of values for θ

(b) $\theta \sim$ Beta(2,5). The mean of this distribution is 0.29 and the mode is 0.20. Given only a month of practice, and the likely range of initial success rates, most of the sample will improve by 0.5 at most, and it is highly unlikely for the average improvement to be above 0.8.

(c) $\theta \sim$ Unif(0,1). After practicing for a month, it is reasonable to expect the students to perform the same as the first trial or better. Restrict the range of θ to reflect this but keep the prior flat as we have no information to guide the range of improvement.

Problem 3

Prior density :

$$p(\theta) \propto \theta^3(1-\theta)^3$$

Likelihood: Let Y be the number of heads

$$\begin{aligned} Pr(Y<3|\theta) &= \binom{10}{0}(1-\theta)^{10} + \binom{10}{1}\theta^1(1-\theta)^9 + \binom{10}{2}\theta^2(1-\theta)^8 \\ &= (1-\theta)^{10} + 10\theta^1(1-\theta)^9 + 45\theta^2(1-\theta)^8 \end{aligned}$$

Posterior density:

$$p(\theta|Y<3) \propto \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11}$$

We require $\int_0^1 p(\theta|Y<3)d\theta = 1$
Now

$$\begin{aligned} &\int_0^1 \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11}\ d\theta \\ &= B(4,14)\int_0^1 \frac{\theta^3(1-\theta)^{13}}{B(4,14)}d\theta + 10B(5,13)\int_0^1 \frac{\theta^4(1-\theta)^{12}}{B(5,13)}d\theta + 45B(6,12)\int_0^1 \frac{\theta^4(1-\theta)^{11}}{B(6,12)} \\ &= B(4,14) + 10B(5,13) + 45B(6,12) \end{aligned}$$

Then the exact posterior density is

$$p(\theta|Y<3) = \frac{(\theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11})}{B(4,14) + 10B(5,13) + 45B(6,12)}$$

or

$$\theta|Y<3 \sim \frac{B(4,14)Beta(4,14) + 10B(5,13)Beta(5,13) + 45B(6,12)Beta(6,12)}{B(4,14) + 10B(5,13) + 45B(6,12)}$$

where $B(a,b)$ is the beta function and Beta(a,b) is the probability density function of a beta distribution with parameters a and b.

Let $k = B(4,14) + 10B(5,13) + 45B(6,12)$

$$E[\theta|Y<3] = \frac{B(4,14)}{k}\times\frac{4}{18} + \frac{10B(5,13)}{k}\times\frac{5}{18} + \frac{45B(6,12)}{k}\times\frac{6}{18} = 0.3046875$$

$$E[\theta^2|Y<3] = \frac{1}{56}\times\frac{4}{18}\times\frac{5}{19} + \frac{10}{56}\times\frac{5}{18}\times\frac{6}{19} + \frac{45}{56}\times\frac{6}{18}\times\frac{7}{19} = \frac{2210}{19152} = 0.1053088$$

$$Var[\theta^2|Y<3] = 0.1053088 - 0.3046875^2 = 0.01247437$$

where if $\theta \sim \text{Beta}(a,b)$, then $E[\theta^2] = \frac{a}{a+b} \times \frac{a+1}{a+b+1}$

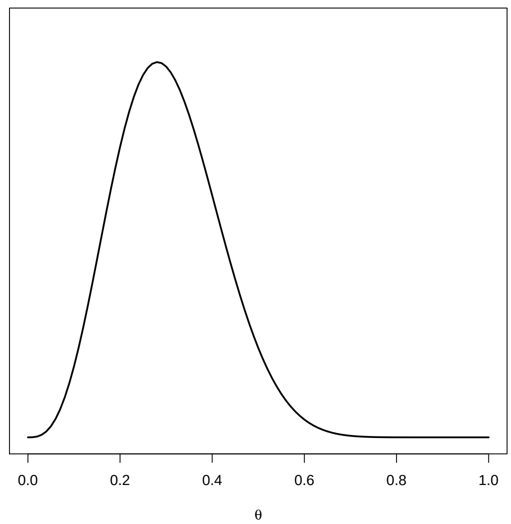

Figure 1: posterior density for θ

Sampling 1000 values of theta from the posterior density $\theta|Y<3$, and taking the 2.5% and 97.5% quantiles, a 95% posterior confidence interval for θ is (0.12,0.54)

Alternatively, we can also compute a 95% HPD interval with the following R code:

```
hpd<-function(x,dx,p){
md<-x[dx==max(dx)]
px<-dx/sum(dx)
pxs<--sort(-px)
ct<-min(pxs[cumsum(pxs)< p])
list(hpdr=range(x[px>=ct]),mode=md) }
ord<-order(-dens)

xdens<-cbind(theta[ord],dens[ord])
xdens<-cbind(xdens,cumsum(xdens[,2])/sum(xdens[,2]))

hpd(xdens[,1],xdens[,2],0.95)$hpdr
```

The 95% HPD interval is (0.10,0.51)

The table below compares posterior summaries for the two different priors:

	Beta(4,4)	Beta(20,20)
$E[\theta \mid Y < 3]$	0.305	0.434
$E[\theta^2 \mid Y < 3]$	0.105	0.194
$Var[\theta \mid Y < 3]$	0.01247	0.00049
95% posterior CI	(0.12, 0.54)	(0.29, 0.59)
95% HPD	(0.10, 0.51)	(0.30, 0.56)

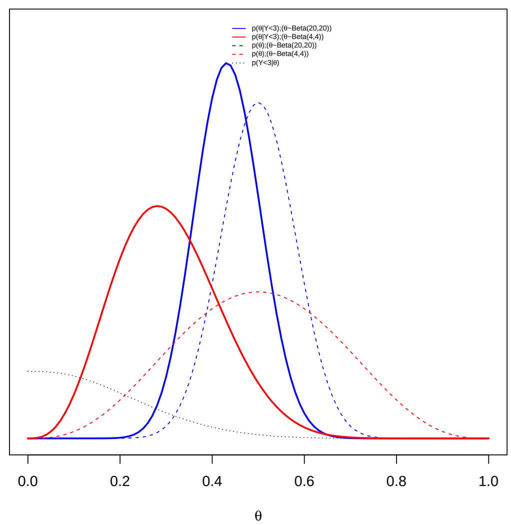

Figure 2: Comparison of posterior density for θ

The sketch in Figure 3 shows a tighter posterior density assuming a Beta(20,20) prior, that is closer to the prior distribution. This makes sense because the prior sample size has more than doubled, so our posterior results are relatively more heavily weighted to the prior distribution assumptions.

Problem 4

(a) $p(\text{data}|N) = \frac{1}{N}$ fpr $N \geq 159$ and zero when $N < 159$.

$$p(N|\text{data}) \propto p(N)p(\text{data}|N) = \frac{1}{N}\frac{1}{200}\left(\frac{199}{200}\right)^{N-1} \propto \frac{1}{N}\left(\frac{199}{200}\right)^{N} \quad \text{for } N \geq 159$$

(b) $p(N|\text{data}) = c\frac{1}{N}\left(\frac{199}{200}\right)^N$. We need to compute the normalizing constant c. Now $\sum_N p(N|\text{data}) = 1$, and so $1/c = \sum_{159}^{\infty} \frac{1}{N}\left(\frac{199}{200}\right)^N$. This sum can be computed analytically (as $c = \sum_0^{\infty} \frac{1}{N}\left(\frac{199}{200}\right)^N - \sum_0^{158} \frac{1}{N}\left(\frac{199}{200}\right)^N$). However, it is easier to approximate by a computer when we do not have a nice formula.

$\sum_{159}^{\infty} \frac{1}{N}\left(\frac{199}{200}\right)^N \approx 0.3125823$ and $c \approx 3.2$.

$$E[N|\text{data}] = c\sum_{159}^{\infty} N\frac{1}{N}\left(\frac{199}{200}\right)^N \approx c\sum_{159}^{\infty}\left(\frac{199}{200}\right)^N = 3.2\frac{\left(\frac{199}{200}\right)^{159}}{1-\left(\frac{199}{200}\right)} = 288.44$$

$$SD[N|\text{data}] = \sqrt{\sum_{159}^{\infty}(N-288.44)^2 c\frac{1}{N}\left(\frac{199}{200}\right)^N} \approx 122.38$$

(see computer code for summation)

(c) There are many possible solutions. One idea that does not work is the improper discrete uniform prior density on N: $p(N) \propto 1$ for N. This density leads to an improper posterior density: $p(N|\text{data}) \propto \frac{1}{N}$ for $N \geq 159$. ($\sum_{159}^{\infty}(1/N) = \infty$). The prior density $p(N) \propto \frac{1}{N}$ is improper, but leads to a proper posterior density, because $\sum_N \frac{1}{N^2}$ is convergent. However, we cannot evaluate the posterior mean or variance, because the required summation is divergent.

Note that if more than one data point is available (that is, if more than one cable car number is observed), then the posterior density is proper under all of the above prior densities, and the posterior mean and variance do exist.

With only one data point, perhaps it would not make much sense in practice to use a noninformative prior distribution here.

Problem 5

(a) The prior distribution of λ is given in the table below:

λ	0.5	1	1.5	2	2.5	3
$g(\lambda)$	0.1	0.2	0.3	0.2	0.15	0.05

We observe y=12 and t=6. The table below evaluates $g(\lambda)\exp(-t\lambda)(t\lambda)^y$ for each value of λ:

λ	0.5	1	1.5	2	2.5	3	Total
$g(\lambda)\exp(-t\lambda)(t\lambda)^y$	2646	1079141	10456372	10956483	5953456	880926	29329024

The normalising constant of the posterior density of λ is 29329024. Therefore, we have the following posterior probabilities for the different rate values:

λ	0.5	1	1.5	2	2.5	3
$g(\lambda\|y)$	0.0000902	0.0368	0.3565	0.3736	0.2030	0.0300

(b) For a seven-day period, t=7, and $g(y = 0|\lambda) = \exp(-7\lambda)$. We need to find the predictive probability that $g(\tilde{y} = 0|y = 12)$. To do so, we can sum over the posterior probabilities of λ as calculated in part (a). This is done in the table below:

λ	0.5	1	1.5	2	2.5	3
$g(\tilde{y} = 0\|\lambda)g(\lambda\|y)$	2.72×10^{-6}	3.36×10^{-5}	9.82×10^{-6}	3.11×10^{-7}	5.10×10^{-9}	2.28×10^{-11}

Summing the values in the table above, the predictive probability that there are no accidents in the next week is 0.0000464 (very small!)

Problem 6

If we knew the coin that was chosen, then the problem would be simple. Let θ be the probability that the coin lands heads, and let N be the number of additional spins required until a head, then

$$E[N|\theta] = 1 \cdot \theta + 2 \cdot (1-\theta)\theta + 3 \cdot (1-\theta)^2\theta + \ldots = \frac{1}{\theta}$$

Let TT denote the event that the first two spins are tails, and let C be the coin that was chosen. By Bayes' rule,

$$\begin{aligned}
Pr(C = C_1|TT) &= \frac{Pr(C = C_1)Pr(TT|C = C_1)}{Pr(C = C_1)Pr(TT|C = C_2) + Pr(C = C_1)Pr(TT|C = C_2)} \\
&= \frac{0.5(0.75)^2}{0.5(0.75)^2 + 0.5(0.25)^2} = 0.90
\end{aligned}$$

The posterior expectation of N is then:

$$\begin{aligned}
E[N|TT] &= E[E[N|TT, C]|TT] \\
&= Pr(C = C_1|TT)E[N|C = C_1, TT] + Pr(C = C_2|TT)E[N|C = C_2, TT] \\
&= 0.9 \times \frac{1}{0.25} + 0.1 \times \frac{1}{0.75} = 3.73
\end{aligned}$$