DSCC/CSC/TCS 462 Assignment 2

Due Thursday, October 6, 2022 by 4:00 p.m.

Qirong Huang

This assignment will cover material from Lectures 6, 7, and 8.

1. Consider random variables X and Y. Calculate $\text{Var}(3X + 2Y)$ given the following information. (Hint: At some point, you may need to use the fact that variance cannot be negative.)
   - $\text{E}(3X + 2) = 8$
   - $\text{E}(4X + 2Y) = 14$
   - $\text{E}(2Y(X + 1)) = 28$
   - $\text{E}(X^2Y^2) = 144$
   - $\text{Cov}(X^2, Y^2) = 36$
   - $\text{E}(X^2 + 2Y^2) = 33$

$$\begin{aligned} E(3X + 2) &= 3E(X) + 2 \\ &= 8 \\ 3E(X) &= 6 \end{aligned}$$

   - $\text{E}(X) = 2$

$$\text{E}(4X + 2Y) = E(4X) + E(2Y) = 4E(X) + 2E(Y) = 14$$

   $4 * 2 + 2E(Y) = 14$
   - $\text{E}(Y) = 3$

$$\text{E}(3X + 2Y) = E(X) + 2E(Y) = 3 * 2 + 2 * 3 = 12$$

$$\text{E}(2Y(X + 1)) = 2E(XY) + 2E(Y) = 28$$

   $\text{E}(XY) + E(Y) = 14$ $\text{E}(XY) + 3 = 14$
   - $\text{E}(XY) = 11$

$$\text{E}(X^2 + 2Y^2) = E(X^2) + E(2Y^2) = 33$$

   $\text{E}(X^2) = 33 - E(2Y^2)$

$$\text{Cov}(X^2, Y^2) = 36 = E(X^2Y^2) - E(X^2)E(Y^2)$$

   $36 = 144 - E(X^2)E(Y^2)$

$$\text{E}(X^2)E(Y^2) = (33 - E(2Y^2))E(Y^2) = 108$$

$$\mathrm{E}(X^2) = 24, E(Y^2) = 9/2, or E(X^2) = 9, E(Y^2) = 12$$
$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - 2^2 = E(X^2) - 4$$
$$\mathrm{Var}(Y) = E(Y^2) - E(Y)^2 = E(Y^2) - 3^2 = E(Y^2) - 9$$

- $\mathrm{Var}(Y) \geq 0$
- $\mathrm{E}(Y^2) \geq 9$

$$\mathrm{E}(X^2) = 9, E(Y^2) = 12$$
$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = 11 - 2 * 3 = 5$$
$$\begin{aligned}\mathrm{Var}(3X + 2Y) &= 9Var(X) + 4Var(Y) + 12Cov(X, Y)\\ &= 9(E(X^2) - E(X)^2) + 4(E(Y^2) - E(Y)^2) + 12(E(XY) - E(X)E(Y))\\ &= 9 * (9 - 2^2) + 4(12 - 3^2) + 12 * (11 - 2 * 3)\\ &= 45 + 12 + 60\\ &= 117\end{aligned}$$

Comment: following equation was applied in the calculation of Var(3X+2Y).

$$\mathrm{E}(aX + b) = aE(X) + b$$
$$\mathrm{Var}(X) = E(X^2) - E(X)^2$$
$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$$
$$\mathrm{Var}(aX + bY) = a^2Var(x) + b^2Var(Y) + 2abCov(X, Y)$$

2. The density function of X is given by $f_X(x) = ax^3 + bx + \frac{2}{3}$ for $x \in [0, 1]$, and $\mathrm{E}(X) = \frac{7}{15}$.

   a. Find a and b.

$$\begin{aligned}\mathrm{E}(X) &= \int_0^1 xf(x)\,dx\\ &= \int_0^1 x(ax^3 + bx + \frac{2}{3})\,dx\\ &= \frac{a}{5} + \frac{b}{3} + \frac{1}{3} = \frac{7}{15}\end{aligned}$$

- $\frac{a}{5} + \frac{b}{3} = \frac{2}{15}$

$$\begin{aligned}\int_0^1 f(x)\,dx &= \int_0^1 (ax^3 + bx + \frac{2}{3})\,dx\\ &= \frac{a}{4} + \frac{b}{2} + \frac{2}{3} = 1\end{aligned}$$

- $b = 2$
- $a = -\frac{8}{3}$

b. Calculate the CDF, $F(X)$.

$$\begin{aligned}F(X) &= Pr(X \leq x)\\ &= \int_0^x f(x), dx\\ &= \int_0^x (-\frac{8}{3}x^3 + 2x + \frac{2}{3}), dx \quad = \frac{2x^4}{3} + x^2 + \frac{2x}{3}\end{aligned}$$

c. Calculate $\Pr(X > 0.75)$

$$\Pr(X > 0.75) = 1 - Pr(X \leqslant 0.75)$$

$$\Pr(X \leqslant 0.75) = \int_0^{0.75} f(x)\,dx = \int_0^{0.75} (-\frac{8}{3}x^3 + 2x + \frac{2}{3})\,dx = 0.8516$$

$$\begin{aligned}\Pr(X > 0.75) &= 1 - Pr(X \leqslant 0.75)\\ &= 1 - 0.8516\\ &= 0.1484\end{aligned}$$

e. Suppose $Y = 1.5X + 2$. Calculate $\mathrm{E}(Y)$.

$$\begin{aligned}E(Y) &= E(1.5X + 2)\\ &= 1.5E(X) + 2\\ &= 1.5 * \frac{7}{15} + 2\\ &= 2.7\end{aligned}$$

3. The distribution of battery life of MacBook laptops is normally distributed with a mean of 8.1 hours and a standard deviation of 1.3 hours. The distribution of Dell laptops is normally distributed with a mean of 6.8 hours with a standard deviation of 0.9 hours.

   a. Calculate the probability that a randomly selected MacBook laptop battery lasts more than 9 hours.

```
1-pnorm(9,8.1,1.3)
```

```
## [1] 0.2443721
```

$\mu_{Mac} = 8.1$, $\sigma_{Mac} = 1.3$,

$$Pr(X \geq x) = 1 - pnorm(x, mean, sd)$$

```
b. Calculate the probability that a randomly selected Dell laptop battery lasts between
```

```
pnorm(8,6.8,0.9)-pnorm(6,6.8,0.9)
```

```
## [1] 0.7217574
```

$\mu_{Del} = 6.8$, $\sigma_{Del} = 0.9$,

$$Pr(6 \leq X \leq 8) = pnorm(8, 6.8, 0.9) - pnorm(6, 6.8, 0.9) = 0.7217574$$

```
c. How long must a MacBook laptop battery last to be in the top 3%?
```

```
qnorm(0.97,8.1,1.3)
```

```
## [1] 10.54503
```

```
d. How long must a Dell laptop battery last to be at the 30th percentile?
```

```r
qnorm(0.3,6.8,0.9)
```

```
## [1] 6.32804
```

```
e. Calculate the probability that a randomly selected MacBook laptop lasts longer than t
```

```r
#25th percentile of Dell laptops battery hours
qnorm(0.25,6.8,0.9)
```

```
## [1] 6.192959
```

```r
#the probability that a randomly selected MacBook laptop lasts longer than the 25th pe
1-pnorm(6.192959,8.1,1.3)
```

```
## [1] 0.9288058
```

```
f. A randomly selected laptop has a battery life of at least 8.5 hours. Calculate the pr
```

$x \geq 8.5$

```r
#the probability of this laptop being a MacBook
1-pnorm(8.5,8.1,1.3)
```

```
## [1] 0.3791582
```

```r
#the probability of this laptop being a Dell
1-pnorm(8.5,6.8,0.9)
```

```
## [1] 0.02945336
```

The probability of this laptop being a Macbook is 37.92%. The probability of this laptop being a Dell is 2.95%. 4. Payton applies for 12 jobs, each of which he has a 70% chance of getting a job offer for. Assume that job offers are independent of each other.

```
a. How many job offers is Payton expected to receive?
```

```r
#n = 12, p = 0.7, E(X) = np
12*0.7
```

```
## [1] 8.4
```

So payton is expected to receive 8 job offers. b. Calculate the probability that Payton receives job offers from all 12 places.

```r
dbinom(12,12,0.7)
```

```
## [1] 0.01384129
```

The probability that Payton receives job offers from all 12 places is about 1.38%. c. Calculate the probability that Payton receives between 5 and 7 (inclusive, i.e., 5, 6, or 7) job offers.

```
a<-pbinom(7,12,0.7)
b<-pbinom(5-1,12,0.7)
a-b
```

```
## [1] 0.2668552
```

The probability that Payton receives between 5 and 7 (inclusive, i.e., 5, 6, or 7) job offers is about 26.69% d. Calculate the probability that Payton receives strictly more than 9 job offers.

```
1-pbinom(9,12,0.7)
```

```
## [1] 0.2528153
```

The probability that Payton receives strictly more than 9 job offers is about 25.28%.

```
e. Calculate the probability that Payton receives strictly fewer than 3 job offers.
```

```
pbinom(2,12,0.7)
```

```
## [1] 0.0002063763
```

The probability that Payton receives strictly fewer than 3 job offers is 0.02% f. Calculate the variance of the number of job offer Payton is expected to receive. $var(X) = np(1-p)$

```
12*0.7*(1-0.7)
```

```
## [1] 2.52
```

5. Suppose a company has three email accounts, where the number of emails received at each account follows a Poisson distribution. Account A is expected to receive 4.2 emails per hour, account B is expected to receive 5.9 emails per hour, and account C is expected to received 2.4 emails per hour. Assume the three accounts are independent of each other.

   a. Calculate the variance of emails received for each of the three accounts.

$$
\begin{aligned}
\sigma^2 &= \lambda = \mu \\
\sigma_A^2 &= 4.2 \\
\sigma_B^2 &= 5.9 \\
\sigma_C^2 &= 2.4
\end{aligned}
$$

   b. Calculate the probability that account A receives at least 8 emails in an hour.

```
1-ppois(8-1,4.2)
```

```
## [1] 0.06394334
```

```
c. Calculate the probability that account B receives exactly 4 emails in an hour.
```

```
dpois(4,5.9)
```

```
## [1] 0.1383118
```

d. Calculate the probability that account C receives at most 3 emails in an hour.

```r
ppois(3,2.4)
```

```
## [1] 0.7787229
```

e. Calculate the probability that account B receives between 2 and 4 emails in an hour.

```r
ppois(4,5.9)-ppois(1,5.9)
```

```
## [1] 0.2797626
```

f. Calculate the probability that the company receives more than 10 emails total in an h

```r
#\lambda = E(A+B+C) = E(A)+E(B)+E(C)
sum(4.2,5.9,2.4)
```

```
## [1] 12.5
```

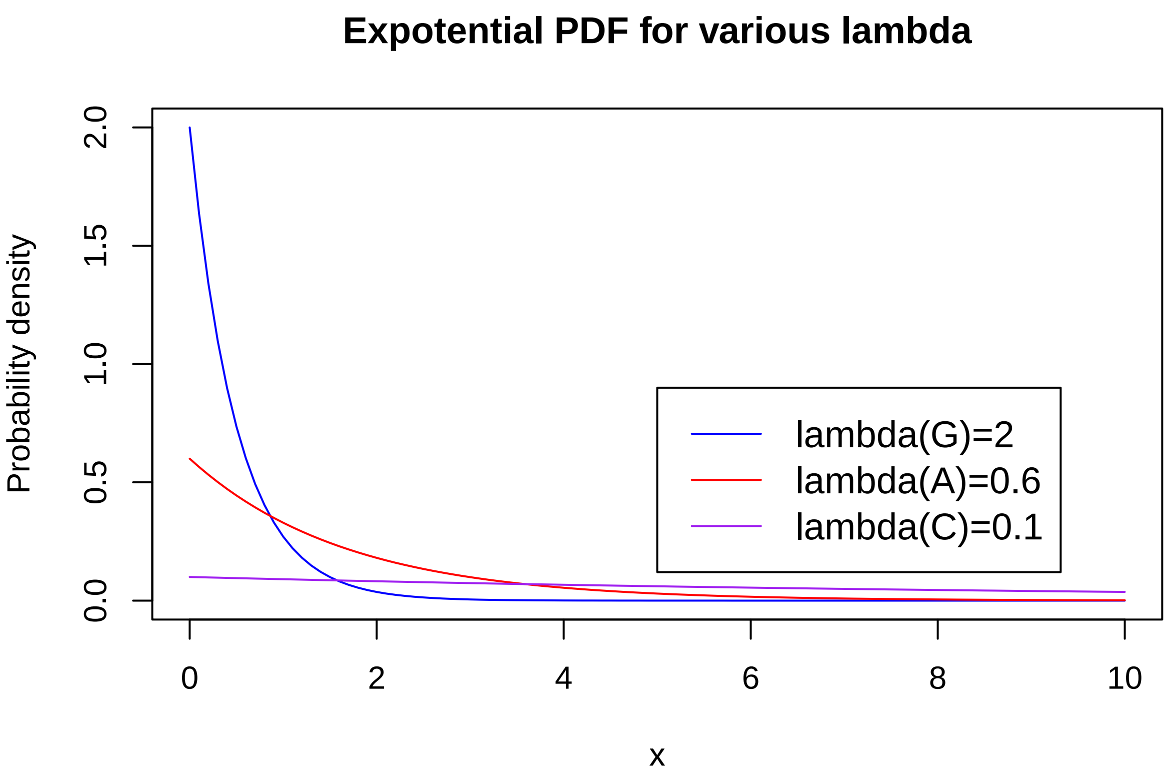
```r
#the probability that the company receives more than 10 emails total in an hour
1-ppois(10,12.5)
```

```
## [1] 0.7029253
```

6. Suppose that we are interested in the length of time before the next lightning strike. There are three types of lightning we are interested in: cloud-to-ground (G), cloud-to-air (A), and cloud-to-cloud (C). For all types of lightning, the length of time before the next strike is distributed according to an exponential distribution, but the exponential distribution has a different parameter for each type of lightning. In particular, $\lambda_G = 2$, $\lambda_A = 0.6$, and $\lambda_C = 0.1$.

    a. On a single plot, visualize the PDFs over the range $x \in [0, 10]$ for each of these exponential distributions. It may be helpful to use the function "dexp" in R.

```r
#plot PDF curves
curve(dexp(x, rate = 2), from=0, to=10, col='blue', ylab="")
curve(dexp(x, rate = 0.6), from=0, to=10, col='red', add=TRUE)
curve(dexp(x, rate = 0.1), from=0, to=10, col='purple', add=TRUE)
#add legend
legend(5, .9, legend=c("lambda(G)=2", "lambda(A)=0.6", "lambda(C)=0.1"),
       col=c("blue", "red", "purple"), lty=1, cex=1.2)
title(main="Expotential PDF for various lambda",
   xlab="x", ylab="Probability density")
```



`b. What are $\E(G)$, $\E(A)$, and $\E(C)$, as well as $\Var(G)$, $\Var(A)$, and $\Var(C)`

- $\mu = \frac{1}{\lambda}$ *

$$\mathrm{E}(G) = \frac{1}{2} = 0.5$$

-

$$\mathrm{E}(A) = \frac{1}{0.6} = \frac{5}{3} = 1.667$$

-

$$\mathrm{E}(C) = \frac{1}{0.1} = 10$$

$$\mathrm{Var} = \frac{1}{\lambda^2}$$

-

$$\mathrm{Var}(G) = 0.25$$

$$\mathrm{Var}(A) = \frac{1}{0.6^2}$$

-

$$\mathrm{Var}(A) = \frac{25}{9} = 2.778$$

$$\mathrm{Var}(C) = \frac{1}{0.1^2}$$

-

$$\mathrm{Var}(C) = 100$$

c. Suppose that we repeatedly sample collections of $n = 100$ observations from the distribution of cloud-to-ground (G) lightning strike timings. What is the mean and variance of this sample distribution?

```
# n>30, CLT
#The mean of sample distribution equals to the population mean: \mu=\frac{1}{\lambda}
1/2
```

```
## [1] 0.5
```

```
#The variance of the distribution of sample distribution equals to \sigma^2/n
0.25/100
```

```
## [1] 0.0025
```

```
d. Now, let us examine the empirical sampling distribution of cloud-to-ground ($G$) ligh
```

```
x_10<-c()
for (m in 1:5000) {
  x_10<-append(x_10,mean(rexp(10,2)))
}
mean(x_10)
```

```
## [1] 0.4990149
```

```
x_100<-c()
for (m in 1:5000) {
  x_100<-append(x_100,mean(rexp(100,2)))
}
mean(x_100)
```

```
## [1] 0.5006304
```

```
x_1000<-c()
for (m in 1:5000) {
  x_1000<-append(x_1000,mean(rexp(1000,2)))
}
mean(x_1000)
```

```
## [1] 0.5001429
```

```
sd(x_10)
```

```
## [1] 0.1586488
```

```
sd(x_100)
```

```
## [1] 0.05040668
```

```
sd(x_1000)
```

```
## [1] 0.01570363
```

```r
#standard deviation of sample mean distribution equals to sigma/sqrt(n)
0.5/sqrt(10)
```

```
## [1] 0.1581139
```

```r
0.5/sqrt(100)
```

```
## [1] 0.05
```

```r
0.5/sqrt(1000)
```

```
## [1] 0.01581139
```

Comments:The results agree with our central limit theorem. Since the population we sampling from is exponential distribution, the population mean equals to $1/\lambda$ (0.5), which is about equals to sample mean in our results. According to the central limit theorem, when sampling time is sufficient large, even through the population we are sampling from is not normal distribution(exponential distribution), the shape of the distribution of sample mean will be normal.Hence,the mean of the sample distribution equals to the population mean. The standard error of the exponential distribution population mean equals to $1/\lambda$=0.5. So according to CLT, when n is sufficiently large, the shape of the sampling distribution is approximately normal.And the standard deviation of the distribution of sample means is equal to standard error of the mean ($\sigma/\sqrt{n}$) which are the same as we got in our results. With n increased, the standard deviation of sample distribution mean decreased.sd(x_10)>sd(x_100)>sd(x_1000).

7. Suppose that we would like to get an idea of how much coffee is consumed by the entire University of Rochester each day. We take a sample of 100 days and find that the average amount of coffee consumed by the University of Rochester per day is 580 gallons.

   a. Assume that coffee consumption comes from a normal distribution with $\sigma = 90$. Find a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day. $\sigma = 90$, $\bar{x} = 580$, $n = 100$,

$$\alpha = 1 - 0.95 = 0.05$$

,

$$((\bar{x} - z_{\frac{\alpha}{2}} * \sigma/\sqrt{n}), (\bar{x} + z_{\frac{\alpha}{2}} * \sigma/\sqrt{n}))$$

```r
580-1.96*90/sqrt(100)
```

```
## [1] 562.36
```

```r
580+1.96*90/sqrt(100)
```

```
## [1] 597.64
```

So the two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day is (562.36, 597.64). b. Assuming the same information as part a, suppose that we now only want a upper-bound confidence interval. Calculate a one-sided 95% upper-bound confidence interval for the average amount of coffee consumed by the University of Rochester each day.

$$\alpha = 1 - 0.95 = 0.05$$

$$\bar{x} + z_\alpha * \sigma/\sqrt{n}$$

```r
qnorm(0.95)
```

```
## [1] 1.644854
```

```r
580+1.645*90/sqrt(100)
```

```
## [1] 594.805
```

The one one-sided 95% upper-bound confidence interval for the average amount of coffee consumed by the University of Rochester each day is 594.805.
c. Now, suppose that we do not know the variance of the true distribution of coffee consumption. However, in our sample, we see that $s = 80$. Find a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day.

$$t = (\bar{x} - \mu)/(s/\sqrt{n})$$

$$\bar{x} \pm t_{\frac{\alpha}{2},df} * s/\sqrt{n}$$

$\bar{x} = 580$, $n = 100$,

```r
580-qt(0.975,df = 99)*80/sqrt(100)
```

```
## [1] 564.1263
```

```r
580+qt(0.975,df = 99)*80/sqrt(100)
```

```
## [1] 595.8737
```

So the two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day is (564.1263, 595.8737).

```
d. Assuming the same information as part c, suppose that we now only want a upper-bound
```

$$\bar{x} \pm t_{\alpha,d} f * s/\sqrt{n}$$

```r
580+qt(0.95,df = 99)*80/sqrt(100)
```

```
## [1] 593.2831
```

So the one-sided 95% upper-bound confidence interval for the average amount of coffee consumed by the University of Rochester each day is 582. e. Assuming the same information as part a (i.e., known population variance), calculate the number of samples needed in order to get a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day of length 16. $X \sim N(\mu, 90)$ $2m = 16$, $m = 8$,

$$m = z_{\frac{\alpha}{2}} * \sigma\sqrt{n}$$

$$n = \lceil z_{\frac{\alpha}{2}}^2 * \sigma^2/m^2 \rceil$$

```
1.96^2*90^2/8^2
```

```
## [1] 486.2025
```

the number of samples needed in order to get a two-sided 95% confidence interval for the average amount of coffee consumed by the University of Rochester each day of length 16 is 486.2025.

Short Answers:

- About how long did this assignment take you? Did you feel it was too long, too short, or reasonable?
- Who, if anyone, did you work with on this assignment?
- What questions do you have relating to any of the material we have covered so far in class?