Michael Leibert
Math 611
Homework 12

1. Exercise 7.25 a, pg. 235, Robert and Casella.

Exercise 7.25 Rao-Blackwellization can be applied to most of the Gibbs samplers in this chapter. For each of the following examples, verify the conditional expectations provided there and compare via an R experiment the empirical average with the Rao-Blackwellization.

a. Example 7.2: $E\left[\theta|x\right] = \dfrac{x+a}{n+a+b}$

Consider the pair of distributions

$$X|\theta \sim \mathcal{B}in(n,\theta), \quad \theta \sim \mathcal{B}e(a,b),$$

with the joint distribution

$$f(x,\theta) = \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{x+a-1} \; (1-\theta)^{n-x+b-1}.$$

With conditional distributions

$$f(\theta|x) = \frac{f(x,\theta)}{f(x)} \propto \theta^{x+a-1} \; (1-\theta)^{n-x+b-1}$$

$$\mathcal{B}e\Big(a+x, n-x+b\Big)$$

$$f(x|\theta) = \frac{f(x,\theta)}{f(\theta)} \propto \binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

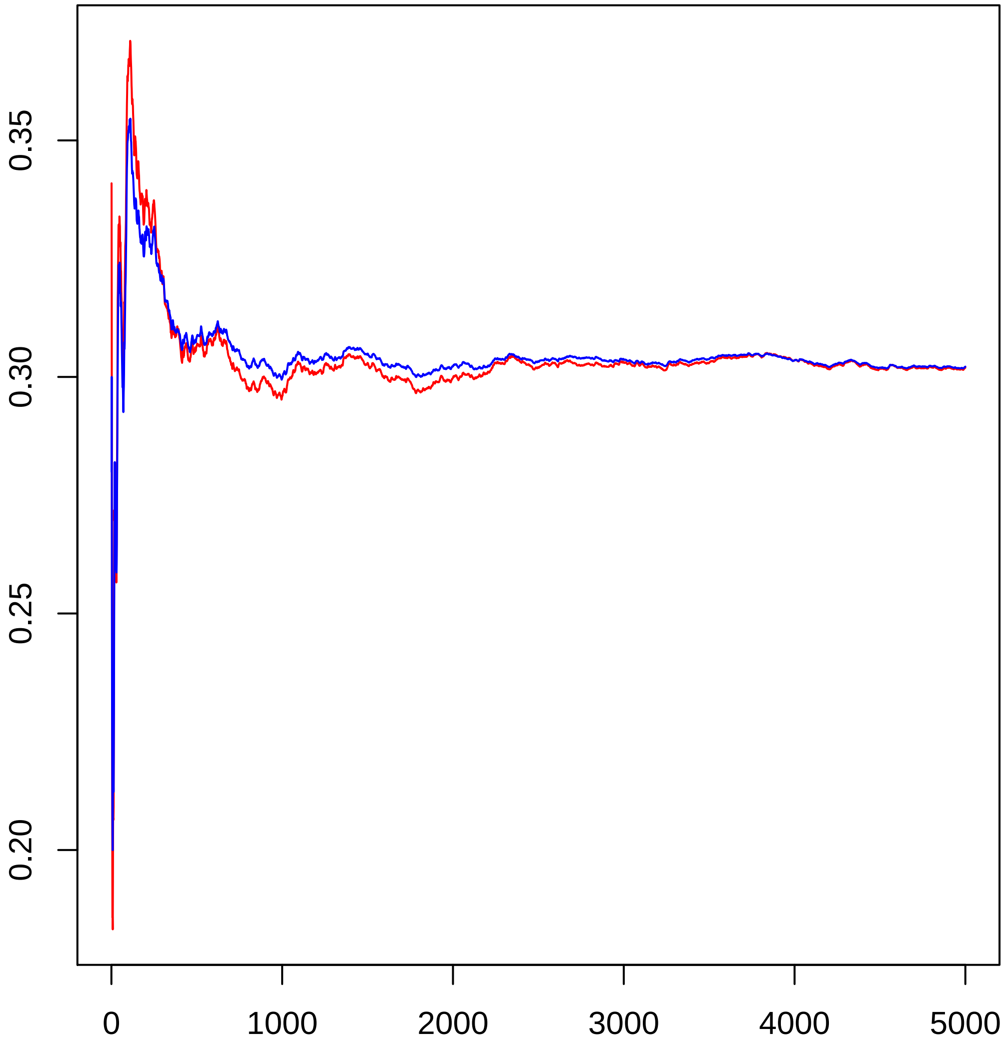
$$\mathcal{B}in\Big(n,\theta\Big).$$

Thus $E\left[\theta|x\right] = \dfrac{x+a}{a+x+n-x+n} = \dfrac{x+a}{a+n+b}$.

Comparison of the empirical average with the Rao-Blackwellization.

```
set.seed(111)
Nsim=5000 #initial values
 n=15
 a=3
 b=7
 X=T=array(0,dim=c(Nsim,1)) #init arrays
 T[1]=rbeta(1,a,b) #init chains
 X[1]=rbinom(1,n,T[1])
 for (i in 2:Nsim){
 X[i]=rbinom(1,n,T[i-1])
 T[i]=rbeta(1,a+X[i],n-X[i]+b) }
```

```
par(mar=c(4,4,1,2))
plot(cumsum(T)/(1:Nsim),type="l",col="red",    xlab="",ylab="")
lines(cumsum((X+a))/((1:Nsim)*(n+a+b)),col="blue")
```



2. Consider the hierarchical Bayes model:

$$Y \sim b(n,p), \quad 0 < p < 1$$
$$p|\theta \sim h(p|\theta) = \theta p^{\theta-1}, \quad \theta > 0$$
$$\theta \sim Gamma(1,a), \quad \text{for a fixed } \alpha > 0$$

a. Show analytically that $g\,(p|y,\theta)$ is $Beta\,(y+\theta, n-y+1)$

We know that the conditional pdf $f(y|p)$ does not depend on θ, therefore,

$$\begin{aligned} g\,(p,\theta|y) &= \frac{g\,(y,\theta,p)}{g(y)} \\ &= \frac{g\,(y|\theta,p)\,g\,(\theta,p)}{g(y)} \\ &= \frac{f\,(y|p)\,h\,(p|\theta)\,\psi(\theta)}{g(y)}. \end{aligned}$$

The joint distribution is given by:

$$
\begin{aligned}
g(y,\theta,p) &= f(y|p)\,h(p|\theta)\,\psi(\theta) \\
&= \binom{n}{y}\ p^{y}\ (1-p)^{n-y}\ \theta\ p^{\theta-1}\ \frac{1}{\alpha}\ \exp\left(-\frac{\theta}{\alpha}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
g(p|y,\theta) &= \frac{g(p,\theta,y)}{g(\theta,y)} \\
&\propto p^{y}\ (1-p)^{n-y} p^{\theta-1} \\
&= p^{y+\theta-1}\ (1-p)^{n-y} \qquad \text{kernel of } Beta\,(y+\theta, n-y+1)\,.
\end{aligned}
$$

b. Show analytically that $g(\theta|y,p)$ is $Gamma\left(2, \left[\frac{1}{\alpha} - \log p\right]^{-1}\right)$

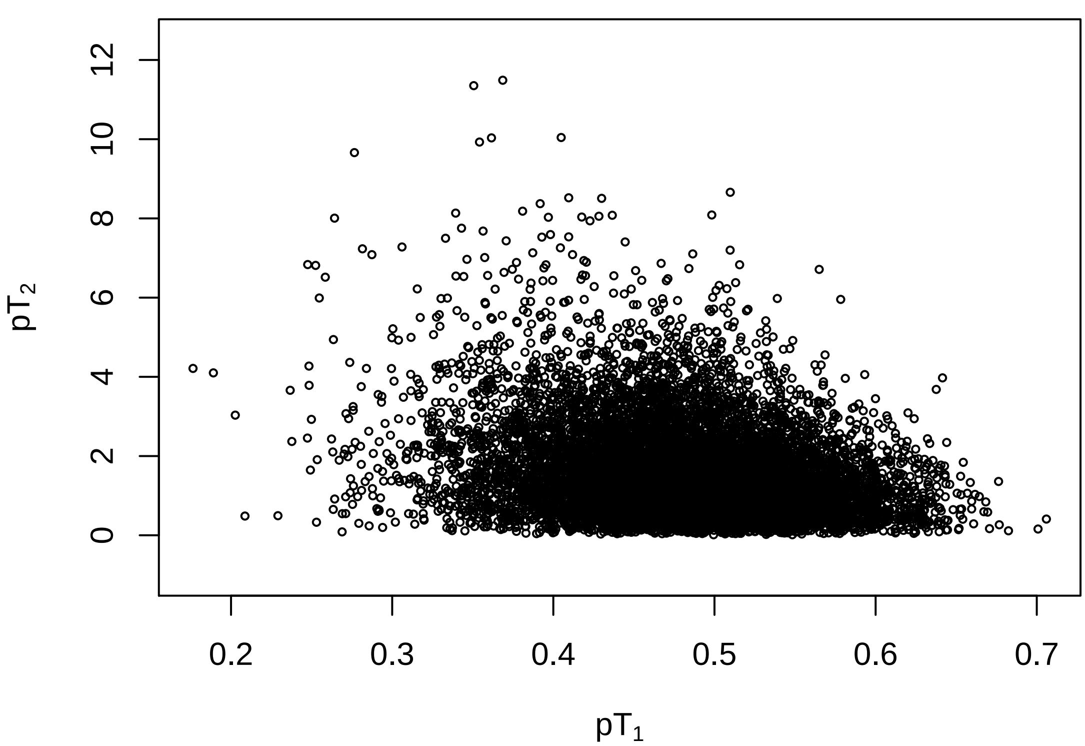
$$
\begin{aligned}
g(\theta|y,p) &= \frac{g(p,\theta,y)}{g(p,y)} \\
&\propto \theta\ p^{\theta}\ \exp\left(-\frac{\theta}{\alpha}\right) \\
&= \theta\ \exp\left(\log p^{\theta}\right)\ \exp\left(-\frac{\theta}{\alpha}\right) \\
&= \theta\ \exp\left(\theta \log p\right)\ \exp\left(-\frac{\theta}{\alpha}\right) \\
&= \theta\ \exp\left(-\theta\left[\frac{1}{\alpha} - \log p\right]\right) \\
&= \theta^{2-1}\ \exp\left(-\frac{\theta}{\left[\frac{1}{\alpha} - \log p\right]^{-1}}\right) \qquad \text{kernel of } Gamma\left(2, \left[\frac{1}{\alpha} - \log p\right]^{-1}\right)
\end{aligned}
$$

c. Write and implement the Gibbs sampler algorithm to obtain the Bayes estimate of p. Show the two dimensional scatter plot of (p,θ) and the density plot of the marginal distribution of p. It is up to you to determine the initial values, number of iterations etc.

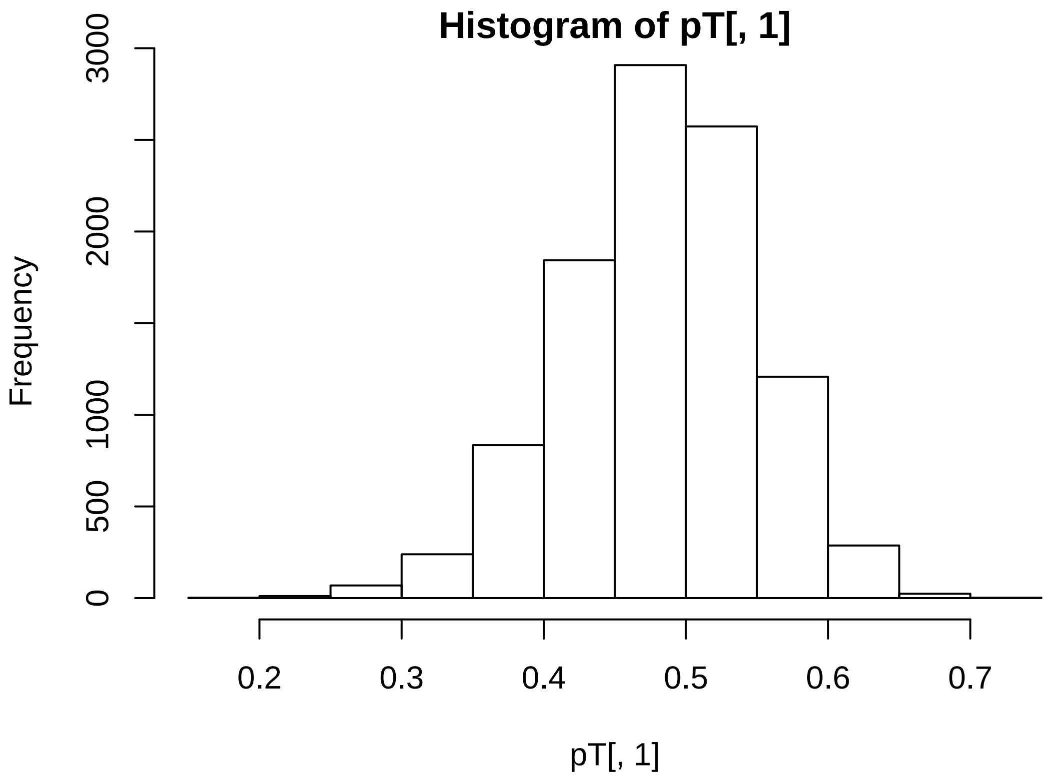
```r
m=10000
burn <- 1500
N <- m +burn
pT <- matrix(0, N, 2)
a=sample(1:100,1)
        n=100
        pT [1,  ] <- c(.5,1)
        y<-rep(NA,N)
        y[1]<-rbinom(1,n,pT[1,1])
        for (i in 2:N) {
                T <- pT[i-1, 2]
                pT [i, 1] <- rbeta(1, y[i-1], n-y+1 )
                p <- pT[i, 1]
                pT[i, 2] <- rgamma(1,2,  1/( (1/a)- log(p) )   )
                y[i]<-rbinom(1,n,p)                         }
        b <- burn + 1
        pT<- pT[b:N, ]
mean(pT[,1])    #E[P]
```

```
## [1] 0.4818253
```

```r
par(mar=c(4,4,1,2))
plot(pT, main="", cex=.5, xlab=bquote(pT[1]), ylab=bquote(pT[2]),
        ylim=(range(pT[,2])+c(-1,1)) , xlim=(range(pT[,1])+c(0,0)))
```




```r
par(mar=c(4,4,1,2))
hist(pT[,1])
```



3. Suppose we want to estimate the median of the $Gamma\,(1, \beta)$ distribution.

a. Determine analytically the value of the median as a function of β.

A gamma distribution with shape parameter $\alpha = 1$ and scale parameter β is an exponential (β) distribution.

CDF of $Exp(\beta)$

$$F_X(x) = P\,(X \leq x) = \int_0^x \beta \exp\left(-\beta t\right) \mathrm{d}t = 1 - e^{-\beta x}$$

To find the median m we need to solve this equation:

$$1 - e^{-\beta m} = \frac{1}{2}.$$

Or equivalently,

$$\begin{aligned} 1 - \left(1 - e^{-\beta m}\right) &= \frac{1}{2} \\ e^{-\beta m} &= \frac{1}{2} \\ \log\left(e^{-\beta m}\right) &= \log\left(\frac{1}{2}\right) \\ -\beta m &= -\log(2) \\ m &= \frac{\log(2)}{\beta} \end{aligned}$$

b. Write an R-function that uses bootstrap to obtain 10000 observations of the sample median (with a $\beta-$value of your choice such that $\beta > 0.5$ and construct the 90% BCa confidence interval. Did it capture the true value of the median?

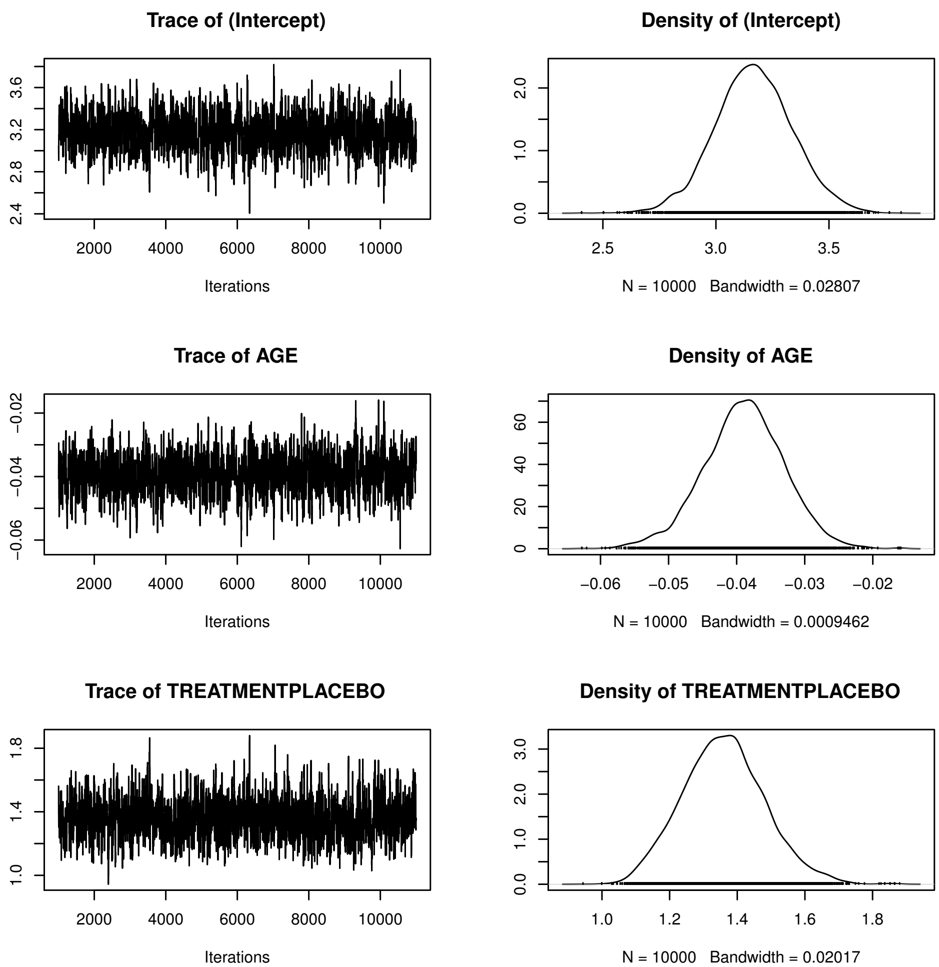
```
BootMedian<-function(B,n=10000){
        EXPmedian<-rep(NA,n)
        for (i in 1:n){EXPmedian[i]<-median(rexp(n/100,B))}
        print(mean(EXPmedian))}

BootMedian(5)
[1] 0.1397232

#True Median
log(2)/5
[1] 0.1386294

 require("bootstrap")

n=10000
 bcanon(rexp(n/10,5),n,median,alpha=c(.1,0.9))$conf
     alpha bca point
[1,]   0.1 0.1334379
[2,]   0.9 0.1478224
```

The BCa confidence interval did capture the true value of the median.

4. Using the FAP trial data, choose a function from 'MCMCpack' to run a Poisson regression model, that predicts the number of polyps based on treatment and age. Present the model summary and plots and comment on your findings.

```
require("MCMCpack")
fap<-read.csv("fap.csv",header=T)
head(fap)
  NUMBER AGE TREATMENT
1      2  16     NSAID
2     17  22     NSAID
3      1  23     NSAID
4     25  17     NSAID
5      3  23     NSAID
6     33  23     NSAID

posterior <- MCMCpoisson(NUMBER~AGE+TREATMENT , burnin = 1000, mcmc = 10000,
  data=fap)
summary(posterior)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
```

```
                      Mean       SD  Naive SE Time-series SE
(Intercept)        3.16679 0.171369 1.714e-03      0.0057783
AGE               -0.03895 0.005923 5.923e-05      0.0001983
TREATMENTPLACEBO   1.35950 0.120937 1.209e-03      0.0039742

2. Quantiles for each variable:

                      2.5%     25%      50%      75%    97.5%
(Intercept)        2.81847  3.0548  3.16620  3.27873  3.51366
AGE               -0.05137 -0.0426 -0.03883 -0.03505 -0.02765
TREATMENTPLACEBO   1.12993  1.2768  1.35688  1.43766  1.60806
```



It looks like age has very little effect on the number of polyps, however there could be an effect from the treatment. If the level of age is held constant, a unit change in treatment looks like the number of polyps increases by 1.36.