Day 13 - Bayesian Day 2

David Carlson

January 12, 2021

Conjugate Priors

- These are more efficient than non-conjugate priors, and can be analyzed with direct draws from the posterior as it is of known form
- This used to be a much bigger deal; we do not really need conjugacy with the samplers we use
- When there is no conjugacy, we ignore the denominator and simulate
- There a lot of algorithms for this, but you do not really need to know them for practical purposes
- Let's start with a very simple example; drawing multivariate normally distributed variables
- Notice this is not an inference problem, but will dramatically help understanding how conditional sampling works

(Sampling from a multivariate normal distribution using Gibbs sampling) The multivariate normal distribution can be drawn directly in R using the mvtnorm package, which uses the Cholesky decomposition. The multivariate normal distribution can also be approximated using a Gibbs sampler. Consider a 3-dimensional multivariate normal with mean vector (1, 2, 3), variances each equal to 4, and an intraclass correlation coefficient of 0.3; that is,

$$Y \sim \mathcal{N}_3 \left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, 4 \times \begin{pmatrix} 1 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{pmatrix} \right)$$

- Solve for the conditional distributions of each component of Y with respect to the other two. You may use the standard decomposition of the multivariate normal, available at http://en.wikipedia.org/wiki/Multivariate_normal_distribution

$$y_1 | y_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$y_1 | y_2, y_3 \sim N \left(1 + (1.2, 1.2) \begin{pmatrix} 4 & 1.2 \\ 1.2 & 4 \end{pmatrix}^{-1} \left(\begin{pmatrix} y_2 \\ y_3 \end{pmatrix} - \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right), 4 - (1.2, 1.2) \begin{pmatrix} 4 & 1.2 \\ 1.2 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1.2 \\ 1.2 \end{pmatrix} \right)$$

$$y_2 | y_1, y_3 \sim N \left(1 + (1.2, 1.2) \begin{pmatrix} 4 & 1.2 \\ 1.2 & 4 \end{pmatrix}^{-1} \left(\begin{pmatrix} y_1 \\ y_3 \end{pmatrix} - \begin{pmatrix} 1 \\ 3 \end{pmatrix} \right), 4 - (1.2, 1.2) \begin{pmatrix} 4 & 1.2 \\ 1.2 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1.2 \\ 1.2 \end{pmatrix} \right)$$

$$y_3 | y_1, y_2 \sim N \left(3 + (1.2, 1.2) \begin{pmatrix} 4 & 1.2 \\ 1.2 & 4 \end{pmatrix}^{-1} \left(\begin{pmatrix} y_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right), 4 - (1.2, 1.2) \begin{pmatrix} 4 & 1.2 \\ 1.2 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 1.2 \\ 1.2 \end{pmatrix} \right)$$

- Write the Gibbs sampler for this distribution by sampling sequentially from each of these conditional distributions. Choose a thinning parameter, burnin factor and total number of iterations that allow you to take 1000 'non-autocorrelated' draws

```
S <- cbind(c(4,1.2,1.2),c(1.2,4,1.2),c(1.2,1.2,4))
set.seed(1934)
y <- rnorm(3)
multinorm <- function(mu, y, iter, Sigma){
```

```
  S <- Sigma
  s <- S [1,-1]%*%solve(S[-1,-1])%*%S[-1,1]
  M <- S[1,-1]%*%solve(S[-1,-1])
  y.final <- NULL
  for (i in 1:iter){
  y[1] <- rnorm(1, mu[1]+M%*%(y[-1]-c(mu[2],mu[3])), sqrt(4-s))
  y[2] <- rnorm(1, mu[2]+M%*%(y[-2]-c(mu[1],mu[3])), sqrt(4-s))
  y[3] <- rnorm(1, mu[3]+M%*%(y[-3]-c(mu[1],mu[2])), sqrt(4-s))
  y.final <- rbind(y.final, y)
  }
  return(y.final)
}
out <- multinorm(mu=c(1,2,3), y=y, iter=15000, Sigma=S)
out <- out[-(1:5000),]
out2 <- out[seq(1,10000,by=10),]

par(mfrow=c(3,1))
acf(out2[,1], main=expression(y[1]))
acf(out2[,2], main=expression(y[2]))
acf(out2[,3], main=expression(y[3]))
```
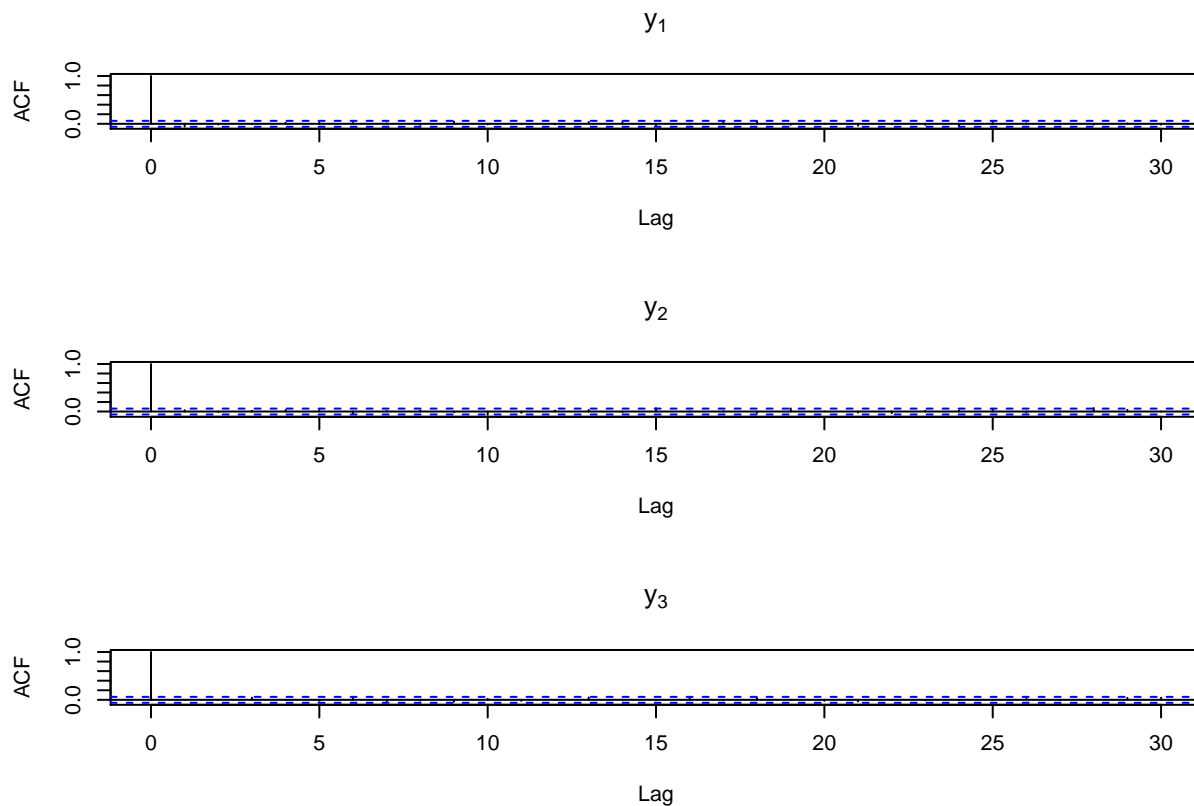

- Use the mcmc() function in the R package coda to summarize your generated sequence. Report the mean and associated Monte Carlo approximation error.

```
library(coda)
mcmc.object <- mcmc(out, start=5001, end=15000, thin=10)
summary(mcmc.object)
```

```
##
```

```
## Iterations = 5001:14991
## Thinning interval = 10
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##        Mean     SD Naive SE Time-series SE
## [1,] 0.9472 2.057  0.06504        0.07881
## [2,] 2.0178 2.081  0.06581        0.07464
## [3,] 2.9823 2.123  0.06714        0.08676
##
## 2. Quantiles for each variable:
##
##        2.5%     25%    50%    75% 97.5%
## var1 -3.113 -0.4235 0.9235 2.441 4.929
## var2 -2.105  0.5954 2.0195 3.484 5.950
## var3 -1.328  1.5800 3.0402 4.445 7.114
```

```r
par(mfrow=c(3,1))
traceplot(mcmc.object)
```
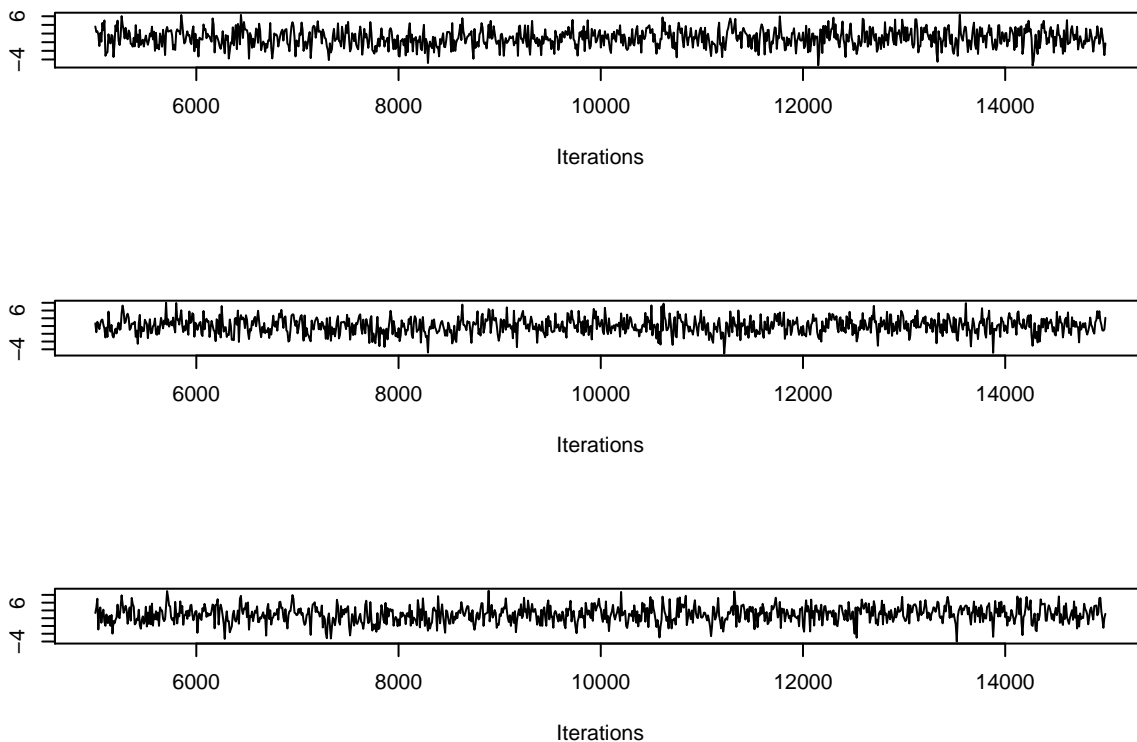

- The Gibbs Sampler (for a simple normal model)

- iter ← 1000

- burnin ← 100

- declare β_0

- declare β_1

- declare ϕ

- $\beta_0[1] \leftarrow beta_1[1] \leftarrow \phi[1] \leftarrow$ arbitray value
- for i = 2: iter do
- $\phi[i] \sim p(\phi|\phi = \phi[i-1], \beta_0 = \beta_0[i-1], \beta_1 = \beta_1[i-1])$ ++ $\beta_0[i] \sim p(\beta_0|\phi = \phi[i], \beta_1 = \beta_1[i-1], \tau_0, \mu_0)$
- $\beta_1[i]\ simp(\beta_1|\phi = \phi[i], \beta_0 = \beta_0[i], \tau_0, \mu_0)$
- return $\phi[burnin:iter], \beta_0[burnin:iter], \beta_1[burnin:iter]$
- Now let's move on to an actual model, specifically a Poisson
- (Poisson Regression) Haberman (1978) considers an experiment involving subjects reporting one stressful event. The collected data are $y_1, \ldots, y_{18}$, where y_i is the number of events recalled i months before the interview. Suppose y_i is Poisson distributed with mean λ_i , where the $\{\lambda_i\}$ satisfy the loglinear regression model:

$$\log \lambda_i = \beta_0 + \beta_1 i$$

The data are given by Months $1 - 18$, $y_i = (15, 11, 14, 17, 5, 11, 10, 4, 8, 10, 7, 9, 11, 3, 6, 1, 1, 4)$

- Show that the logarithm of the posterior density is given, up to an additive constant, by

$$\log g(\beta_0, \beta_1 | data) = \sum_{i=1}^{18}[y_i(\beta_0 + \beta_1 i) - \exp(\beta_0 + \beta_1 i)]$$

$$p(\lambda|y) \propto \exp(-\sum_{i=1}^{18} \lambda_i) \prod_{i=1}^{18} \lambda_i^{y_i}$$

$$\log(p(\lambda|y)) \propto \log(\exp(-\sum_{i=1}^{18} \lambda_i) \prod_{i=1}^{18} \lambda_i^{y_i})$$

$$\propto -\sum_{i=1}^{18} \lambda_i + \log(\prod_{i=1}^{18} \lambda_i^{y_i})$$

$$\propto -\sum_{i=1}^{18} \exp(\beta_0 + \beta_1 i) + \sum_{i=1}^{18} y_i(\beta_0 + \beta_1 i)$$

$$\propto \sum_{i=1}^{18}(y_i(\beta_0 + \beta_1 i) - \exp(\beta_0 + \beta_1 i))$$

- Find the posterior mean and standard deviation of β_1 by the Metropolis-Hastings algorithm using the proposal being the normal approximation given by the function laplace() in R package LearnBayes.

```r
library(LearnBayes)
set.seed(1919)
log.post <- function(data, params){
  sum(sapply(1:length(data$data), function(x) y[x]*(params[1] + (params[2]*x))-exp(params[1]+(params[2]
}
prop.f<-function(data){
  laplace(log.post, mode=start, data=data)$mode
}
met <- function(data, start, iters){
    chain <- array(dim = c(iters+1,length(start)))
    chain[1,] <- start
    for (i in 1:iters){
        proposal <- prop.f(data)
```

```
        prob <- exp(log.post(data, proposal)) - log.post(data,chain[i,])
        if (runif(1) < prob){
            chain[i+1,] <- proposal
        }else{
            chain[i+1,] <- chain[i,]
        }
    }
    return(chain)
}
y <- c(15,11,14,17,5,11,10,4,8,10,7,9,11,3,6,1,1,4)
start <- rnorm(2)
data <- list(data=y, par=start)
out <- met(data,start, 1000)
mean(out[,2])
```

[1] -0.08400987

```
sd(out[,2])
```

[1] 0.009363812

```
#lambdas
exp(mean(out[,1])+mean(out[,2])*1:18)
```

[1] 15.115528 13.897552 12.777717 11.748117 10.801479 9.931120 9.130892
[8] 8.395144 7.718682 7.096727 6.524889 5.999127 5.515731 5.071285
[15] 4.662652 4.286946 3.941513 3.623915

- Suppose your prior distribution for β, the proportion of Californians who support the death penalty is beta(α, β) with mean 0.6 and standard deviation 0.3.
- Determine the parameters α and β of your prior

$$E[x] = \frac{\alpha}{\alpha + \beta} = .6$$

$$\beta = \frac{\alpha}{.6} - \alpha = \frac{2\alpha}{3}$$

$$var(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = .09$$

$$\frac{\alpha\frac{2\alpha}{3}}{(\alpha + \frac{2\alpha}{3})^2(\alpha + \frac{2\alpha}{3} + 1)} = .09$$

$$\frac{\frac{2\alpha^2}{3}}{\frac{25\alpha^2}{9}(\frac{5\alpha}{3} + 1)} = .09$$

$$\alpha = 1$$

$$\beta = \frac{2}{3}$$

- A random sample of 1000 Californians is taken, and 65% support the death penalty. What are your posterior mean and variance for θ? For comparison, draw the prior and posterior density function in the same plot using R

$$p(\theta|data) \propto (data|\theta)p(\theta)$$
$$\propto \theta^{650}(1-\theta)^{350}(1-\theta)^{-1/3}$$
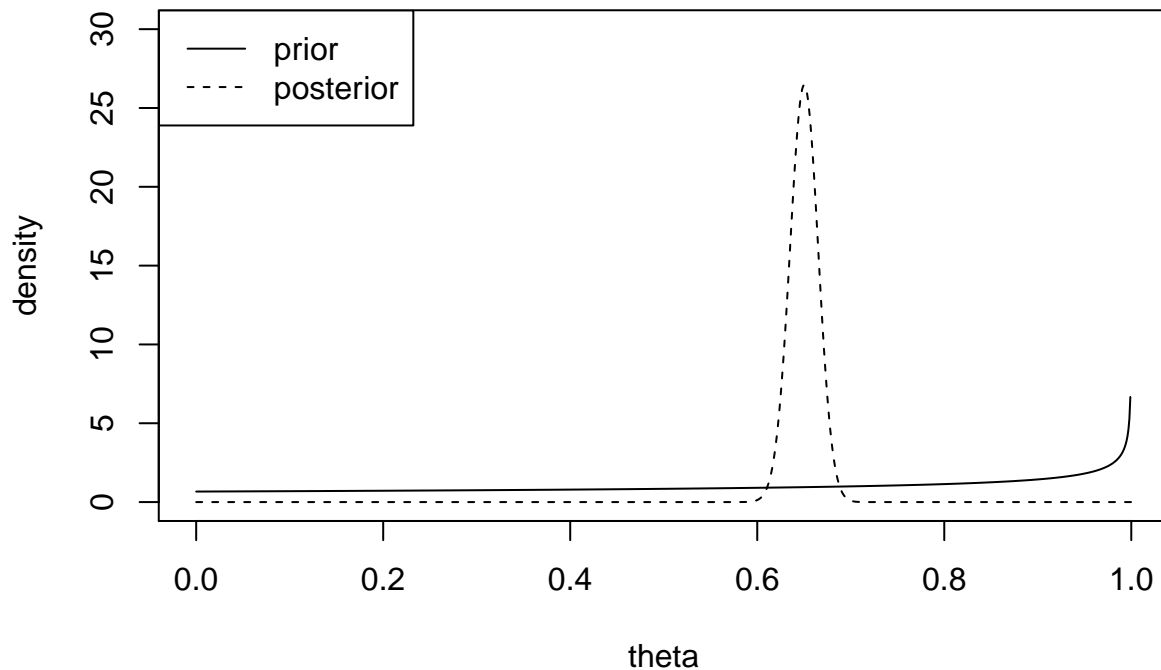$$\propto \theta^{650}(1-\theta^{350-1/3})$$
$$\propto Beta(651, 350 + 2/3)$$

$$E[x] = \frac{\alpha}{\alpha + \beta} \approx .6499$$

$$var(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \approx .0002269$$

```r
theta <- seq(0,1,by=.001)
plot(theta, dbeta(theta,1,2/3), type='l', ylab='density', ylim=c(0,30))
lines(theta, dbeta(theta, 651, 350+2/3), lty=2)
legend('topleft', c('prior', 'posterior'), lty=c(1,2))
```


- Exponential model with conjugate prior distribution
- Show that if $y_i|\theta, i = 1, \ldots, n$ are i.i.d. exponentially distributed with rate θ, i.e. the pdf is

$$f(y_i|\theta) = \theta e^{-\theta y_i},$$

then the conjugate prior of θ is a gamma distribution.

$$p(y_i|\theta) = \theta e^{-\theta y}$$
$$\theta \sim Gamma(\alpha, \beta)$$
$$p(y_1, \ldots, y_n|\theta) = \theta^n e^{-\theta \sum_{i=1}^{n} y_i}$$
$$p(\theta|y) \propto p(y_1, \ldots y_n|\theta)p(\theta)$$
$$\propto \theta^n e^{-\theta \sum_{i=1}^{n} y_i} \theta^{\alpha-1} e^{-\beta\theta}$$
$$\propto \theta^{n+\alpha-1} e^{-\theta(\beta + \sum_{i=1}^{n} y_i)}$$

$$\therefore \theta|y \sim Gamma(n + \alpha, \beta + \sum_{i=1}^{n} y_i)$$

- The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. A random sample of light bulbs is to be tested and the lifetime of each obtained. Suppose that the prior information tells the coefficient of variation is 0.5. (The coefficient of variation is the standard deviation divided by the mean.) Suppose that we would like to reduce the coefficient of variation to 0.1 in the posterior. How many light bulbs need to be tested? Use the conjugate prior for this problem

$$\frac{\sqrt{\alpha\beta^2}}{\alpha\beta} = .5$$
$$\frac{1}{\sqrt{\alpha}} = .5$$
$$\alpha = 4$$

$$\frac{1}{\sqrt{n+\alpha}} = .1$$
$$\frac{1}{\sqrt{n+4}} = .1$$
$$n + 4 = 100$$
$$n = 96$$