

Theoretical Exercises

1. Assume $y_i \sim L(\mu, \sigma)$. We can represent this model as a mixture of a normal likelihood and inverse-gamma priors. Thus if we let $y_i \sim N\left(\mu, \frac{4\sigma^2}{\alpha_i}\right)$ where $\sigma^2 \sim IG(a, b)$ and $\alpha_i \stackrel{iid}{\sim} IG(1, 1/2)$ for fixed hyper-parameters a and b . Using this specification and a flat prior on μ , $\pi(\mu) \propto 1$, state the full posterior and determine the conditional posterior distributions for all model parameters, μ , σ^2 , and $\alpha_1, \dots, \alpha_n$. Write out the steps of a Gibbs Sampler you could use to draw posterior samples. (Hint: the conditional posterior for α_i , with some manipulation, should be recognizable as an inverse-Gaussian.)

$$\begin{aligned} \mathcal{L}(y_i | \mu, \sigma^2, \alpha_i) &\propto \prod_{i=1}^n \left(\frac{4\sigma^2}{\alpha_i} \right)^{-\frac{1}{2}} \exp \left[-\frac{1}{2 \frac{4\sigma^2}{\alpha_i}} (y_i - \mu)^2 \right] \\ &\propto (\sigma^2)^{-\frac{n}{2}} \left[\prod_{i=1}^n \sqrt{\alpha_i} \right] \exp \left[-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 \right] \end{aligned}$$

$$\begin{aligned} P(\mu, \sigma^2, \alpha_i | y_i) &\propto (\sigma^2)^{-\frac{n}{2}} \left[\prod_{i=1}^n \alpha_i^{\frac{1}{2}} \right] \exp \left[-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 \right] (\sigma^2)^{-(a+1)} \exp \left(-\frac{b}{\sigma^2} \right) \left[\prod_{i=1}^n \alpha_i^{-2} \exp \left(-\frac{1}{2\alpha_i} \right) \right] \\ &\propto (\sigma^2)^{-\left(\frac{n}{2} + a + 1\right)} \left[\prod_{i=1}^n \alpha_i^{-\frac{3}{2}} \right] \exp \left[-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 - \frac{b}{\sigma^2} - \sum_{i=1}^n \frac{1}{2\alpha_i} \right] \end{aligned}$$

$$\begin{aligned} P(\sigma^2 | \mu, \alpha_i, y_i) &\propto (\sigma^2)^{-\left(\frac{n}{2} + a + 1\right)} \exp \left[-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 - \frac{b}{\sigma^2} \right] \\ &\propto (\sigma^2)^{-\left(\frac{n}{2} + a + 1\right)} \exp \left[-\frac{1}{\sigma^2} \left(\frac{1}{8} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 + b \right) \right] \\ \sigma^2 | \mu, \alpha_i, y_i &\sim IG \left(\frac{n}{2} + a, \frac{1}{8} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 + b \right) \end{aligned}$$

$$\begin{aligned}
p(\mu|\sigma^2, \alpha_i, y) &\propto \exp \left[-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i (y_i - \mu)^2 \right] \\
&\propto \exp \left[-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i (y_i - \bar{y} + \bar{y} - \mu)^2 \right] \\
&\propto \exp \left(-\frac{1}{8\sigma^2} \sum_{i=1}^n \alpha_i \left[(y_i - \bar{y})^2 + 2(y_i - \bar{y})(\bar{y} - \mu) + (\bar{y} - \mu)^2 \right] \right) \\
&\propto \exp \left[-\frac{1}{8\sigma^2} \left(2(\bar{y} - \mu) \sum_{i=1}^n \alpha_i (y_i - \bar{y}) + (\bar{y} - \mu)^2 \sum_{i=1}^n \alpha_i \right) \right] \quad \text{Let } \alpha_i (y_i - \bar{y}) = z_i \\
&\propto \exp \left[-\frac{n}{8\sigma^2} \left(2(\bar{y} - \mu) \bar{z} + (\bar{y} - \mu)^2 \bar{\alpha} \right) \right] \\
&\propto \exp \left[-\frac{n}{8\sigma^2} \left(2\bar{z}\mu - 2\bar{\alpha}\bar{y}\mu + \bar{\alpha}\mu^2 \right) \right] \\
&\propto \exp \left[-\frac{n}{8\sigma^2} \left(-\mu (2\bar{\alpha}\bar{y} - 2\bar{z}) + \bar{\alpha}\mu^2 \right) \right] \\
&\propto \exp \left[-\frac{n\bar{\alpha}}{8\sigma^2} \left(\mu - \frac{2\bar{\alpha}\bar{y} - 2\bar{z}}{2\bar{\alpha}} \right) - \frac{(2\bar{z} - 2\bar{\alpha}\bar{y})^2}{4\bar{\alpha}} \right] \\
&\propto \exp \left[-\frac{n\bar{\alpha}}{8\sigma^2} \left(\mu - \frac{\bar{\alpha}\bar{y} - \bar{z}}{\bar{\alpha}} \right) \right] \\
\mu|\sigma^2, \alpha_i, y &\sim N \left(\frac{\bar{\alpha}\bar{y} - \bar{z}}{\bar{\alpha}}, \frac{4\sigma^2}{n\bar{\alpha}} \right)
\end{aligned}$$

$$\begin{aligned}
p(\alpha_k | \mu, \sigma^2, \alpha_{i \neq k}, y) &\propto \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{1}{8\sigma^2} \alpha_k (y_k - \mu)^2 - \frac{1}{2\alpha_k} \right] \\
&= \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{1}{2} \left(\frac{\alpha_k^2 (y_k - \mu)^2 + 4\sigma^2}{4\alpha_k \sigma^2} \right) \right] \\
&= \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{1}{2} \left(\frac{\alpha_k^2 (y_k - \mu)^2 - 4\sigma(y_k - \mu)\alpha_k + 4\sigma^2 + 4\sigma(y_k - \mu)\alpha_k}{4\alpha_k \sigma^2} \right) \right] \\
&= \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{1}{2} \left(\frac{(y_k - \mu)^2 \left(\alpha_k - \frac{4\sigma(y_k - \mu)}{2(y_k - \mu)^2} \right)^2}{4\alpha_k \sigma^2} + \frac{4\sigma^2 - \frac{(4\sigma(y_k - \mu))^2}{4(y_k - \mu)^2} + 4\sigma(y_k - \mu)\alpha_k}{4\alpha_k \sigma^2} \right) \right] \\
&= \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{1}{2} \left(\frac{(y_k - \mu)^2 \left(\alpha_k - \frac{2\sigma}{(y_k - \mu)} \right)^2}{4\alpha_k \sigma^2} + \frac{\cancel{4\sigma^2} - \cancel{4\sigma^2} + 4\sigma(y_k - \mu)\alpha_k}{\cancel{4\sigma^2}\alpha_k} \right) \right] \\
&\propto \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{1}{2} \left(\frac{(y_k - \mu)^2 (y_k - \mu)^{-2} \left(\alpha_k - \frac{2\sigma}{y_k - \mu} \right)^2}{4\alpha_k (y_k - \mu)^{-2} \sigma^2} \right) \right] \\
&\propto \alpha_k^{-\frac{3}{2}} \exp \left[-\frac{\left(\alpha_k - \frac{2\sigma}{y_k - \mu} \right)^2}{2\alpha_k \left(\frac{2\sigma}{y_k - \mu} \right)^2} \right] \\
\alpha_k | \mu, \sigma^2, \alpha_{i \neq k}, y &\sim \text{Inverse-Gaussian} \left(\frac{2\sigma}{y_k - \mu}, 1 \right)
\end{aligned}$$

$$(\sigma^2)^{(b)} \sim IG \left(\frac{n}{2} + a, \frac{1}{8} \sum_{i=1}^n \alpha_i^{(b-1)} (y_i - \mu^{(b-1)})^2 + b \right)$$

$$\mu^{(b)} \sim N \left(\frac{\frac{\bar{y}}{n} \sum \alpha_i^{(b)} - \sum \alpha_i^{(b)} (y_i - \bar{y})}{\frac{1}{n} \sum \alpha_i^{(b)}}, \frac{4\sigma^{2(b)}}{\sum \alpha_i^{(b)}} \right)$$

$$\alpha_1^{(b)} \sim \text{Inverse-Gaussian} \left(\frac{2\sigma^{2(b)}}{y_1 - \mu^{(b)}}, 1 \right)$$

⋮

$$\alpha_n^{(b)} \sim \text{Inverse-Gaussian} \left(\frac{2\sigma^{2(b)}}{y_n - \mu^{(b)}}, 1 \right)$$

2. Suppose we wish to build a more general Bayesian model for a binomial sample. Let $X \sim \text{Binom}(N, p)$. Further, let $p \sim \text{Beta}(\alpha, \beta)$ where $\alpha \sim \text{Gamma}(a_1, b_1)$ and $\beta \sim \text{Gamma}(a_2, b_2)$. Find the likelihood, posterior, and the full conditionals. If a full conditional is recognizable, state its name. If they are not recognizable, suggest a potential proposal distribution.

$$\mathcal{L}(X|N, p,) \propto p^x(1-p)^{N-x}$$

$$P(p, \alpha, \beta|N, X) \propto p^x(1-p)^{N-x} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \alpha^{a_1-1} \exp(-\alpha b_1) \beta^{a_2-1} \exp(-\beta b_2)$$

$$\begin{aligned} P(p|\alpha, \beta, N, X) &\propto p^x(1-p)^{N-x} p^{\alpha-1}(1-p)^{\beta-1} \\ &= p^{x+\alpha-1}(1-p)^{N+\beta-x-1} \\ p &\sim \text{Beta}(x + \alpha, N + \beta - x) \end{aligned}$$

$$P(\alpha|p, \beta, N, X) \propto \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} p^\alpha \alpha^{a_1-1} \exp(-\alpha b_1)$$

$$P(\beta|p, \alpha, N, X) \propto \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)} (1-p)^\beta \beta^{a_2-1} \exp(-\beta b_2)$$

The conditionals for α and β have a form that is not exactly the kernel of a gamma distribution, but shares some similarities. For these two conditionals I would suggest a gamma for their proposal distributions.

Computing Exercises

1. The Kumaraswamy distribution is a distribution that, like the Beta, can be used to model probabilities. It has as its pdf the following:

$$p(\theta) = ab\theta^{a-1}(1 - \theta^a)^{b-1}, \theta \in (0, 1).$$

We wish to draw samples from $p(\theta)$ when $a = 2$ and $b = 2$ using a Metropolis-Hastings algorithm. Compare the following proposal densities to each other based on acceptance rate, ACF, and the resulting sampled density: $Beta(1, 1)$, $Beta(2, 1)$, $Beta(2, 2)$, and $Beta(3, 2)$. Select the proposal density you think is best out of these four and provide the criteria by which you made your selection. Set the seed to 1218 and take $B = 20000$ samples. Discard the burn-in before examining ACF and the sampled density. Without thinning, do you notice any differences between proposals?

```
## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)
## ## Copyright (C) 2003-2019 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park
## ##
## ## Support provided by the U.S. National Science Foundation
## ## (Grants SES-0350646 and SES-0350613)
## ##

Kumaraswamy <- function(x) { 2*2*x^(2-1)*(1-x^2)^(2-1) }

B <- 10000
xs <- ar <- rep(0, B)
ar <- vector("numeric", B)

Xs <- Ar <- list()

a <- c(1,2,2,3)
b <- c(1,1,2,2)

j = 2

for( j in 1:length(a)) {
  x <- .5
  xs <- ar <- rep(0, B)
  set.seed(1218)
  for( i in 2:B){
    xstar <- rbeta(1, a[j], b[j])
    rho <- ( Kumaraswamy(xstar) / Kumaraswamy(x) ) *
      ( dbeta( x , a[j], b[j] ) / dbeta( xstar, a[j], b[j] ) )
    rho <- min(1, rho )

    if ( runif(1) < rho ){ x <- xstar; ar[i] <- 1}
    xs[i] <- x }
  Xs[[j]] <- xs[-(1:(B/2))]
  Ar[[j]] <- ar
```

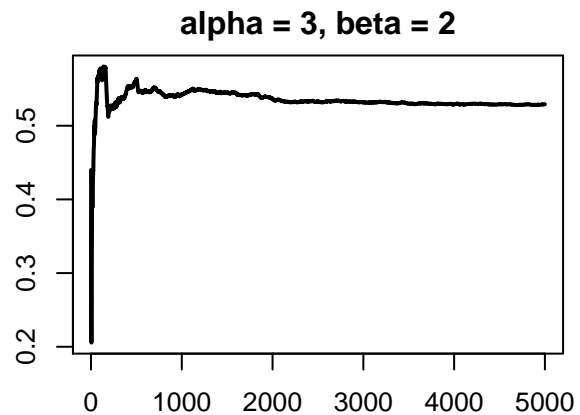
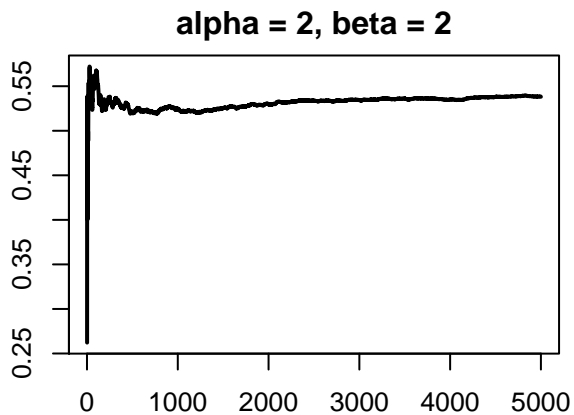
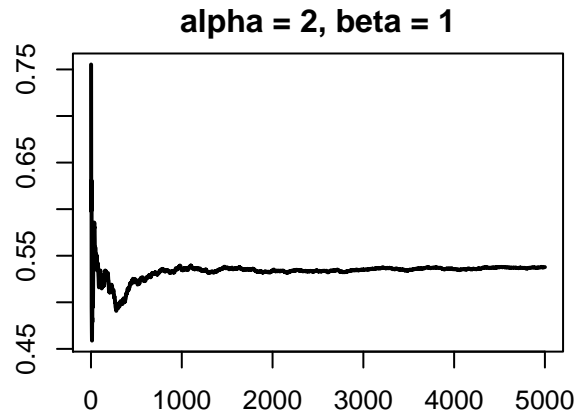
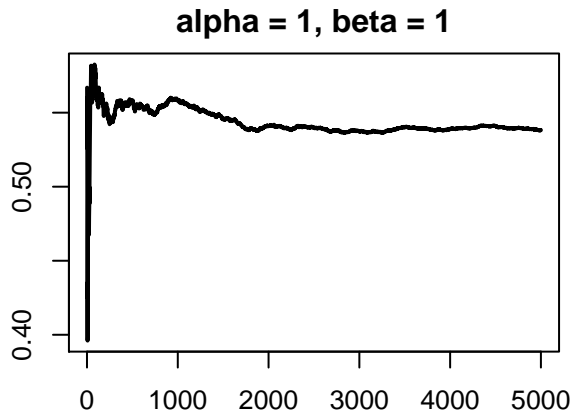
```

}

j = 4

par(mfrow=c(2,2) , mar=c(2.1,2.1,2.1,2.1) )
for(k in 1:4){ plot(cumsum(Xs[[k]] )/(1:(B/2)), type = 'l',
  ylab = 'Running Mean', xlab = 'B', lwd = 2,
  main = paste0("alpha = ",a[k], ", beta = ",b[k]) ) }

```



```

unlist( sapply( Xs, geweke.diag )[1,] )

```

```

##      var1      var1      var1      var1
## 1.683033 -1.052533 -1.543329 1.501628

```

```

sapply( Ar, function(X) sum(X)/10000 )

```

```

## [1] 0.7330 0.6714 0.9106 0.8326

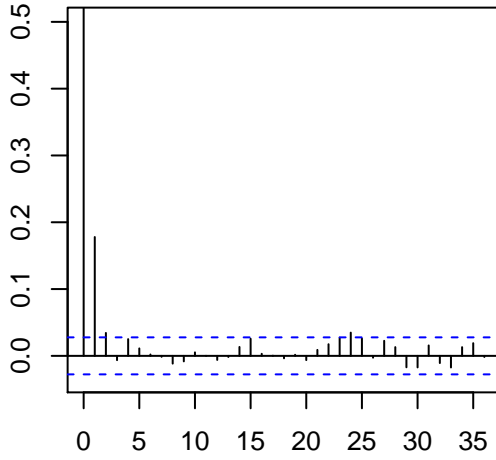
```

```

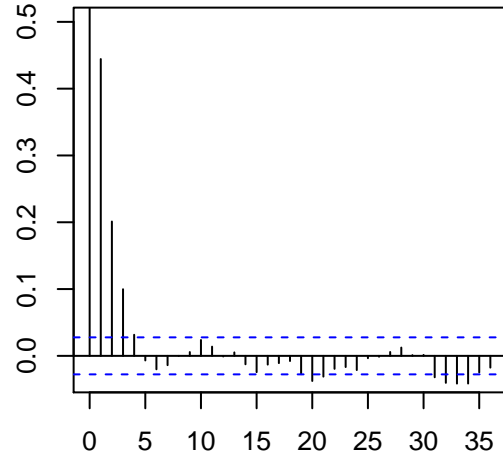
par(mfrow=c(2,2) , mar= rep(3,4) )
for( k in 1:4){ acf(Xs[[k]], main = paste0("alpha = ",a[k], ", beta = ",b[k]), ylim = c(-.1/3,.5))}

```

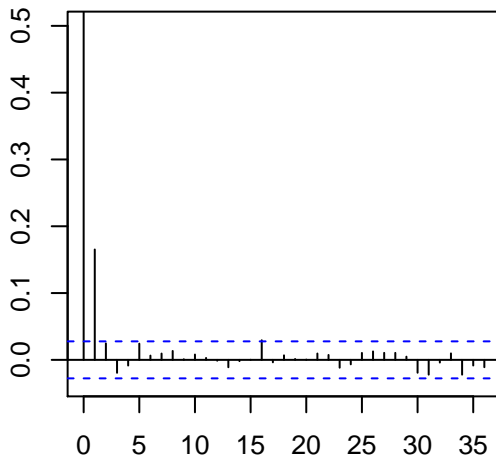
alpha = 1, beta = 1



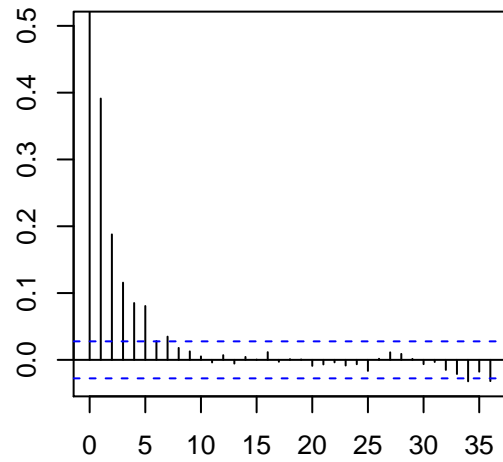
alpha = 2, beta = 1



alpha = 2, beta = 2



alpha = 3, beta = 2



They all have relatively high acceptance rates. I would focus on $\alpha = 2, \beta = 1$, but it seems to have some autocorrelation in the later samples. The next two lowest Geweke values are the ones $\alpha = 2, \beta = 2$ and $\alpha = 3, \beta = 2$. The samples with the $\beta = 3$ do appear to have high autocorrelation in the beginning and slight autocorrelation towards the end. The acceptance rate is the highest for $\alpha = 2, \beta = 2$. I would choose $\alpha = 2, \beta = 2$.

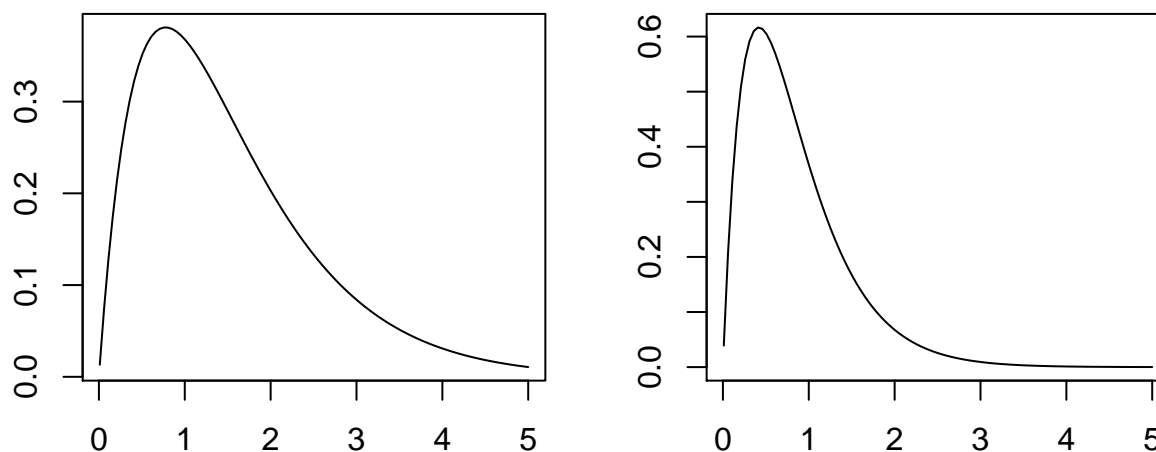
Analysis Exercises

- Returning to the binomial model from the second theoretical exercise, write a sampler to generate posterior estimates for p , α , and β . Note that you may need to use a metropolis-hastings step for some of the model components. Test your sampler by setting the seed to 1789 and generating 100 Bernoulli random variables with true probability of success 0.27, i.e. run `x <- rbinom(100, 1, 0.27)`. Discuss any acceptance rates that are needed as well as model convergence. For your test run, generate 20000 samples and set $a_1 = b_1 = a_2 = b_2 = 1$. Describe how well your model predicts the truth. (Hint: you may need the function `gamma()` in R.)

```
fx <- function( Alpha, Beta, p, N ) {
  ( gamma(Alpha + Beta) / gamma( Alpha ))*
  p^(Alpha-1) * Alpha^(1-1)* exp( -Alpha * 1 ) }

gx <- function( Alpha, Beta, p, N ) {
  ( gamma(Alpha + Beta) / gamma( Beta))*
  (1-p)^(Beta-1) * Beta^(1-1)* exp( -Beta * 1 ) }

par(mfrow=c(1,2) , mar=c(2.1,2.1,2.1,2.1) )
curve( fx( x , 1 , .75 , 1 ) , from = .01 , to = 5 )
curve( gx( 1 , x , .75 , 1 ) , from = .01 , to = 5 )
```



```
X <- rbinom(100, 1, 0.27)
N <- 1
B <- 10000*2
xs <- Ar <- Br <- rep(0, B)
x <- As <- Bs <- p <- Alpha <- Beta <- rep(NA,B);
Alpha[1] <- Beta[1] <- 1; x[1] <- 1;p[1] <- .2

set.seed(1789)
for( i in 2:B){

  Astar <- rgamma(1, Alpha[i-1], Beta[i-1] )
  rho <- ( fx(Astar, Beta[i-1], p[i-1], N ) /
```



```

      fx( Alpha[i-1] , Beta[i-1], p[i-1], N ) ) *
      ( dgamma( Alpha[i-1] , Astar, Beta[i-1] ) /
        dgamma( Astar, Alpha[i-1], Beta[i-1] ) )
rho <- min(1, rho )

if( runif(1) < rho) { Alpha[i] <- Astar; Ar[i] <-1 } else{
  Alpha[i] <- Alpha[i-1] ) }

Bstar <- rgamma(1, Alpha[i ], Beta[i-1] )
rho <- ( gx(Beta[i-1], Bstar, p[i-1], N ) /
  gx( Alpha[i-1] , Beta[i-1], p[i-1], N ) ) *
  ( dgamma( Beta[i-1] , Alpha[i ],Bstar ) /
    dgamma( Bstar,Alpha[i ], Beta[i-1] ) )
rho <- min(1, rho )

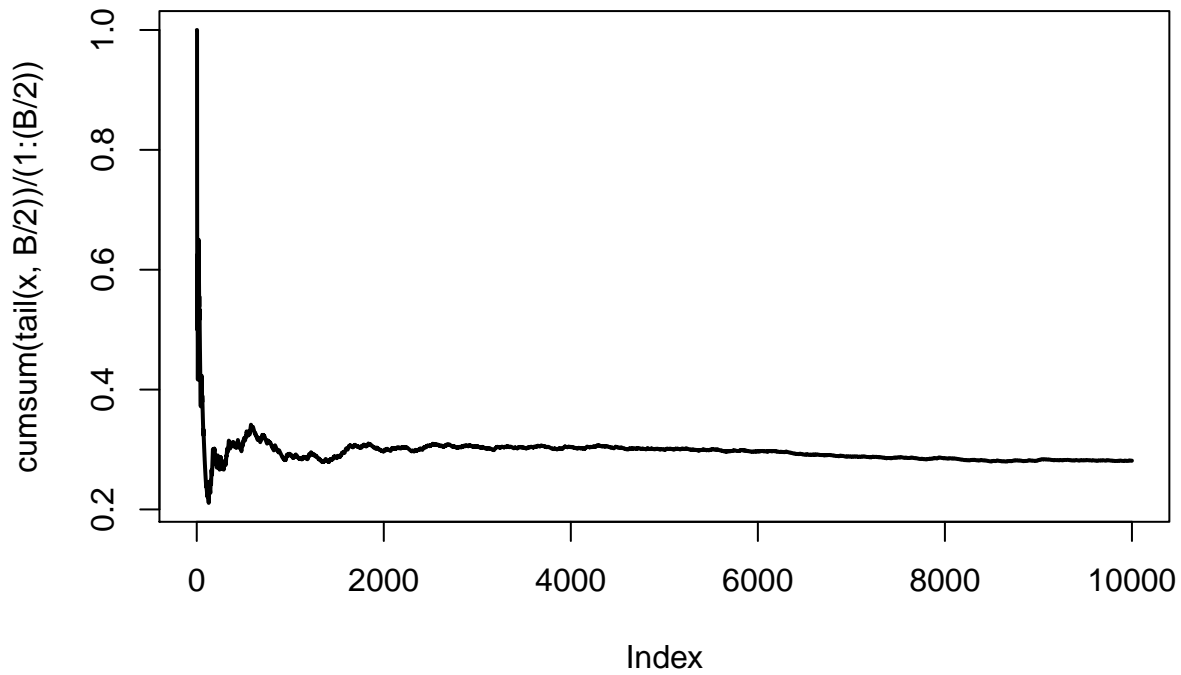
if( runif(1) < rho) { Beta[i] <- Bstar; Br[i] <-1 } else{
  Beta[i] <- Beta[i-1] ) }

Bs[i] <- Beta[i]

p[i] <- rbeta(1, x[i-1] + Alpha[i] , N + Beta[i] - x[i-1] )
x[i] <- rbinom(1,N,p[i])
}

plot(cumsum( tail( x , B/2 ) )/(1:(B/2)), type = 'l', lwd = 2 )

```



2. We are interested in estimating the country-specific variability of the risk of coup during the 1980s. Coup risk is calculated in the Rulers, Elections, and Irregular Governance or REIGN dataset¹. We are interested in modeling each nation's standard deviation in log transformed risk. The summarized data can be found in the file `coupsd.txt`. A plausible model for standard deviations is the inverse-Gaussian which has a pdf of

$$p(s_i) = \left(\frac{\lambda}{2\pi s_i^3} \right)^{1/2} \exp \left[-\frac{\lambda(s_i - \mu)^2}{2\mu^2 s_i} \right].$$

Using a joint prior of $\pi(\lambda, \mu) \propto \lambda^{-1}$, determine the posterior and find the full conditionals. The full conditional for λ should be immediately recognizable. For the full conditional of μ , implement a Metropolis-Hastings step using a normal proposal density with mean equal to the sample mean of the data and standard deviation tuned to achieve an acceptance rate between 0.23 and 0.3. Set the starting value of $\mu^{(1)}$ equal to the sample mean the starting value of $\lambda^{(1)}$ equal to the inverse of the sample standard deviation. Using a seed of 1980, take $B = 20000$ posterior samples for both parameters. After dropping the burn-in, prepare ACF plots for the remaining samples. Next, thin those remaining samples by 5 (i.e. keep every fifth sample) and generate the ACF plot. Then thin by 10 and generate the ACF plot. Finally, thin by 20 and generate the ACF plot. What do you notice each time you increase your thinning? What level of thinning attains the most desirable ACF plot? How many samples are left?

```
rm(list = ls())
coup <- read.table("coupsd.txt", header = F); colnames(coup) <- "s"
#require( "rmutil" )

B <- 20000

mu <- lambda <- s <- rep(NA, B )
mu[1] <- s[1] <- mean(coup$s)
lambda[1] <- (sd(coup$s))(-1)
Ar <- rep(0,B)
pmu <- function(M, L, S ) { exp( (-L/(2*M2)) * sum( ((S-M)2)/S)) }
n <- nrow(coup)
SD <- .1

set.seed(1980)
for( i in 2:B){

  mustar <- rnorm( 1, mu[1] , SD )
  rho <- ( pmu( mustar, lambda[i-1], coup$s ) /
    pmu( mu[i-1] , lambda[i-1], coup$s ) ) *
    ( dnorm( mu[i-1] , mu[1] , SD ) / dnorm( mustar, mu[1] , SD ) )
  rho <- min(1, rho )

  if( runif(1) < rho ){ mu[i] <- mustar; Ar[i] <- 1 } else {
    mu[i] <- mu[i-1] }

  lambda[i] <- rgamma( 1 , n/2, (1/(2*(mu[i]2))) *
    sum( ((coup$s-mu[i])2/coup$s ) )

  #s[i] <- rinvgauss( 1 , mu[i] , 1/lambda[i] )
}
```

¹Bell, Curtis. 2016. The Rulers, Elections, and Irregular Governance Dataset (REIGN). Broomfield, CO: OEF Research. Full data available at oefresearch.org.

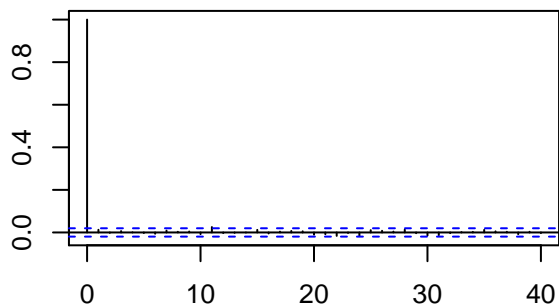
```

mu<- tail(mu,10000)
lambda<- tail(lambda,10000)

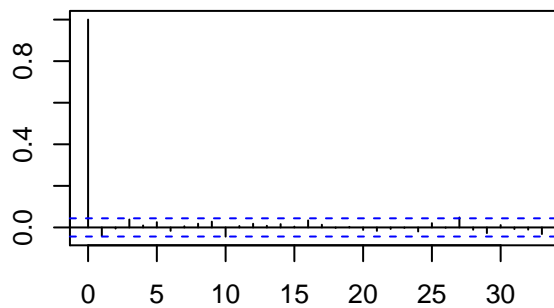
par(mfrow=c(2,2) , mar=c(3.1,2.1,3.1,2.1) )
acf(lambda , main = "lambda, no thinning")
acf(lambda[seq(1, length(lambda), 5) ] , main = "lambda, thin by 5")
acf(lambda[seq(1, length(lambda), 10)] , main = "lambda, thin by 10")
acf( lambda[seq(1, length(lambda), 20)] , main = "lambda, thin by 20")

```

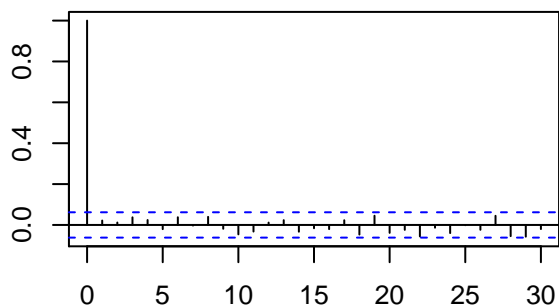
lambda, no thinning



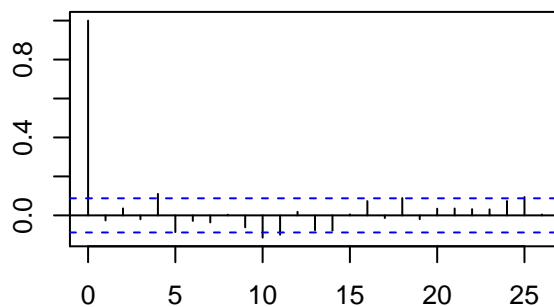
lambda, thin by 5



lambda, thin by 10



lambda, thin by 20



```
length( lambda[seq(1, length(lambda), 5)] )
```

```
## [1] 2000
```

```
length( lambda[seq(1, length(lambda), 10)] )
```

```
## [1] 1000
```

```
length( lambda[seq(1, length(lambda), 20)] )
```

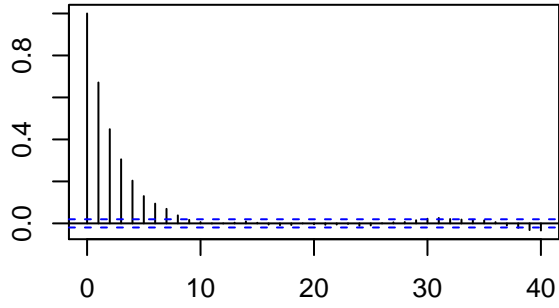
```
## [1] 500
```

```

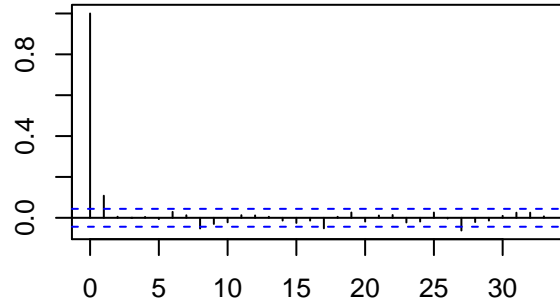
par(mfrow=c(2,2) , mar=c(3.1,2.1,3.1,2.1) )
acf(mu , main = "mu, no thinning")
acf(mu[seq(1, length(mu), 5) ] , main = "mu, thin by 5")
acf(mu[seq(1, length(mu), 10)] , main = "mu, thin by 10")
acf( mu[seq(1, length(mu), 20)] , main = "mu, thin by 20")

```

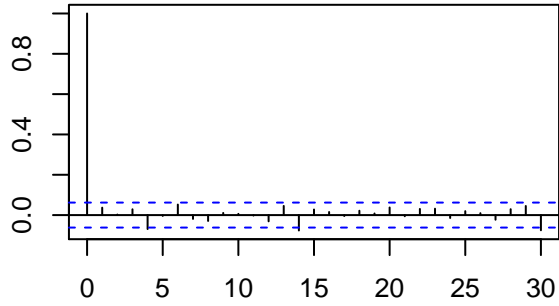
mu, no thinning



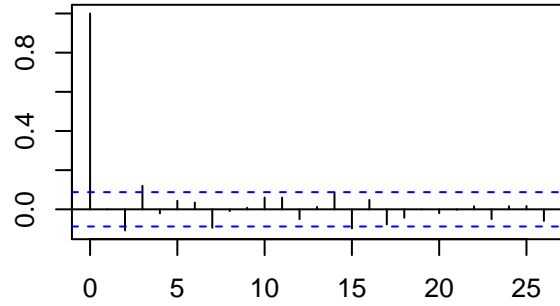
mu, thin by 5



mu, thin by 10



mu, thin by 20



It looks like for λ , it looks like the proper amount of thinning is every 10 units, there is some autocorrelation at 5 units and possibly more at 20. For μ , it appears that 20 is the proper amount, it has the least autocorrelation among the remaining choices. For λ if 10 is picked we have 1000 samples, and if 20 is picked for μ there is only 500 samples left.