STA 721 HW 1.

1. Exercise 1.11 Here $M = P_X = X(X^TX)^{-1}X^T$

(a) $E(\hat{e}) = E((I-P_X)Y) = (I-P_X)E(Y) = (I-P_X)\mu = 0.$

(b) $cov(\hat{e}) = cov((I-P_X)Y) = (I-P_X)\, cov(Y)(I-P_X)^T = (I-P_X)\sigma^2 I(I-P_X)^T = \sigma^2(I-P_X)$

(c) $cov(\hat{e}, P_X Y) = cov((I-P_X)Y, P_X Y) = (I-P_X)\, cov(Y) P_X^T = (I-P_X)\sigma^2 I P_X = 0.$

(d) $E(\hat{e}'\hat{e}) = E(Y^T(I-P_X)^T(I-P_X)Y)$

$= tr((I-P_X)^T(I-P_X)\sigma^2 I) + \mu^T(I-P_X)^T(I-P_X)\mu$

$= \sigma^2 tr((I-P_X))$

$= (n-p)\sigma^2.$

(e) $\hat{e}'\hat{e} = Y^T(I-P_X)^T(I-P_X)Y = Y^T(I-P_X)Y = Y^TY - (Y^TP_X)Y$

(f) $cov(\hat{e}, P_X Y) = cov(Y - X\hat{\beta}, X\hat{\beta}) = 0$

$\hat{e}'\hat{e} = (Y-X\hat{\beta})'(Y-X\hat{\beta}) = Y'Y - Y'X\hat{\beta} - (X\hat{\beta})'Y + (X\hat{\beta})'(X\hat{\beta})$

$= Y'Y - Y'MY - Y'M'Y + Y'M^2Y = Y'Y - (Y'M)Y.$

2. Exercise 1.5.2

(a) $y_1 = \begin{pmatrix} 1&0&0\\0&0&0\\0&0&0 \end{pmatrix} Y$, so $y_1 \sim N\left(\begin{pmatrix} 1&0&0\\0&0&0\\0&0&0 \end{pmatrix}\begin{pmatrix}5\\6\\7\end{pmatrix}, \begin{pmatrix} 1&0&0\\0&0&0\\0&0&0 \end{pmatrix}\begin{pmatrix} 2&0&1\\0&3&2\\1&2&4 \end{pmatrix}\begin{pmatrix} 1&0&0\\0&0&0\\0&0&0 \end{pmatrix}^T\right)$,

which is $y_1 \sim N(5,2)$.

Or, since marginal of MVN is still normal, we only need to calculate $E(y_1)$ and $Var(y_1)$.

$E(y_1) = 5$, $Var(y_1) = cov(y_1, y_1) = 2$, so $y_1 \sim N(5,2)$

(b) $E\begin{pmatrix}y_1\\y_2\end{pmatrix} = \begin{pmatrix}5\\6\end{pmatrix}$, $cov\left(\begin{pmatrix}y_1\\y_2\end{pmatrix}, \begin{pmatrix}y_1\\y_2\end{pmatrix}\right) = \begin{pmatrix}2&0\\0&3\end{pmatrix}$.

So $\begin{pmatrix}y_1\\y_2\end{pmatrix} \sim N\left(\begin{pmatrix}5\\6\end{pmatrix}, \begin{pmatrix}2&0\\0&3\end{pmatrix}\right)$.

(c) $\begin{pmatrix}y_3\\y_1\\y_2\end{pmatrix} \sim N\left(\begin{pmatrix}7\\5\\6\end{pmatrix}, \begin{pmatrix}4&1&2\\1&2&0\\2&0&3\end{pmatrix}\right)$

$y_3 | y_1 = u_1, y_2 = u_2 \sim N\left(7 + (1\ 2)\begin{pmatrix}2&0\\0&3\end{pmatrix}^{-1}\begin{pmatrix}u_1-5\\u_2-6\end{pmatrix}, 4 - (1\ 2)\begin{pmatrix}2&0\\0&3\end{pmatrix}^{-1}\begin{pmatrix}1\\2\end{pmatrix}\right)$

$= N\left(\frac{1}{2}u_1 + \frac{2}{3}u_2 + \frac{1}{2}, \frac{13}{6}\right)$

(d) $\begin{pmatrix} y_3 \\ y_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 7 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}\right)$

Hence $y_3 | y_1 = u_1 \sim N(7 + 1 \cdot \frac{1}{2} \cdot (u_1 - 5),\ 4 - 1 \cdot \frac{1}{2} \cdot 1)$

$$= N\left(\frac{1}{2}u_1 + \frac{9}{2},\ \frac{7}{2}\right).$$

(e) $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix}\right)$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \Big| y_3 = u_3 \sim N\left(\begin{pmatrix} 5 \\ 6 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \frac{1}{4}(u_3 - 7), \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} \frac{1}{4} (1 \ \ 2)\right)$$

$$= N\left(\begin{pmatrix} \frac{1}{4}u_3 + \frac{13}{4} \\ \frac{1}{2}u_3 + \frac{5}{2} \end{pmatrix}, \begin{pmatrix} \frac{7}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 2 \end{pmatrix}\right)$$

(f) $\rho_{12} = 0$, $\rho_{13} = \frac{\sigma_{13}}{\sqrt{\sigma_{11}\sigma_{33}}} = \frac{\sqrt{2}}{4}$, $\rho_{23} = \frac{\sigma_{23}}{\sqrt{\sigma_{22}\sigma_{33}}} = \frac{2}{\sqrt{3 \times 4}} = \frac{\sqrt{3}}{3}$

(g) Let $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, $b = \begin{pmatrix} -15 \\ -18 \end{pmatrix}$.

Then: $AY \sim N\left(A\begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}, A\begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix}A^T\right)$

$$= N\left(\begin{pmatrix} 16 \\ 18 \end{pmatrix}, \begin{pmatrix} 11 & 11 \\ 11 & 15 \end{pmatrix}\right)$$

Hence $Z = AY + b \sim N\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 11 & 11 \\ 11 & 15 \end{pmatrix}\right)$.

(h) Let $J = \begin{pmatrix} Y \\ Z \end{pmatrix}$. Then $J = \begin{pmatrix} I_3 \\ A \end{pmatrix} Y + \begin{pmatrix} 0 \\ b \end{pmatrix} \sim N\left(\begin{pmatrix} I_3 \\ A \end{pmatrix} Y + \mu, \begin{pmatrix} I_3 \\ A \end{pmatrix} V \begin{pmatrix} I_3 \\ A \end{pmatrix}^T\right)$.

$$= N\left(\begin{pmatrix} 5 \\ 6 \\ 7 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 1 & 4 & 3 \\ 0 & 3 & 2 & 3 & 5 \\ 1 & 2 & 4 & 4 & 7 \\ 4 & 3 & 4 & 11 & 11 \\ 3 & 5 & 7 & 11 & 15 \end{pmatrix}\right).$$

$$=: N(\mu_0, \Sigma_0)$$

Then for $\forall t \in \mathbb{R}^5$, $t^T J \sim N(t^T \mu_0, t^T \Sigma_0 t)$.

Then the characteristic function f_J for J is

$$f_J(t) = E(e^{it^T J}) = f_{t^T J}(1) = e^{it^T \mu_0 - \frac{1}{2} t^T \Sigma_0 t}$$

3. (a) $P_X = X(X^TX)^{-1}X^T$

$$= U\Lambda V^T(V\Lambda U^T U\Lambda V^T)^{-1} V\Lambda U^T$$
$$= U\Lambda V^T(V\Lambda^2 V^T)^{-1} V\Lambda U^T$$
$$= U\Lambda(V^T V)\Lambda^{-2}(V^T V)\Lambda U^T$$
$$= U\Lambda\Lambda^{-2}\Lambda U^T$$
$$= UU^T$$

$\hat{Y} = P_X Y = UU^T Y.$

$\hat{\beta} = (X^TX)^{-1}X^T = (V\Lambda U^T U\Lambda V^T)^{-1} V\Lambda U^T = (V\Lambda^2V^T)^{-1}V\Lambda U^T = V^{-T}\Lambda^{-2}V^{-1}V\Lambda U = V\Lambda^{-1}U$

(b) $P_X = X(X^TX)^{-1}X^T$

$$= QR(R^TQ^TQR)^{-1}R^TQ^T$$
$$= QR(R^T I_p R)^{-1}R^TQ^T$$
$$= QRR^{-1}R^{-T}R^TQ^T$$
$$= QQ^T$$

Normal equations: $(X^TX)\beta = X^TY$

$$\Leftrightarrow (R^TQ^TQR)\beta = R^TQ^TY$$
$$\Leftrightarrow R^TR\beta = R^TQ^TY$$

Since R is full-rank, we can eliminate R^T from both sides.

$$\Leftrightarrow R\beta = Q^TY \equiv Z.$$

Let $R = \begin{pmatrix} r_{11} & \cdots & r_{1p} \\ & r_{22} \cdots & r_{2p} \\ & & \ddots \\ & & & r_{pp} \end{pmatrix}$, then the above equation is equivalent to

$$\begin{cases} r_{11}\beta_1 + \cdots + r_{1p}\beta_p = Z_1 & (1) \\ r_{21}\beta_2 + \cdots + r_{2p}\beta_p = Z_2 & (2) \\ \vdots \\ r_{pp}\beta_p = Z_p. & (p) \end{cases}$$

And it can be solved by first solve (p) and get $\hat{\beta}_p$. then plug $\beta_p = \hat{\beta}_p$ into (p-1), and obtain $\hat{\beta}_{p-1}$. And so on.

1(c) (Solve $Lw = Z_1$)

$w_1 = Z_1 / L_{11}$.

for $i = 2, \cdots, p$:

$$w_i = Z_i - \sum_{j=1}^{i-1} L_{ij} w_j .$$

end

(Solve $L^T \beta = w$)

$\beta_p = w_p / L_{pp}$

for $i = 1, \cdots, p-1$:

$$\beta_{p-i} = w_{p-i} - \sum_{j=p-i+1}^{p} L_{j,p-i} \beta_j .$$

end

In the psuedo-code above, no matrix inversion is used.

1(d) Let $P_1 = UU^T$, $P_2 = QQ^T$.

Since columns of U and columns of Q are two ONBs of $C(X)$, for $\forall v \in C(X)$, let $v = Xk$, then

$P_1 v = UU^T Xk = UU^T U \Lambda V^T k = U \Lambda V^T k = Xk = v$,

$P_2 v = QQ^T Xk = QQ^T QRk = QRk = Xk = v$.

For $\forall \alpha \perp C(X)$, we have $v^T \alpha = 0$, $\forall v \in C(X)$, then

$P_1 \alpha = UU^T \alpha = 0$. $P_2 \alpha = QQ^T \alpha = 0$.

From the above, it follows that P_1 and P_2 are perpendicular projections onto $C(X)$. Proposition B.34. of Christensen states such perpendicular projection is unique. Hence $P_1 = P_2$.

(For 3e and 3f, see the attached file written with knitr)

4. (a) $Y \sim N(\mu, LL^T) = N(\mu, \Sigma)$

(b) It doesn't matter. Since the distribution of Y is $N(\mu, LL^T)$, the only way that A would affect Y is that it determines the covariance function of Y. Hence whatever A is, as long as $\Sigma = AA^T$ remains the same, the distribution of Y remains the same.

(For 4c~4f, see the attached file written with knitr)

STA721 HW1

Zhuoqun Wang

September 9, 2019

3(e) Answer

```
#input data
X<-matrix(c(1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,1,0,0,0,0,0,0,1,1),
          nrow = 6,byrow = FALSE)
Q<-qr.Q(qr(X))[,1:3]
U<-svd(X)$u[,1:3]
round(Q,7)

##            [,1]       [,2]       [,3]
## [1,] -0.4082483 -0.4082483  0.0000000
## [2,] -0.4082483 -0.4082483  0.0000000
## [3,] -0.4082483 -0.4082483  0.0000000
## [4,] -0.4082483  0.4082483  0.8164966
## [5,] -0.4082483  0.4082483 -0.4082483
## [6,] -0.4082483  0.4082483 -0.4082483

round(U,7)

##            [,1]       [,2]       [,3]
## [1,] -0.4490638 -0.3487210  0.1003428
## [2,] -0.4490638 -0.3487210  0.1003428
## [3,] -0.4490638 -0.3487210  0.1003428
## [4,] -0.3280033  0.1542044 -0.9320058
## [5,] -0.3791035  0.5529023  0.2248990
## [6,] -0.3791035  0.5529023  0.2248990
```

We see that $Q \neq U$.

3(f) Answer

Calculate the projection matrix use U and Q respectively:

```
P_svd<-U[,1:3]%*%t(U[,1:3])
P_qr<-Q[,1:3]%*%t(Q[,1:3])
M<-matrix(c(1/3,1/3,1/3,0,0,0,1/3,1/3,1/3,0,0,0,
            1/3,1/3,1/3,0,0,0,0,0,0,1,0,0,0,0,0,
            0,1/2,1/2,0,0,0,0,1/2,1/2),nrow=6)
round(P_svd,7)
```

```
##           [,1]      [,2]      [,3] [,4] [,5] [,6]
## [1,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [2,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [3,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [4,] 0.0000000 0.0000000 0.0000000    1  0.0  0.0
## [5,] 0.0000000 0.0000000 0.0000000    0  0.5  0.5
## [6,] 0.0000000 0.0000000 0.0000000    0  0.5  0.5
```

```
round(P_qr,7)
```

```
##           [,1]      [,2]      [,3] [,4] [,5] [,6]
## [1,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [2,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [3,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [4,] 0.0000000 0.0000000 0.0000000    1  0.0  0.0
## [5,] 0.0000000 0.0000000 0.0000000    0  0.5  0.5
## [6,] 0.0000000 0.0000000 0.0000000    0  0.5  0.5
```

```
M
```

```
##           [,1]      [,2]      [,3] [,4] [,5] [,6]
## [1,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [2,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [3,] 0.3333333 0.3333333 0.3333333    0  0.0  0.0
## [4,] 0.0000000 0.0000000 0.0000000    1  0.0  0.0
## [5,] 0.0000000 0.0000000 0.0000000    0  0.5  0.5
## [6,] 0.0000000 0.0000000 0.0000000    0  0.5  0.5
```

The projection matrix obtained from SVD and QR decomposition are the same as M in 1.5.8.

4(c) Answer

The Cholesky decomposition of Y is calculated as follows:

```
V<-matrix(c(2,0,1,0,3,2,1,2,4),nrow=3)
L<-t(chol(V))
L
```

```
##            [,1]     [,2]    [,3]
## [1,] 1.4142136 0.000000 0.00000
## [2,] 0.0000000 1.732051 0.00000
## [3,] 0.7071068 1.154701 1.47196
```

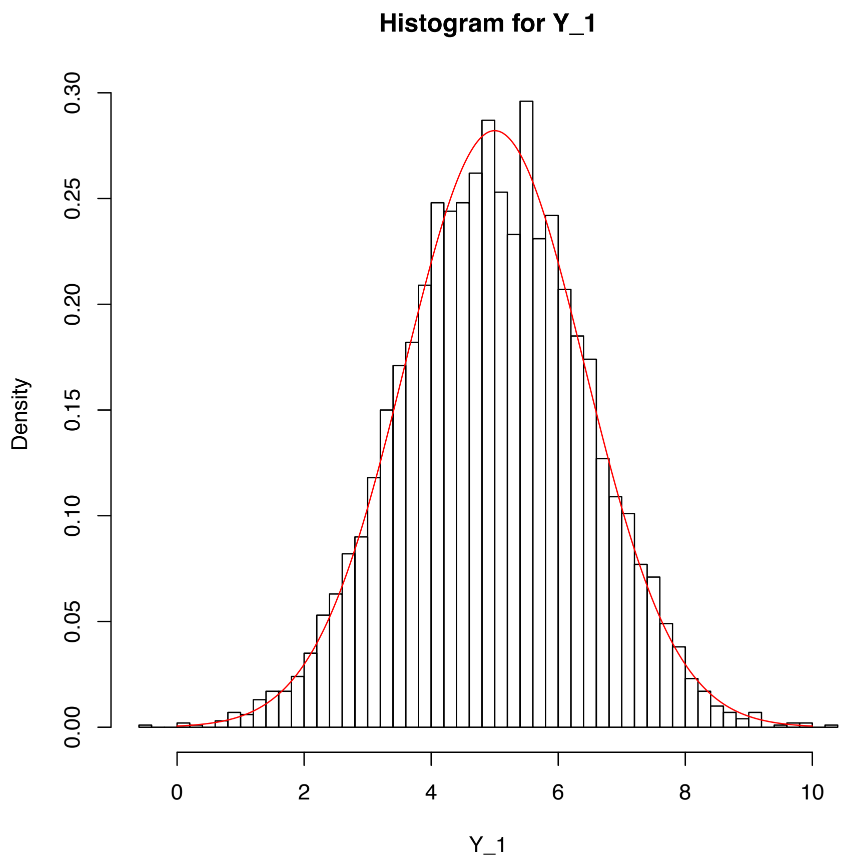
4(d) and 4(e) Answer

Generate 5000 samples of Y:

```
set.seed(123)
Z<-matrix(rnorm(3*5000),nrow=3)
mu<-c(5,6,7)
Y<-apply(Z, 2, function(x){L%*%x+mu})
```

In 1.5.2(a), we have $Y_1 \sim N(5,2)$. Create a histogram for the marginal distribution of Y_1 and overlay the actual density $N(5,2)$:

```
hist(Y[1,],freq = FALSE,breaks = 50,
     main = "Histogram for Y_1",xlab = "Y_1")
p<-dnorm(seq(0,10,length.out = 1000),mean = 5,sd = sqrt(2))
lines(seq(0,10,length.out = 1000),p,col='red')
```



The histogram of Y_1 looks like $N(5,2)$ distribution.

4(f) Answer

We calculate the sample mean, variance and covariance of Z:

```
A<-matrix(c(2,1,1,1,0,1),nrow=2)
b<-matrix(c(-15,-18),ncol=1)
Z_new<-apply(Y, 2, function(x){A%*%x+b})
Z_mean<-apply(Z_new,1,function(x){mean(x)})
Z_variance<-apply(Z_new,1,function(x){var(x)})
Z_covariance<-mean((Z_new[1,]-Z_mean[1])*(Z_new[2,]-Z_mean[2]))
Z_mean

## [1] 0.973299677 0.002561618

Z_variance
```

```
## [1] 11.30531 15.06970
Z_covariance
## [1] 11.25377
```

In 1.5.8, we have $Z \sim N(\mu_Z, \Sigma_Z)$, where $\mu_Z = (1, 0)^T$,

$$\Sigma_Z = \begin{pmatrix} 11 & 11 \\ 11 & 15 \end{pmatrix}. \tag{1}$$

Hence the estimates obtained with simulation are consistent with the results in 1.5.8.