# STAT430 Homework #3: Due Friday, February 15, 2019.

**Name: Oliver Shanklin**

---

0. We finish Chapter 7 in this homework: remainder of **Section 7.2**, **Section 7.3-7.4** on the Central Limit Theorem (CLT), and **Section 7.5** on the normal approximation to the binomial. First midterm is coming up. Be sure you are strong so far on distribution of maximum and minimum order statistic; expectation, variance and covariance calculations; normal probability computations; the "origin stories" of the $\chi^2$, $t$, and $F$ distributions; and the Central Limit Theorem and its applications.

1. (a) Use the Central Limit Theorem to establish an approximate normal distribution for a $\chi^2_n$ random variable with $n$ large, specifying the mean and variance of your approximating distribution. (b) Check your normal approximation for each of $n = 3, 9, 27, 81$ by (i) setting your random number seed to 4302019 and simulating 10,000 random $\chi^2_n$ random variables; (ii) plotting the empirical cdf of your simulated random variables; and (iii) adding a curve of the approximating normal cdf. (That is, four different empirical cdf's, each with its own approximating normal cdf.) Comment on the quality of your approximation.

---

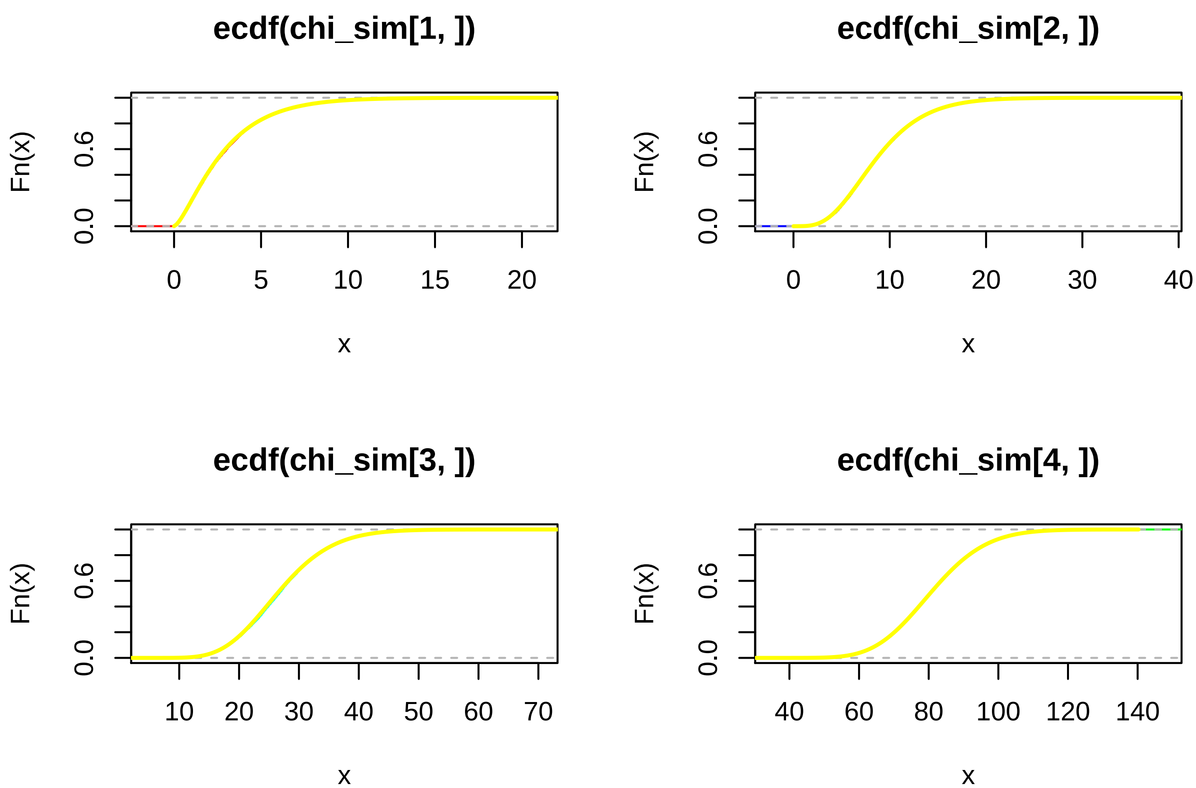
**Answer:**

```r
set.seed(4302019)
n <- c(3,9,27,81)
chi_sim <- matrix(data=0, nrow=4,ncol=10000,byrow = T)

for(i in seq(1:4)){

  chi_sim[i,] <- rchisq(10000, n[i], ncp=0)

}
x_vec <- seq(0,max(chi_sim), length = 1000)
par(mfrow = c(2,2))
plot(ecdf(chi_sim[1,]),col="red")
lines(x_vec, pchisq(x_vec, df=3), lwd="2", col="Yellow")
plot(ecdf(chi_sim[2,]),col="blue")
lines(x_vec, pchisq(x_vec, df=9), lwd="2", col="Yellow")
plot(ecdf(chi_sim[3,]),col="cyan")
lines(x_vec, pchisq(x_vec, df=27), lwd="2", col="Yellow")
plot(ecdf(chi_sim[4,]),col="green")
lines(x_vec, pchisq(x_vec, df=81), lwd="2", col="Yellow")
```




The quality of the approximation is pretty close to the actual cdf of the $\chi^2_n$. They yellow line is the actual cdf and the colors under the yellow line is the approximation for the $\chi^2_n$.
*****************************************************************************

2. Complete **Exercise 7.68**, using `R` instead of the suggested applet to compute the exact (binomial) probability and the approximating normal probability. Be sure to use the appropriate continuity correction.

---

**Answer:**

**a)**

\begin{align*}

$$P(Y \geq 29) = 1 - P(Y < 29) = 1 - P(Y \leq 28) = 1 - P(Y \leq 28.5)$$

$$= 1 - P(Z \leq \frac{28.5 - 50 * 0.52}{\sqrt{50 * .52 * .48}})$$

```r
# Approx
1-pnorm(((28.5-50*.52)/(sqrt(50*.52*.48))),0,1)
```

```
## [1] 0.2395741
```

```r
# Exact
1 - pbinom(29,50,.52)
```

```
## [1] 0.1609333
```

---

3. Let $Y_1 \sim N(1,1)$, $Y_2 \sim N(2,4)$, and $Y_3 \sim N(3,9)$ be independent random variables. Use these random variables to construct (a) a $\chi^2$ random variable with 3 degrees of freedom (df); (b) a $t$ random variable with 2 df; (c) an $F$ random variable with 1 numerator df and 2 denominator df; (d) an $F$ random variable with 2 numerator and 1 denominator df.

---

**Answer:**

We need to make $Z_i$ for each $Y_i$:

$$Z_1 = \frac{Y_1 - 1}{1}$$
$$Z_2 = \frac{Y_2 - 2}{2}$$
$$Z_3 = \frac{Y_3 - 3}{3}$$

Where $Z_i \sim N(0,1)$.

**a)**

$$W = (Z_1^2 + Z_2^2 + Z_3^2) \sim \chi_3^2$$

**b)**

$$W = Z_1^2 + Z_2^2 \sim \chi_2^2$$
$$T = \frac{Z_3}{\sqrt{W/2}} \sim t_2$$

Since $Z_3$ is independent of $Z_1, Z_2$.

**c)**

$W_1 = Z_1^2 \sim \chi_1^2$, $W_2 = Z_2^2 + Z_3^2 \sim \chi_2^2$

$$F_{1,2} = \frac{W_1/1}{W_2/2} \sim \mathcal{F}_{\infty,\in}$$

**d)**

$W_1 = Z_1^2 + Z_2^2 \sim \chi_2^2$, $W_2 = Z_3^2 \sim \chi_1^2$

$$F_{2,1} = \frac{W_2/2}{W_1/1} \sim \mathcal{F}_{\in,\infty}$$

---

4. Refer to **Example 7.7**, page 363-364. If we take independent samples of sizes $n_1 = 6$ and $n_2 = 10$ from two normal populations with equal population variances, use `R` to find:

- (a) $P(S_1^2/S_2^2 > 2)$;

- (b) $P(S_1^2/S_2^2 < 0.5)$; and
- (c) the probability that one of the sample variances is at least twice as big as the other. Then check your answers via simulation as follows:
- (d) set your random number seed to 4302019, simulate 10,000 $F$ random variables using `rf`, and check (a);
- (e) using the simulated $F$ random variables from (d), check your computation in (b);
- (f) using the simulated $F$ random variables from (d), check your computation in (c).

---

**Answer:**

**a)**

$$P(\frac{(5S_1^2/\sigma^2)}{(9S_2^2/\sigma^2)} \geq 2) = P(F_{5,9} \geq 2)$$

```
1-pf(2, df1=5, df2=9)
```

```
## [1] 0.1727099
```

**b)**

$$P(F_{5,9} < 0.5)$$

```
pf(0.5, df1=5, df2=9)
```

```
## [1] 0.2304087
```

**c)**

$$P(S_1^2 \geq 2S_2^2) + P(S_2^2 \geq 2S_1^2)$$
$$P(F_{5,9} \geq 2) + P(F_{9,5} \geq 2)$$

```
(1-pf(2,df1=5,df2=9))+(1-pf(2,df1=9,df2=5))
```

```
## [1] 0.4031185
```

**d) e) f)**

```
set.seed(4302019)
f_sim <- rf(10000,5,9,ncp=0)


mean((f_sim >= 2))
```

```
## [1] 0.1725
```

```
mean((f_sim < 0.5))
```

```
## [1] 0.2296
```

These results are very similar to the exact probabilities.

---

5. Recently, researchers analyzed the 700 most popular songs since 2012 (as ranked on Billboard's Hot 100), and found that a small group of just 10 songwriters was responsible for 23% of these songs. Suppose you sample $n = 35$ songs randomly, with replacement, from the 700 most popular songs. What is the exact probability that your 35 songs will include at least 11 songs from the small group of songwriters? What is the approximate probability, using the CLT (with continuity correction)?

---

**Answer:**

Let $X \sim Bin(35, .23)$

We need $P(X \geq 11)$, or if we standardize it to satisfy CLT,

$$P(Z \geq \frac{10.5 - 35 * 0.23}{\sqrt{35 * 0.23 * 0.77}}) = P(Z \geq 0.98406) = 1 - P(Z < 0.98406)$$

```
1-pnorm(0.98406,0,1)
```

```
## [1] 0.162543
```

---

6. Complete **Exercise 7.58** of the text.

---

**Answer:**

$W_i = X_i - Y_i, i = 1, 2, \ldots, n)$ where $W_i$ are independent because $X_i$ and $Y_i$ are independent.

$E(W_i) = E(X_i) - E(Y_i) = \mu_1 - \mu_2 = \mu_W$

$V(W_i) = \sigma_1^2 + \sigma_2^2 = \sigma_W^2 < \infty$

$$U_n = \frac{\bar{W} - \mu_W}{\sqrt{\sigma_W^2 / n}}$$

Since $W_i$ are iid and variance is finite, the $W_i$ follow CLT.

---

7. Complete **Exercise 7.62** of the text.

---

**Answer:**

$$P(X \geq 4) = 1 - P(X < 4)$$

$X_i$ = service time for $i^{th}$ customer.

$$P(\sum_{i=1}^{100} X_i \geq 240) = 1 - P(\sum_{i=1}^{100} X_i < 240)$$
$$= 1 - P(\bar{X} \leq 2.4)$$

Where, $\bar{X} \sim N(\mu_{\bar{X}} = 2.5, \sigma_{\bar{X}}^2 = 4/100)$

So, after standardizing,

$$1 - P(Z \leq \frac{2.4 - 2.5}{2/\sqrt{100}} = 1 - P(Z \leq -0.5) = 0.6915$$

---

8. Complete **Exercise 7.63** of the text, finding the number of customers $n$ such that the probability that the orders of all $n$ customers can be processed in less than 2 hours is approximately 0.1 (as close as possible, without going over 0.1).

---

**Answer:**

$$P(X \leq 2(hours)) = 0.1$$

$$P(\sum_{i=1}^{n} X_i \leq 120) = 0.1$$

$$P(\bar{X} \leq 120/n) = 0.1$$

Now Standardize,

$$P(Z \leq \frac{120/n - 2.5}{2/\sqrt{n}}) = 0.1$$

After standardizing, we get that,

$$P(Z \leq -1.28) = 0.1$$

So we need to set,

$$\frac{120/n - 2.5}{2/\sqrt{2}} = -1.28$$

After using Wolfram Alpha to solve for $n$, we get $n \approx 55.638$ and we can round up to $n = 56$.

---