MATH 180A HOMEWORK #8. SOLUTIONS.

FALL 2019

1. (*ASV*[*], *Exercise 8.2*) - 2 points.

Solution.
Let X_k be the outcome of rolling k-sided die, for $k \in \{4, 6, 12\}$. By linearity,

$$E(X_4 + X_6 + X_{12}) = E(X_4) + E(X_6) + E(X_{12}).$$

For any k,

$$E(X_k) = \frac{k+1}{2}.$$

We conclude that

$$E(X_4 + X_6 + X_{12}) = \frac{4+1}{2} + \frac{6+1}{2} + \frac{12+1}{2} = \frac{25}{2}.$$

2. (*ASV, Exercise 8.8*) - 2 points.

Solution.
Let X be arrival time of plumber and Y be the time during which the plumber will be working. It is given that $X \sim \text{Unif}[1, 7]$ and $Y \sim \text{Exp}(2)$.
By linearity,

$$E(X + Y) = E(X) + E(Y) = 4 + \frac{1}{2} = 4.5.$$

By independence,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 3 + \frac{1}{4} = 3.25.$$

3. (*ASV, Exercise 8.26*) - 3 points.

Solution.
(a) Let I_i be indicator of $A_i \cap B_{i+1}$, where A_i is the even"i-th ball is green" and B_i is the event "i-th ball is yellow". Then $X_n = \sum_{i=1}^{n-1} I_i$, and

$$E(X_n) = E(\sum_{i=1}^{n-1} I_i) = \sum_{i=1}^{n-1} E(I_i)$$

[*] *Introduction to Probability*, by David F. Anderson, Timo Seppäläinen, and Benedek Valkó

by linearity. Next, by using independence, we have that

$$E(I_i) = P(A_i \cap B_{i+1}) = P(A_i)P(B_{i+1}) = \frac{4}{9} \cdot \frac{3}{9} = \frac{4}{27}.$$

(b) Let A_i be the event that k-th ball is green, and B_i be the event that there are no white balls among the first i balls. Denote by I_i the indicator of $A_i \cap B_{i-1}$. Then $Y = \sum_{i=1}^{\infty} I_i$ and by linearity and independence

$$E(Y) = E(\sum_{i=1}^{\infty} I_i) = \sum_{i=1}^{\infty} E(I_i) = \sum_{i=1}^{\infty} P(A_i \cap B_{i-1}) = \sum_{i=1}^{\infty} P(A_i)P(B_{i-1}).$$

Since $P(A_i) = \frac{4}{9}$ and $P(B_{i-1}) = (\frac{7}{9})^{i-1}$, we conclude that

$$E(Y) = \sum_{i=1}^{\infty} \frac{4}{9}\left(\frac{7}{9}\right)^{i-1} = 2.$$

4. (*ASV, Exercise 8.48*) - 2 points.

Solution.
The joint p.m.f. of X and Y is given by

$$p(1,0) = \frac{9}{100}, \quad p(2,0) = \frac{81}{100}, \quad p(2,1) = \frac{9}{100}, \quad p(3,3) = \frac{1}{100}.$$

We can now compute the marginal p.m.f. of X

$$p_X(1) = \frac{9}{100}, \quad p_X(2) = \frac{90}{100}, \quad p_X(3) = \frac{1}{100},$$

and the marginal p.m.f. of Y

$$p_Y(0) = \frac{90}{100}, \quad p_Y(1) = \frac{9}{100}, \quad p_Y(2) = \frac{1}{100}.$$

From this we can compute

$$E(X) = \frac{48}{25}, \quad E(Y) = \frac{11}{100}, \quad E(XY) = \frac{6}{25},$$

and

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{18}{625}.$$

5. (*ASV, Exercise 8.55*) - 3 points.

Solution.
Note that $I_{A^C} = 1 - I_A$ and similarly, $I_{B^C} = 1 - I_B$. Now, by the properties of correlation,

$$\mathrm{Corr}(I_{A^C}, I_{B^C}) = \mathrm{Corr}(1 - I_A, 1 - I_B) = -1 \cdot \mathrm{Corr}(I_A, 1 - I_B) = \mathrm{Corr}(I_A, I_B).$$

6. (*ASV, Exercise 9.4*) - NOT TO BE TURNED IN.

The European style roulette wheel has the following probabilities: a red number appears with probability $\frac{18}{37}$, a black number appears with probability $\frac{18}{37}$, and a green number appears with probability $\frac{1}{37}$. Ben bets exactly 1 dollar on black each round. Explain why this is not a good long term strategy.

Solution. Let X_i denote Ben's net winnings in the i-th game. Then X_i has the following distribution: $P(X_i = 1) = \frac{18}{37}$, $P(X_i = -1) = \frac{19}{37}$ (i.e., if black number appears, Ben gets 2 dollars and the net winning is equal to $2 - 1 = 1$ (subtract 1 dollar bet); if red or green number appears, Ben loses his 1 dollar bet, so the net winning is -1). We conclude that for any i,

$$E(X_i) = 1 \cdot \frac{18}{37} + (-1) \cdot \frac{19}{37} = -\frac{1}{37}.$$

The total net winnings after n games is then equal to

$$S_n = X_1 + \cdots + X_n.$$

By the Law of Large Numbers (note that X_i's are independent), for any small $\varepsilon > 0$

$$\lim_{n\to\infty} P\Big(\Big|\frac{S_n}{n} + \frac{1}{37}\Big| \le \varepsilon\Big) = 1.$$

In particular, if ε is equal to, say, $\frac{1}{2\cdot 37}$

$$\lim_{n\to\infty} P\Big(\Big|\frac{S_n}{n} + \frac{1}{37}\Big| \le \frac{1}{2\cdot 37}\Big) = \lim_{n\to\infty} P\Big(-\frac{1}{37} - \frac{1}{2\cdot 37} \le \frac{S_n}{n} \le -\frac{1}{37} + \frac{1}{2\cdot 37}\Big) = 1.$$

We get that

$$\lim_{n\to\infty} P\Big(-\frac{3n}{2\cdot 37} \le S_n \le -\frac{n}{2\cdot 37}\Big) = 1,$$

which means that after playing many games with the chosen strategy, with very high probability Ben's balance will be negative and proportional to the number of games he played (the longer he plays, the worse his balance becomes).

7. (*ASV, Exercise 9.7*) - NOT TO BE TURNED IN.

A car insurance company has 2500 policy holders. The expected claim paid to a policy holder during a year is 1000 dollars with a standard deviation of 900 dollars. What premium should the company charge each policy holder to assure that with probability 0.999, the premium income will cover the cost of the claims? Compute the answer both with Chebyshev's inequality and with the CLT.

Solution. Let X_i, $1 \le i \le 2500$ denote the amount of money paid to the i-th policyholder. It is given that $E(X_i) = 1000$, $\sqrt{\mathrm{Var}(X_i)} = 900$ and we assume that these payments are independent. Denote by p the size of the premium. Then the required condition on the size of the premium can be written as

$$P\Big(\sum_{i=1}^{2500} X_i \le 2500 \cdot p\Big) \ge 0.999.$$

In order to use Chebyshev's inequality, note that

$$E\Big(\sum_{i=1}^{2500} X_i\Big) = 2500 \cdot 1000, \quad \mathrm{Var}\Big(\sum_{i=1}^{2500} X_i\Big) = 2500 \cdot 900^2,$$

where we used independence in the second equality.

Then, by the Chebyshev's inequality,

$$P\Big(\sum_{i=1}^{2500} X_i \ge 2500 \cdot p\Big) = P\Big(\sum_{i=1}^{2500} X_i - 2500 \cdot 1000 \ge 2500 \cdot (p - 1000)\Big) \le \frac{\mathrm{Var}\Big(\sum_{i=1}^{2500} X_i\Big)}{(2500 \cdot (p-1000))^2}$$

and, thus, by taking the complement,

$$P\Big(\sum_{i=1}^{2500} X_i \le 2500 \cdot p\Big) \ge 1 - \frac{\mathrm{Var}\Big(\sum_{i=1}^{2500} X_i\Big)}{(2500 \cdot (p-1000))^2}.$$

We want the number on the right-hand side to be at least 0.999, which means that p should satisfy

$$\frac{\mathrm{Var}\Big(\sum_{i=1}^{2500} X_i\Big)}{(2500 \cdot (p-1000))^2} = \frac{2500 \cdot 900^2}{(2500 \cdot (p-1000))^2} \le 0.001.$$

This leads to

$$\begin{aligned} 2500 \cdot 900^2 \cdot 1000 &\le 2500^2 (p-1000)^2, \\ 900^2 \cdot 1000 &\le 2500 (p-1000)^2, \\ 900 \cdot 10\sqrt{10} &\le 50(p-1000), \\ 180\sqrt{10} &\le (p-1000), \\ 1000 + 180\sqrt{10} &\le p. \end{aligned}$$

Therefore, if we use the Chebyshev's inequality to estimate the probability, the size of the premium should be at least $1000 + 180\sqrt{10}$ dollars.

Estimating the same probability via the CLT gives a much better bound. First, centralize and normalize the random variable $\sum_{i=1}^{2500} X_i$

$$P\Big(\sum_{i=1}^{2500} X_i \le 2500 \cdot p\Big) = P\Big(\frac{\sum_{i=1}^{2500} X_i - 2500 \cdot 1000}{\sqrt{2500} \cdot 900} \le \frac{2500 \cdot p - 2500 \cdot 1000}{\sqrt{2500} \cdot 900}\Big).$$

By the CLT, the probability on the right-hand side of the above equality is approximately equal to

$$\Phi\Big(\frac{2500 \cdot p - 2500 \cdot 1000}{\sqrt{2500} \cdot 900}\Big) = \Phi\Big(\frac{50 \cdot (p-1000)}{900}\Big).$$

If we want this probability to be at least 0.999, we need that

$$\Phi\Big(\frac{50\cdot(p-1000)}{900}\Big)=\Phi\Big(\frac{p-1000}{18}\Big)\geq 0.999.$$

Using the table of values of Φ, this leads to the following inequality for p

$$\frac{p-1000}{18}\geq 3.08.$$

Therefore, using the CLT instead of Chebyshev's inequality gives that the premium of size $1000+18\cdot 3.08$ will assure (with probability at least 0.999) that the premium income will cover the cost of the claims.

Note that $1000+18\cdot 3.08$ is much smaller than $1000+180\sqrt{10}$, i.e., the estimate obtained by the CLT is more precise than the estimate obtained by the Chebyshev's inequality.

8. (*ASV, Exercise 9.15*) - NOT TO BE TURNED IN.

Omar and Cheryl each run a food truck on a university campus. Suppose that the number of customers that come to Omar's truck on a given day has a Poisson(220) distribution, whereas the number of customers that come to Cheryl's truck on a given day has a Poisson(210) distribution. What can be said about the probability that more customers will come to Omar's truck than Cheryl's truck in the long run? Provide random variables for the number of customers arriving to each food truck each day and give a clear mathematical statement. Be precise about the assumptions you are making.

Solution. Let X_i be the number of customers coming to Omar's food truck on i-th day, and let Y_i be the number of customers coming to Cheryl's food truck on the same day. Then during the period of n days, there are $\sum_{i=1}^n X_i$ customers coming to Omar's truck, and $\sum_{i=1}^n Y_i$ customers coming to Cheryl's.

You are asked to estimate the probability, that $\sum_{i=1}^n X_i > \sum_{i=1}^n Y_i$ as n becomes big. Rewrite the probability of this event as

$$P\Big(\sum_{i=1}^n X_i > \sum_{i=1}^n Y_i\Big)=P\Big(\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i > 0\Big)=P\Big(\sum_{i=1}^n (X_i-Y_i) > 0\Big).$$

It is given, that $X_i\sim$ Poisson(220) and $Y_i\sim$ Poisson(210), so that

$$E(X_i)=220,\quad E(Y_i)=210,\quad E(X_i-Y_i)=220-210=10,$$

where in the last equality we used the linearity of the expectation.

Assume that the numbers of customers coming to the food trucks on differents days are independent. Under this assumption we can apply the LLN to the sequence of random variables X_i-Y_i, $i\in\{1,2,3,\ldots\}$:

$$P\Big(\sum_{i=1}^n (X_i-Y_i) > 0\Big)=P\Big(\frac{\sum_{i=1}^n (X_i-Y_i)}{n}-10>-10\Big)\to 1,\quad n\to\infty.$$

We conclude, that the probability that Omar will have more customers during the period of n days than Cheryl tends to 1 as the parameter n (length of the period of observation) goes to infinity (becomes longer and longer).

9. (*ASV, Exercise 9.16*) - NOT TO BE TURNED IN.
Every morning I take either but number 5 or bus number 8 to work. Every morning the waiting time for the number 5 is exponential with mean 10 minutes, which the waiting time for the number 8 is exponential with mean 20 minutes. Assume all waiting times are independent of each other. Let S_n be the total amount of bus-waiting (in minutes) that I have done during n mornings, and let T_n be the number of times I have taken the number 5 bus during n morning.
(a) Find the limit $\lim_{n\to\infty} P(S_n \le 7n)$.
(b) Find the limit $\lim_{n\to\infty} P(T_n \ge 0.6n)$.

Solution.
Denote by U_i the waiting time for the bus number 5 on the i-th morning, and by V_i the waiting time for the bus number 8 on the same morning. Then $U_i \sim \text{Exp}(1/10)$ and $V_i \sim \text{Exp}(1/20)$.
Now the random variable $I_i = \min\{U_i, V_i\}$ gives the waiting time on morning i (you wait until any of the two buses arrives), and the random variable

$$J_i = \begin{cases} 1 & \text{if } U_i < V_i, \\ 0 & \text{otherwise} \end{cases},$$

is the indicator of the event that on morning i you take the bus number 5 (the waiting time for bus number 5 is strictly smaller then the waiting time for the bus number 8). Therefore,

$$S_n = \sum_{i=1}^{n} I_i, \quad T_n = \sum_{i=1}^{n} J_i.$$

Since for each i, U_i and V_i are independent, $\min\{U_i, V_i\} \sim \text{Exp}(\frac{1}{10} + \frac{1}{20})$, i.e., $I_i \sim \text{Exp}(3/20)$. On the other hand, $J_i \sim \text{Ber}(p)$ with $p = P(U_i < V_i)$.

$$P(U_i < V_i) = \int_0^\infty \int_x^\infty \frac{1}{10} e^{-\frac{1}{10}x} \frac{1}{20} e^{-\frac{1}{20}y} dy dx = \int_0^\infty \frac{1}{10} e^{-\frac{1}{10}x} e^{-\frac{1}{20}x} dx = \frac{\frac{1}{10}}{\frac{1}{10} + \frac{1}{20}} = \frac{2}{3},$$

so that $J_i \sim \text{Ber}(2/3)$.
Now we have all the necessary information to compute the limits in (a) and (b).
(a)

$$\lim_{n\to\infty} P(S_n \le 7n) = \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^n I_i}{n} \le 7\Big) = \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^n I_i}{n} - \frac{20}{3} \le 7 - \frac{20}{3}\Big) = 1$$

by the LLN (I_i's are i.i.d. with mean $20/3$, and $7 - 20/3 > 0$).
(b)

$$\lim_{n\to\infty} P(T_n \ge 0.6n) = \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^n J_i}{n} \ge 0.6\Big) = \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^n J_i}{n} - \frac{2}{3} \ge 0.6 - \frac{2}{3}\Big) = 1$$

by the LLN (J_i's are i.i.d. with mean $2/3$ and $0.6 - 2/3 < 0$).

10. (*ASV, Exercise 9.20*) - NOT TO BE TURNED IN.

Let $X_1, X_2, X_3, \dots$ be i.i.d. random variables with mean zero and finite variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Determine the limits below, with precise justifications.

(a) $\lim_{n\to\infty} P(S_n \geq 0.01n)$.

Solution.

$$\lim_{n\to\infty} P(S_n \geq 0.01n) = \lim_{n\to\infty} P\Big(\frac{S_n}{n} \geq 0.01\Big) = 0$$

by the LLN (X_i's are i.i.d. with mean 0).

(b) $\lim_{n\to\infty} P(S_n \geq 0)$.

Solution.

$$\lim_{n\to\infty} P(S_n \geq 0) = \lim_{n\to\infty} P\Big(\frac{S_n}{\sqrt{n}\sigma} \geq 0\Big) = 1 - \Phi(0) = 0.5$$

by the CLT (X_i's are i.i.d. with mean 0 and finite variance).

(c) $\lim_{n\to\infty} P(S_n \geq -0.01n)$.

Solution.

$$\lim_{n\to\infty} P(S_n \geq -0.01n) = \lim_{n\to\infty} P\Big(\frac{S_n}{n} \geq -0.01\Big) = 1$$

by the LLN (X_i's are i.i.d. with mean 0).