MATH 680 - Assignment #4

Kevin McGregor

November 30th, 2015

Question 1

(a)

In this problem we have $f(\beta) = \|y - X\beta\|^2$, $g(\beta) = R^\top \beta$, $R = (0, 0, \ldots, 1)^\top$. As seen in earlier written assignments, $\nabla f(\beta) = -2X^\top y + 2X^\top X\beta$. Similarly, $\nabla g(\beta) = R$. So we wish to find $(\bar{\beta}, u)$ such that:

$$
\begin{aligned}
-X^\top y + X^\top X\bar{\beta} + \frac{u}{2}R &= 0 \\
R^\top \bar{\beta} &\leq 0 \\
uR^\top \bar{\beta} &= 0 \\
u &\geq 0.
\end{aligned}
$$

We then have that $\bar{\beta} = (X^\top X)^{-1}(X^\top y - \frac{u}{2}R)$. Plugging this into the third KKT condition gives,

$$
\begin{aligned}
uR^\top (X^\top X)^{-1}(X^\top y - \frac{u}{2}R) &= 0 \\
u &= \frac{2R^\top (X^\top X)^{-1}(X^\top y)}{R^\top (X^\top X)^{-1}R},
\end{aligned}
$$

provided $2R^\top (X^\top X)^{-1}(X^\top y) > 0$. Otherwise, we have $u = 0$. So, if OLS finds a negative value of β_p, then nothing more needs to be done. If OLS finds a positive value for β_p, then choosing u as stated above will ensure the second KKT condition.

(b)

The R function for this is called fitBetaNeg, and is found in the file fitBetaNeg.R. Testing for parts (b) and (c) is found in the file a4_q1.R. The KKT point is

$$(\bar{\beta}, u) = \left(\begin{bmatrix}4.9 & 4.1 & 2.5 & 2.4 & 0\end{bmatrix}^\top, 122.6177\right)$$

. The KKT conditions are met at this point.

(c)

The KKT point is

$$(\bar{\beta}, u) = \left(\begin{bmatrix}3.6 & 3.3 & 2.1 & 1.6 & -1.9\end{bmatrix}^\top, 0\right)$$

. The KKT conditions are met at this point.

Question 2

(a)

We'll take our proposal distribution to be:

$$g(x) = \frac{\beta^a x^{a-1}}{\Gamma(a)} e^{-x},$$

where $a = \lfloor \alpha \rfloor$. Then we have that,

$$\begin{aligned} \frac{f(x)}{g(x)} &= \frac{\beta^{\alpha-a} x^{\alpha-a}}{\Gamma(\alpha)} \Gamma(a) e^{-(\beta-1)x} \\ &\propto x^{\alpha-a} e^{-(\beta-1)x}, \end{aligned}$$

and so,

$$\nabla \frac{f(x)}{g(x)} \propto (\alpha - a) x^{\alpha-a-1} e^{-(\beta-1)x} - x^{\alpha-a}(\beta-1) e^{-(\beta-1)x}.$$

Setting the derivative to zero results in $x = \frac{\alpha-a}{\beta-1}$. We can then conclude that:

$$\begin{aligned} \frac{f(x)}{g(x)} &\leq \frac{\beta^{\alpha-a}}{\Gamma(\alpha)} \left(\frac{\alpha-a}{\beta-1} \right)^{\alpha-a} \Gamma(a) e^{-(\alpha-a)} \\ &= c. \end{aligned}$$

So, we just draw a sample of size $\lfloor \alpha \rfloor$ from $Exp(1)$, using the inverse distribution function $-\log(1-x)$. The sum of these random variables, Z, will follow $Gamma(\lfloor \alpha \rfloor, 1)$. Then we draw $U \sim Unif(0,1)$ and accept if $U < \frac{f(Z)}{cg(Z)}$. This will generate a sample from $Gamma(\alpha, 1)$. Then, multiply the observation by $1/\beta$ to get $Gamma(\alpha, \beta)$.

(b)

See R function gammaSamp.R.

(c)

The code to set up the simulation is found in the R file a4_q2c.R. I ran the algorithm with $\alpha \in \{1.5, 5.3, 10.9, 50.2, 75.2\}$, and $\beta = 2$. The estimated acceptance probabilites are found in Table 1.

α	Prob
1.5	0.90
5.3	0.65
10.9	0.66
50.2	0.03
75.2	0.01

Table 1: Estimated acceptance probabilities for various values of α.

Question 3

(a)

$$\begin{aligned} f(\theta|x) &= \frac{f(x|\theta)f(\theta)}{f(x)} \\ &= \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2}}\frac{1}{\pi(1+\theta^2)}}{\int \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2}}\frac{1}{\pi(1+\theta^2)}d\theta} \\ &\propto e^{-\frac{(x-\theta)^2}{2}}\frac{1}{\pi(1+\theta^2)}. \end{aligned}$$

(b)

Let

$$g(\theta) = \frac{1}{\pi s\left(1+\left(\frac{\theta-m}{s}\right)\right)}$$

. We have that,

$$\frac{\tilde{f}(\theta)}{g(\theta)} = e^{-\frac{(x-\theta)^2}{2}}\frac{s\left(1+\left(\frac{\theta-m}{s}\right)\right)}{1+\theta^2}$$

If we choose $m = 0$ and $s = 1$, then we have that:

$$\frac{\tilde{f}(\theta)}{g(\theta)} \leq 1$$

So the algorithm proceeds as follows:

1. Draw a sample, Z, from $Cauchy(0, 1)$
2. If $U < \frac{\tilde{f}(\theta)}{g(\theta)}$, where $U \sim Unif(0, 1)$, then accept the sample. Reject otherwise.

(c)

The function is called normCauchyPrior, and it is found in the file normCauchyPrior.R. I also ran a simulation to test the function, found in a4_q3.R. I take a sample of size 100 from this posterior distribution and calculate the expected value. I replicated this 1000 times, and the mean of the expected values was 0.5570, and the variance, 0.0054.

(d)

Since we do not know the normalizing constant, we need to use weighted importance sampling. So, if we take a random sample Z_i, $i = 1, \ldots, n$ from g, this estimator takes the form:

$$\hat{E}(\theta|X = x) = \frac{\sum_{i=1}^{n} z_i e^{-\frac{(z_i-x)^2}{2}}}{\sum_{i=1}^{n} e^{-\frac{(z_i-x)^2}{2}}}$$

(e)

The function is called importNormCauchy, and is found in the file normCauchyPrior.R. Testing for this function is found in a4_q3.R. Using the estimator from part (d), and running 1000 replications, we get the mean of the estimator to be 0.5013 and a smaller variance of 0.0042.

Question 4

(a)

$$
\begin{aligned}
E(\hat{E}_1) &= E\left[\frac{1}{T}\sum_{i=1}^{T} h(X_i)\right] \\
&= \frac{1}{T}\sum_{i=1}^{T} E\left[h(X)\right] \\
&= E\left[h(X)\right]
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{E}_2) &= E\left[\frac{1}{n-T}\sum_{i=1}^{n-T} \frac{(c-1)h(R_i)f(R_i)}{cg(R_i)-f(R-i)}\right] \\
&= \frac{1}{n-T}\sum_{i=1}^{n-T}\left[(c-1)E\left[\frac{h(R_i)f(R_i)}{cg(R_i)-f(R-i)}\right]\right]
\end{aligned}
$$

Now, the density of the rejected samples is $K(cg(R)-f(R))$, and so,

$$
K = \frac{1}{\int cg(r)-f(r)dr} = \frac{1}{c-1}
$$

. Therefore,

$$
\begin{aligned}
E\left[\frac{h(R_i)f(R_i)}{cg(R_i)-f(R-i)}\right] &= \int \frac{h(r)f(r)}{cg(r)-f(r)}\frac{cg(r)-f(r)}{c-1}dr \\
&= \frac{1}{c-1}\int h(r)f(r)dr \\
&= \frac{1}{c-1}E\left[h(X)\right].
\end{aligned}
$$

Therefore, $E(\hat{E}_2) = E(h(X))$. So both $\hat{E}_1$ and $\hat{E}_1$ are unbiased estimators of $E(h(X))$. Also, since the X_i's and R_i's were generated independently, then functions of these random variables will also be indepdendent. Since $\hat{E}_1$ only involves the X_i and $\hat{E}_2$ only involves the R_i, we have that $\hat{E}_1$ and $\hat{E}_2$ are independent.

(b)

$$
\begin{aligned}
var\left[\hat{E}_3(b)\right] &= var\left[b\hat{E}_1 + (1-b)\hat{E}_2\right] \\
&= b^2 var(\hat{E}_1) + (1-b)^2 var(\hat{E}_2)
\end{aligned}
$$

To find the optimal value of b, we differentiate:

$$
\frac{d}{db} var\left[\hat{E}_3(b)\right] = 2b var(\hat{E}_1) - 2(1-b) var(\hat{E}_2)
$$

Setting the derivative to zero gives us

$$
\hat{b} = \frac{var(\hat{E}_2)}{var(\hat{E}_1) + var(\hat{E}_2)}
$$

Now,

$$\begin{aligned} var(\hat{E}_1) &= var\left(\frac{1}{T}\sum_{i=1}^{T} h(X_i)\right) \\ &= \frac{1}{T^2}\sum_{i=1}^{T} var(h(X_i)) \\ &= \frac{1}{T} var(h(X)). \end{aligned}$$

$$\begin{aligned} var(\hat{E}_2) &= var\left(\frac{1}{n-T}\sum_{i=1}^{n-T} \frac{h(R_i)(c-1)f(R_i)}{cg(R_i)-f(R_i)}\right) \\ &= \frac{1}{(n-T)^2}\sum_{i=1}^{n-T}(c-1)^2 var\left(\frac{h(R_i)f(R_i)}{cg(R_i)-f(R_i)}\right) \\ &= \frac{1}{(n-T)}\left\{E\left[\left(\frac{h(R)f(R)}{cg(R)-f(R)}\right)^2\right] - \left(E\left[\frac{h(R)f(R)}{cg(R)-f(R)}\right]\right)^2\right\} \end{aligned}$$

And, we can evaluate:

$$\begin{aligned} E\left[\left(\frac{h(R)f(R)}{cg(R)-f(R)}\right)^2\right] &= \int \frac{h(r)^2 f(r)^2}{(cg(r)-f(r))^2}\left(\frac{cg(r)-f(r)}{\sqrt{f(r)}(c-1)}\right)^2 dr \\ &= \frac{1}{(c-1)^2}\int h(r)^2 f(r) dr \\ &= E(h(X)^2). \end{aligned}$$

So, we have that,

$$\begin{aligned} var(\hat{E}_2) &= \frac{1}{n-T}(E(h(X)^2) - [E(h(X))]^2) \\ &= \frac{1}{n-T} var(h(X)). \end{aligned}$$

Finally, we can conclude that,

$$\begin{aligned} \hat{b} &= \frac{var(\hat{(E)}_2)}{var(\hat{(E)}_1) + var(\hat{(E)}_2)} \\ &= \frac{var(h(X))/(n-T)}{var(h(X))/T + var(h(X))/(n-T)} \\ &= \frac{T}{n}. \end{aligned}$$