Tutorial 8 Solutions

STAT 3013/4027/8027

---

1. Write out Example B in 8.6.

We are interested in modeling data where:

$$X_1, \ldots, X_2 \overset{\text{iid}}{\sim} \text{normal}(\theta, \xi)$$

$$f_X(x \big| \theta, \xi) = \left(\frac{\xi}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2}\xi(x-\theta)^2\right)$$

Where $\xi = \frac{1}{\sigma^2}$. As we are considering Bayesian inference, we need to have priors on both parameters (which are considered random in this framework). Here we will model the priors as being independent.

$$p(\theta, \xi) = p(\theta)p(\xi)$$

The prior for θ is:

$$\theta \sim \text{normal}(\theta_0, \xi_{prior})$$

and the prior for ξ is:

$$\xi \sim \text{gamma}(\alpha, \lambda)$$

- For the first case let's consider that ξ is know $\xi = \xi_0$. This leads to the following posterior distribution:

$$
\begin{aligned}
p(\theta \big| \boldsymbol{x}, \xi_0) &\propto p(\boldsymbol{x} \big| \theta, \xi_0) p(\theta) \\
&\propto \exp\left(-\frac{1}{2}\left[\xi_0 \sum (x_i - \theta)^2 + \xi_{prior}(\theta - \theta_0)^2\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\xi_0 \sum x_i^2 - 2\theta\xi_0 \sum x_i + \theta^2 n\xi_0 + \xi_{prior}\theta^2 - 2\theta\theta_0\xi_{prior} + \xi_{prior}\theta_0^2\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\theta^2(n\xi_0 + \xi_{prior}) - 2\theta(\xi_0 \sum x_i + \theta_0\xi_{prior})\right]\right) \\
&= exp\left(-\frac{1}{2}\left[\theta^2 a - 2\theta b\right]\right) \\
&= exp\left(-\frac{1}{2}a\left[\theta^2 - 2\theta b/a + b^2/a^2 - b^2/a^2\right]\right) \\
&\propto exp\left(-\frac{1}{2}a\left[\theta^2 - 2\theta b/a + b^2/a^2\right]\right) \\
&= exp\left(-\frac{1}{2}a\left[\theta^2 - b/a\right]^2\right)
\end{aligned}
$$

We see that the posterior for θ is proportional to a normal distribution with a variance of:

$$v* = 1/a = (n\xi_0 + \xi_{prior})^{-1}$$

and a mean of:

$$m^* = b/a = \frac{(\xi_0 \sum x_i + \theta_0\xi_{prior})}{(n\xi_0 + \xi_{prior})}$$

- Now let's consider the case where θ is known $\theta = \theta_0$:

$$
\begin{aligned}
p(\theta \big| \boldsymbol{x}, \xi_0) &\propto p(\boldsymbol{x} \big| \theta, \xi_0) p(\theta) \\
&\propto \xi^{n/2} \exp\left(-\frac{1}{2}\xi\left[\sum (x_i - \theta_0)^2\right]\right) \xi^{\alpha-1} \exp(-\lambda\xi) \\
&\propto \xi^{\alpha+n/2-1} \exp\left(-\left[\frac{1}{2}\sum (x_i - \theta_0)^2 + \lambda\right]\xi\right)
\end{aligned}
$$

So we see the posterior for ξ is proportional to a gamma distribution with parameters:

$$a^* = \alpha + n/2 \qquad b^* = \left[\frac{1}{2}\sum(x_i - \theta_0)^2 + \lambda\right]$$

2. Based on Section 8.6.3 and using the GDP 2013 data (take the log of the data), in R code the Gibbs sampling procedure. Let it run for 1,000 iterations. Note: The Gibbs sampling procedure is a Metropolis algorithm that accepts with probability 1.

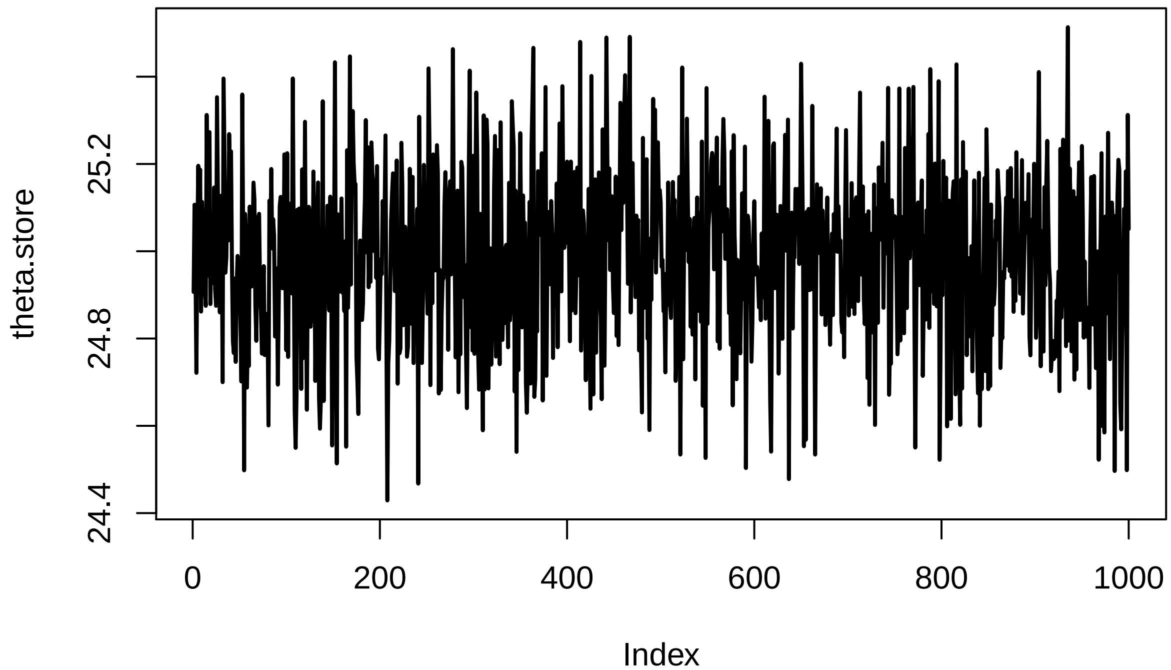
- Using the results derived above, we can construct a Markov chain in the parameters through the full conditional distributions:
    1. Set values for the prior parameters: $\theta_0 = 0, \xi_{prior} = 0.0001, \alpha = 1, \lambda = 1$. You can try other values for the priors parameters.
    2. Set a starting value for $\xi = 1$.
    3. Generate a random draw for θ from $[\theta|\boldsymbol{x}, \xi]$.
    4. Generate a random draw for ξ from $[\xi|\boldsymbol{x}, \xi]$.
    5. Repeat steps 3 & 4 until convergence of the Markov chain, and continue until you have enough samples from the join posterior.

```
x <- read.csv("gdp2013.csv")
x <- log(na.omit(x$X2013))
n <- length(x)

##
theta.0 <- 0
xi.prior <- 0.0001
alpha <- lambda <- 1

##
theta.store <- NULL
xi.store <- NULL

##
xi <- 1
```

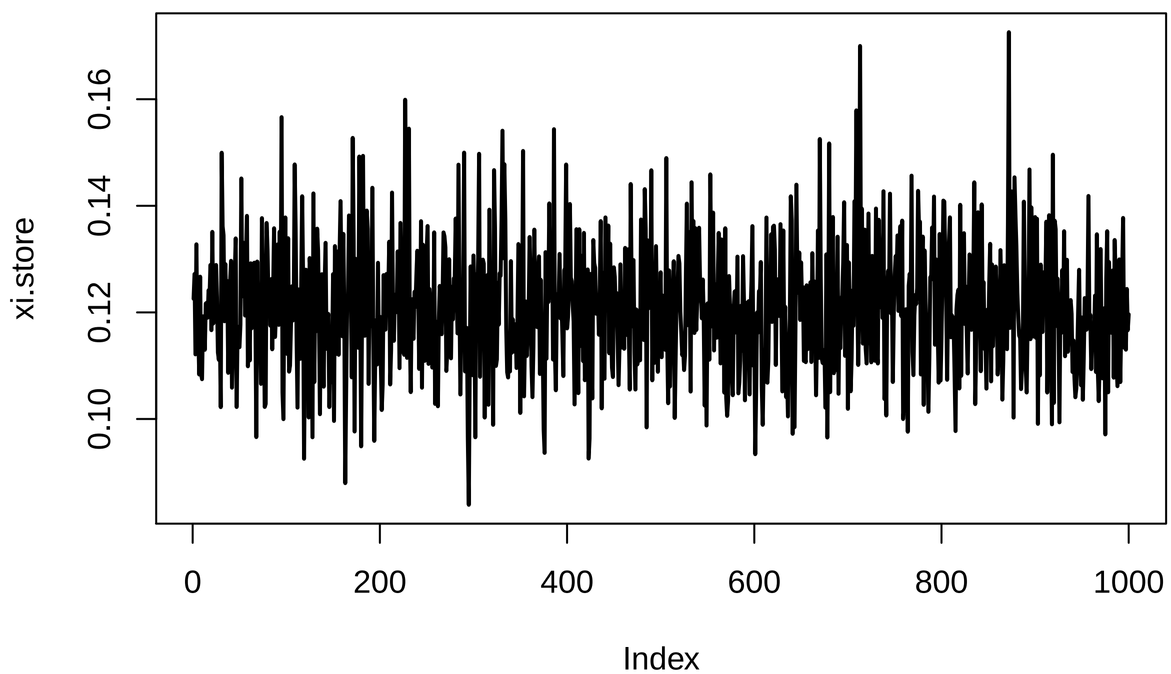
```r
## Start the chain
S <- 1000
for(s in 1:S){

##
v <- 1/(n*xi + xi.prior)
m <- (xi *sum(x) + theta.0*xi.prior)/(n*xi + xi.prior)
theta <- rnorm(1, m, sqrt(v))

##
a <- alpha + n/2
b <- sum( (x-theta)^2)/2 + lambda
xi <- rgamma(1, a, b)


theta.store <- c(theta.store, theta)
xi.store <- c(xi.store, xi)
}
```

- Let's first examine the trace plots to look for signs of non-convergence and poor mixing:

```r
plot(theta.store, type="l", lwd=2)
```



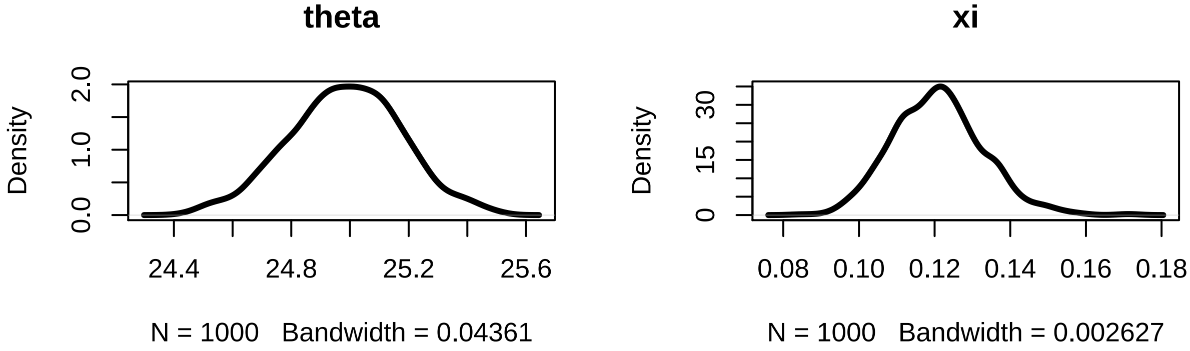

```r
plot(xi.store, type="l", lwd=2)
```

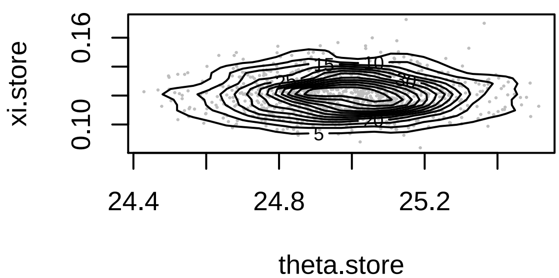


From the figures, it appears that the chains converged and are mixing well. Let's examine the marginal densities and the joint density:

```r
par(mfrow=c(2,2))
library(LaplacesDemon)
plot(density(theta.store), lwd=3, main="theta")
```

```r
plot(density(xi.store), lwd=3, main="xi")
joint.density.plot(theta.store, xi.store, contour=TRUE)
```





Considering we have data coming from a normal distribution and we know the sample mean and variance ($\bar{X}$ and S^2) are independent, it should not be surprising that the joint posterior between θ and ξ suggests independence. Finally let's get the mean and variance of the marginal posteriors:

```r
mean(theta.store)
```

```
## [1] 24.9845
```

```r
var(theta.store)
```

```
## [1] 0.0372147
```

```r
mean(xi.store)
```

```
## [1] 0.120894
```

```r
var(xi.store)
```

```
## [1] 0.0001424202
```

3. Answer question 53 in Chapter 8. Let's consider the following data and model:

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{uniform}(0, \theta)$$

.

a. For the Method of Moments estimator, we want to set the distributional first moment (the mean) equal to the sample first moment:

$$\begin{aligned} E[X] = \frac{\theta + 0}{2} &= \bar{X} \\ \tilde{\theta} &= 2\bar{X} \end{aligned}$$

- Now let's get the mean of the estimator:

$$\begin{aligned} E[\tilde{\theta}] = E[2\bar{X}] &= 2E[X] \\ &= 2\frac{\theta + 0}{2} = \theta \end{aligned}$$

We can see that the estimator is unbiased.

- Let's get the variance of the estimator:

$$\begin{aligned} V[\tilde{\theta}] = V[2\bar{X}] &= \frac{4}{n}V[X] \\ &= \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n} \end{aligned}$$

b. Now let's consider the MLE. Let's get the likelihood:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}$$

Let's try our standard approach, differentiate the log likelihood and set it equal to zero.

$$
\begin{aligned}
\ell(\theta) &= -nlog(\theta) \\
\ell'(\theta) &= -n/\theta = 0 \\
&\Rightarrow \frac{1}{\theta} = 0
\end{aligned}
$$

We see that this has no solution. Let's go back to the likelihood:

$$
L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}
$$

We know we want to make this as large as possible to maximize it. This suggests that we need to make θ as small as possible. But given a set of $x_1, \ldots, x_n$ and knowing that θ can not be smaller than any of those value we find the maximum of the likelihood to be:

$$
\hat{\theta} = max(X_1, \ldots, X_n)
$$

c. Let's determine the distribution of the MLE. Let's use the CDF method for the transformation (See Rice Section 3.7). Note: If the maximum value of X is less the c, then all values of X are less than c. Also the CDF of a uniform(a,b) is $\frac{x-a}{b-a}$.

$$
\begin{aligned}
P(\hat{\theta} < c) &= P(max(X_1, \ldots, X_n) < c) \\
&= P(X_1 < c, X_2 < c, \ldots, X_n < c) \\
&= P(X_1 < c) \times \cdots \times P(X_n < c) \\
&= \frac{c}{\theta} \times \cdots \times \frac{c}{\theta} \\
&= \left(\frac{c}{\theta}\right)^n \Rightarrow \left(\frac{x}{\theta}\right)^n
\end{aligned}
$$

Now let's differentiate this to get the pdf:

$$
f_{\hat{\theta}}(x) = \frac{d}{dx}\left(\frac{x}{\theta}\right)^n = n\left(\frac{x^{n-1}}{\theta^n}\right); \quad 0 \leq x \leq \theta
$$

- Now let's get the mean:

$$
\begin{aligned}
E[\hat{\theta}] &= \int_0^\theta xn\left(\frac{x^{n-1}}{\theta^n}\right)dx \\
&= \frac{n}{\theta^n}\int_0^\theta x^n dx \\
&= \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta
\end{aligned}
$$

We see that the MLE is biased.

- Now let's get the variance:

$$
\begin{aligned}
V[\hat{\theta}] &= E(X^2) - [E(X)]^2 \\
&\Rightarrow E(X^2) = \int_0^\theta x^2 n\left(\frac{x^{n-1}}{\theta^n}\right)dx \\
&\quad = \frac{n}{\theta^n}\int_0^\theta x^{n+1}dx \\
&\quad = \frac{n}{\theta^n}\frac{\theta^{n+2}}{n+2} \\
&= \frac{n}{n+2}\theta^2
\end{aligned}
$$

So the variance is:

$$
\begin{aligned}
V[\hat{\theta}] &= \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 \\
&= \theta^2\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) \\
&= \theta^2\left(\frac{n}{(n+2)(n+1)^2}\right)
\end{aligned}
$$

- The MSE for the MOM is $\frac{\theta^2}{3n}$.
- The MSE for the MLE is:

$$
\begin{aligned}
MSE[\hat{\theta}] &= V(\hat{\theta}) + Bias(\hat{\theta})^2 \\
&= \theta^2 \left(\frac{n}{(n+2)(n+1)^2} \right) + \left[\frac{n}{n+1}\theta - \theta \right]^2 \\
&= \frac{2\theta^2}{(n+2)(n+1)}
\end{aligned}
$$

- For $n > 2$ the MSE of the MLE is smaller than the MSE for the MoM.

d. To make the MLE unbiased consider the following estimator:

$$\hat{\gamma} = \frac{n+1}{n}\hat{\theta}$$

$$E[\hat{\gamma}] = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta$$