STAT 630 Fall 2014
Homework 7 Solution

4.6.1

(a) Since X_1 and X_2 have the normal distributions and they are independent, thus U and V are also normal random variables. Since $E[U] = 3 + 40 = 43, Var(U) = Var(X_1) + 5^2 Var(X_2) = 629$ and $E[V] = -18 + C * (-8) = -8C - 18, Var(V) = 6^2 Var(X_1) + C^2 Var(X_2) = 25C^2 + 144$, thus $U \sim N(43, 629), V \sim N(-18 - 8C, 144 + 25C^2)$.

(b) Since U, V are normal random variables, thus if they are uncorrelated, then they are independent.So let $cov(U, V) = cov(X_1 - 5X_2, -6X_1 + CX_2) = -6Var(X_1) + Ccov(X_1, X_2) + 30cov(X_1, X_2) - 5Cvar(X_2) = -24 - 125C = 0$ we can obtain $C = -\frac{24}{125}$.

4.6.3

Let $X \sim N(3, 5)$ and $Y \sim N(-7, 2)$ are independent. Use the fact that $Z = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$ where $X_1, \cdots, X_n$ are i.i.d. standard normal random variables. Thus let $C_1 = 1/5, C_2 = -3$, then$C_1(X + C_2)^2 \sim \chi^2(1)$. Similarly, let $C_3 = 1/2, C_4 = 7$, then$C_3(Y + C_4)^2 \sim \chi^2(1)$. Thus, $C_1(X + C_2)^2 + C_3(Y + C_4)^2 \sim \chi^2(2)$., that is, $C_5 = 2$.

4.6.5

The moment generating function for X is $m_X(t) = E(e^{tX}) = (1 - 2t)^{-n/2}$, for $t < 1/2$. Similary, we have $m_Y(t) = (1 - 2t)^{-m/2}, t < 1/2$. According to the theorem 3.4.5, we have $m_{X+Y}(t) = m_X(t)m_Y(t) = (1 - 2t)^{-(n+m)/2}$. With theorem 3.4.6 (uniqueness theorem), we proved $X + Y \sim \chi^2(n + m)$.

4.6.6

Since we know that $X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$ and $X_n + 1^2 + X_n + 2^2 + \cdots + X_4 n^2 \sim \chi^2(3n)$, then $\frac{(X_1^2 + X_2^2 + \cdots + X_n^2)/n}{(X_n + 1^2 + X_n + 2^2 + \cdots + X_4 n^2)/3n} \sim F(n, 3n)$. Therefore, $C = \frac{1/n}{1/3n} = 3$

4.6.7

We can see $X_2^2 + \cdots + X_{n+1}^2$ has a chi-squared distribution with degree of freedom n. So according to the definition of student t distribution, $C = \sqrt{n}$.

4.6.8

Let $X \sim N(3,5)$ and $Y \sim N(-7,2)$ are independent. Use the fact that $\frac{Z}{U}$ has T distribution with 1 degrees of freedom, where Z and U are indpendent $N(0,1)$. Thus let $C_2 = -3, C_3 = 1$, then $\frac{1}{\sqrt{5}} \cdot (X + C_2)^{C_3} \sim N(0,1)$. Similarly, let $C_5 = 1, C_4 = 7$, then $\frac{1}{\sqrt{2}}(Y + C_4)^{C_5} \sim N(0,1)$. Since $\frac{\frac{1}{\sqrt{5}}(X+C_2)^{C_3}}{\frac{1}{\sqrt{2}}(Y+C_4)^{C_5}} = \frac{\sqrt{\frac{2}{5}}(X+C_2)^{C_3}}{(Y+C_4)} \sim T_1$, that is, $C_1 = \sqrt{\frac{2}{5}}, C_5 = 1, C_6 = 1$. Therefore, $C_1 = \sqrt{\frac{2}{5}}, C_2 = -3, C_3 = 1, C_4 = 7, C_5 = 1, C_6 = 1$

4.6.10

(a) $X_1^2 \sim \chi_1^2$.

(b) $X_3^2 + X_5^2 \sim \chi_2^2$.

(c) Since the numerator is the standard normal random variable, the denominator is the square root of sum of squared standard normal random variables divided by its degree of freedom, thus it has a t distribution with degree of freedom 3.

(d) $\frac{3X_{10}^2}{[X_{20}^2+X_{30}^2+X_{40}^2]} = \frac{X_{10}^2/1}{[X_{20}^2+X_{30}^2+X_{40}^2]/3}$, so its distribution is $F_{1,3}$.

(e) Similarly to (d), we know its distribution is $F_{70,30}$.

4.6.12

(a) Since $D_i \sim N(40, 5^2)$, $i = 1, 2, \cdots, 20$ and all measurements are independent, thus the sample mean $\bar{D}$ has the normal distribution with mean $E[D_i] = 40$ and variance $Var(D_i)/20 = 5/4 = 1.25$.

(b) Similarly, the sample mean of Compaq desktop computers has the normal distribution with mean $E[C_i] = 45$ and variance $Var(C_i)/30 = 32/15 = 2.13$.

(c) First the measurements of dell desktop computers should be independent with the measurements of Compaq desktop computers, thus C_i and D_i are independent. Since $\bar{C}, \bar{D}$ are sample averages, then they are independent normal variables. Therefore, $Z = \bar{C} - \bar{D}$ has the normal distribution with mean $E[\bar{C}] - E[\bar{D}] = 45 - 40 = 5$ and variance $Var(\bar{C}) + Var(\bar{D}) = 5/4 + 32/15 = \frac{203}{60} = 3.38$.

(d) $P(\bar{C} < \bar{D}) = P(Z < 0) = P\left(\frac{Z-5}{\sqrt{203/60}} < \frac{-5}{\sqrt{203/60}}\right) = \Phi\left(\frac{-5}{\sqrt{203/60}}\right) = 0.003281$.

(e) First we notice that $D_i \sim N(40, 5^2)$. Then $U = (n-1)S^2$. Thus $\frac{(n-1)S^2}{\sigma^2} = \frac{U}{5^2}$ has a chi-square distribution with degree of freedom $n - 1 = 19$. Thus $P(U > 633.25) = P(\frac{U}{5^2} > \frac{633.25}{25}) = 1 - P(\chi_{19}^2 < \frac{633.25}{25}) = 0.1499642$.

5.1.11

First we generate large samples of X from the standard normal distribution and calculate the Y to obtain the empirical distribution of Y. Then we can estimate the probability $P(Y \in A)$ for any interval A. Here is the R code:

```
x=rnorm(1000)
y=x^4+2*x^3-3
sum(y>1 & y<2)/1000
```

One run result for $P(Y \in (1,2)) = 0.0175$.

5.3.11

Since $exp(\psi) = \theta/(1-\theta)$, we can solve it for θ and obtain: $\theta = \frac{exp(\psi)}{1+exp(\psi)}$. Then the probability function for X_i is

$$\left(\frac{exp(\psi)}{1+exp(\psi)}\right)^{x_i}\left(\frac{1}{1+exp(\psi)}\right)^{(1-x_i)}$$

where $x_i \in \{0,1\}$. Since $\theta \in [0,1]$, thus the parameter space for ψ is $(-\infty,+\infty)$ with $\psi = -\infty$ when $\theta = 0$ and $\psi = +\infty$ when $\theta = 1$.

5.4.11

(a) Use the rnorm function in R to generate 1000 random numbers, $(x_1, \cdots, x_{1000})$, from $N(3,2)$. Then sort it to obatin the order statistics $(x_{(1)}, \cdots, x_{(1000)})$. Record the proportion of data values less than or equal to each value. Then $F_X(x)$ equals the largest value i/n, such that $x_{(i)} \leq x$.
R code:

```
y=rnorm(1000,mean=3,sd=sqrt(2))
plot.ecdf(y)
x=seq(-2,8,length=1001)
lines(x,pnorm(x,mean=3,sd=sqrt(2)),col=2)
```

The empirical distribution is plotted below.

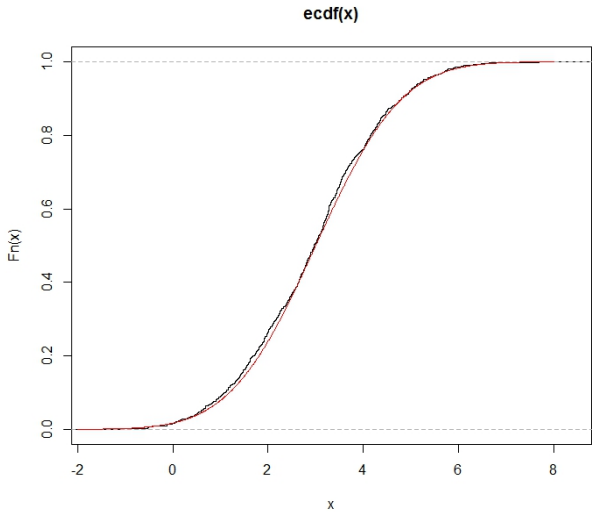

Figure 1: Empirical distribution of X

(b) R code:

```
hist(y, breaks=seq(-5, 10, by=1),freq=FALSE )
x=seq(-5,10,length=1001)
lines(x,dnorm(x,mean=3,sd=sqrt(2)),col=2)
```

(c) R code: Change "`by=1`" to "`by=0.1`".
The density histograms of (b) and (c) are shown below.

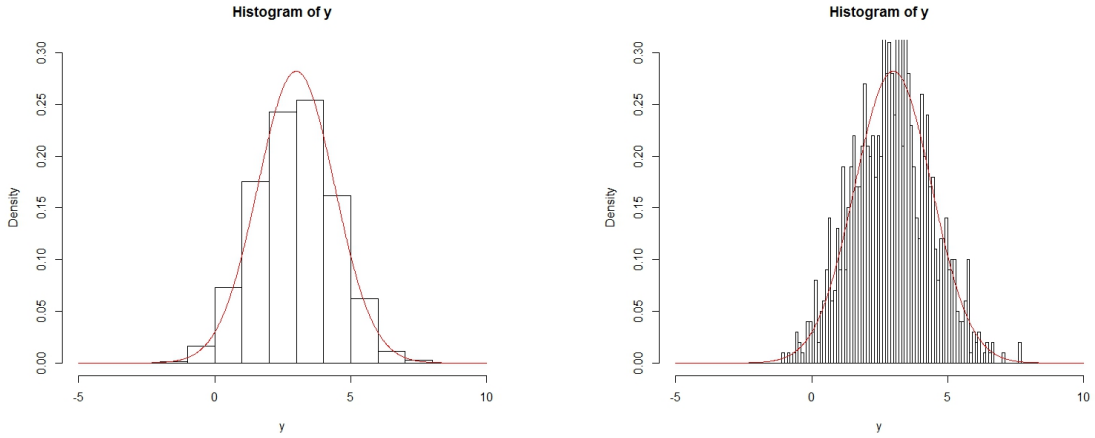

Figure 2: Density histogram with intervals of length = 1 and 0.1

(d) We can see the histogram in (c) looks much more erratic than in (b). We can attribute it to the the sampling error.

(e) If we use very short intervals, then the number of data falling into each interval is small. So the histogram will not have any kind of recognizable shape. This over-fitting problem is caused by making the intervals too small.

5.5.5

The empirical distribution function is shown below.

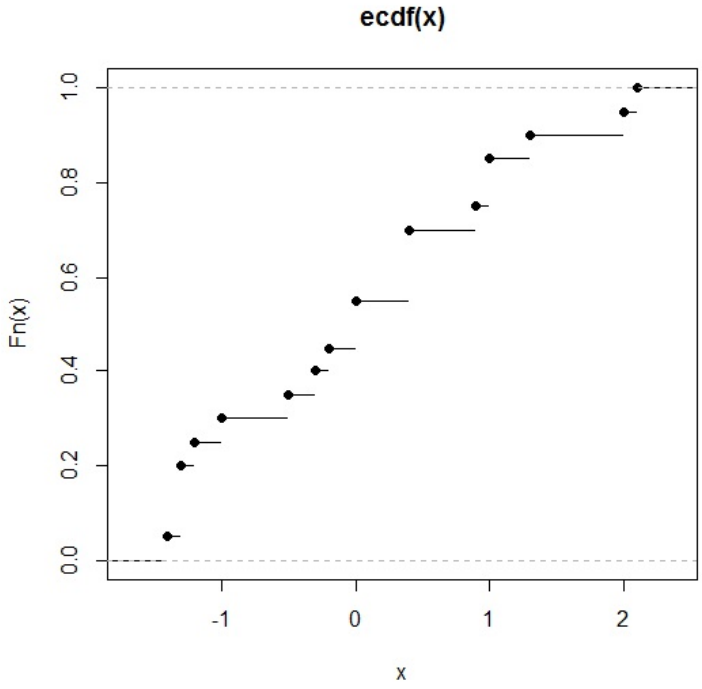

Figure 3: Empirical distribution function

R code:

```
x=c(1.0,-1.2,0.4,1.3,-0.3,-1.4,0.4,-0.5,-0.2,-1.3,
    0.0,-1.0,-1.3,2.0,1.0,0.9,0.4,2.1,0.0,-1.3)
quantile(x,type=4)
IQR(x,type=4)
plot.ecdf(x)
```

The sample median is 0, the first and third quartiles are -1.2 and 0.90 respectively and the IQR is 2.1. Since there are 17 numbers less and equal than 1, the estimate of $F(1)$ equals $\frac{17}{20}$.

5.5.19

(a) The boxplot is shown below: R code:

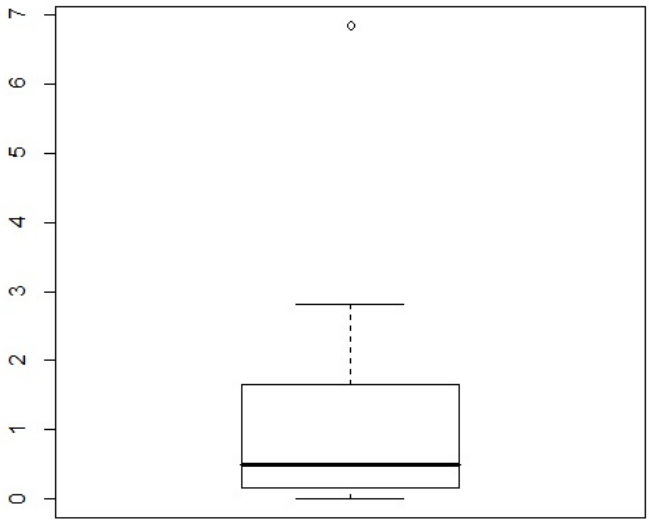

Figure 4: Boxplot of 50 Chi-square(1) samples

```
x=rchisq(50,1)
boxplot(x)
```

(b) We can see there is an outlier in the data set. So the distribution will skew to the right.

(c) The median is more appropriate to measure the location of the distribution and the IQR is more appropriate to measure the spread of the distribution. Because these measures are more robust and less unaffected by the outliers and skewness.

5.5.20

(a) My simulation result is 5.427074.

(b) Here is one simulation result: 5.144087.

(c) If the normality assumption is true, then the estimate in (b) is more appropriate because it uses the distribution assumption. R codes:

```
x=rnorm(50,4,1)
x=sort(x)
x0.9_a=quantile(x,0.9,type=4)
m=mean(x)
s=sd(x)
x0.9_b=m+s*qnorm(0.9)
```