45-750 PROBABILITY AND STATISTICS
FALL 2019

PROBLEM SET 1: BASIC PROBABILITY, DISCRETE RANDOM VARIABLES

This first assignment tests your understanding of basic probability models, sample space and the various rules for computing probabilities of events represented as sets, conditional probability and independence, variance, correlation, and the binomial and Poisson distributions. It is based on Chapters 4 and 5 of the text.

Problem 1 (**Sample Space**). You toss a fair coin (fair: toss returns head and tail with equal probability, that is $P(H) = P(T) = \frac{1}{2}$) until you get exactly two heads. For example the sequence HTH represents the outcome where the first toss was head, the second tail, and the third head (reading the sequence from left to right); the experiment stops since we tossed exactly two heads; therefore the sequence HTH represents a basic outcome of the experiment.

(1.1) **Easy.** Which of the following sequences are basic outcomes of this experiment? TTH, HTH, HHT, $THTH$

1.1 Answer: HTH, $THTH$ are basic outcomes. TTH is not a basic outcome because the sequence does not have enough heads to stop. HHT is not a basic outcome because the experiment should have been stopped after two heads tossed.

(1.2) **Moderate.** Describe the sample space, namely the collection of all basic outcomes.

1.2 Answer: The sample space consists of sequences which contain exactly 2 H, and end with H. In other word, let (T^*) denote a sequence of any number $(0, 1, 2, \ldots)$ of T, then a basic outcome has format: $(T^*)H(T^*)H$.

(1.3) **Moderate.** What is the probability that only two tosses are required in this experiment.

1.3 Answer: The only basic outcome with only two tosses is HH. Each toss gives H with probability $\frac{1}{2}$, and all tosses are independent with each other. So the probability of getting outcome HH is $P(HH) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$.

(1.4) **Challenging.** What is the probability that exactly three tosses are required in this experiment.

1.4 Answer: There are two basic outcomes requiring exactly three tosses: HTH, THH. Using same argument as in part (1.3), $P(HTH) = P(THH) = \frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{8}$. Therefore,

$$P(\text{exactly three tosses are required}) = P(HTH) + P(THH) = \frac{1}{4}.$$

Problem 2 (**Events, Probability Table, Independence**). 90% of American adults have a high school education, 35% have a four-year college degree, and 13% have an advanced degree (Master's, professional degree or doctorate). We assume that adults who have an advanced degree also have a four-year college degree, and those who have a four-year college degree also have a high school education. 50% of American adults earn \$30,000 or more per year. But only 15% of those without a high school education and 30% of those without a four-year college degree earn more than \$30,000 or more per year.

(2.1) **Easy.** Let A denote the event that an American adult has no high school education, B the event that the adult has a high school education but no four-year college degree, C the event that the adult has a four-year college degree but no advanced degree, and D the event that the adult has an advanced degree. Let M denote the event that the adult earns \$30,000 or more per year, and L the event that the adult earns less. What is the fraction of American adults who earn less than $30,000 per year and have no high school education. (Hint: You are asked to compute $P(A \cap L)$.) Are A and L independent events?

2.1 Answer: Since 15% of those without a high school education earn $30,000 or more per year, $P(M|A) = 0.15$. Using definition of conditional probability, we have

$$P(M|A) = \frac{P(M \bigcap A)}{P(A)} = 0.15$$
$$\Rightarrow P(M \bigcap A) = 0.15 * P(A) = 0.15 * (1 - 0.9) = 0.015$$

An adult with no high school education either earns less than $30,000 per year or more than that, hence $P(A) = P(A \bigcap M) + P(A \bigcap L)$. So $P(A \bigcap L) = P(A) - P(A \bigcap M) = 0.1 - 0.015 = 0.085$. 8.5% of Americans earn less than $30,000 per year and have no high school education.
A, L are not independent events because $P(A) * P(L) = 0.1 * 0.5 = 0.05 \neq P(A \bigcap L)$.

(2.2) **Moderate.** What is the fraction of Americans who have a high school education, do not have a four-year college degree and earn $30,000 or more per year. (Hint: You are asked to compute $P(B \cap M)$.)

2.2 Answer: Let E denote the event that an American adult does not have a four-year college education, then $P(E) = 1 - 0.35 = 0.65$. Since 30% of those without a four-year college degree earn $30,000 or more per year, we have conditional probability $P(M|E) = 0.3$. Similarly, with definition of conditional probability,

$$P(M|E) = \frac{P(M \bigcap E)}{P(E)} = 0.3 \Rightarrow P(M \bigcap E) = 0.3 * P(E) = 0.3 * 0.65 = 0.195$$

By definition, event E is a disjoint union of events A and B, so $P(M \bigcap E) = P(M \bigcap A) + P(M \bigcap B)$, then $P(B \bigcap M) = P(M \bigcap E) - P(M \bigcap A) = 0.195 - 0.015 = 0.18$.
18% of Americans have a high school education, do not have a four-year college degree and earn $30,000 or more per year.

(2.3) **Moderate.** What is the fraction of Americans who have a four-year college degree and earn $30,000 or more per year.

2.3 Answer: We computed in 2.2 that the fraction of American adults who have no four-year college degree and earn $30,000 or more per year is $P(M \bigcap E) = 0.195$. It is given that 50% of American adults earn $30,000 or more per year, which means $P(M) = 0.5$.
Let F denote the event that an American adult has a four-year college education, then $P(M) = P(M \bigcap E) + P(M \bigcap F)$. So $P(M \bigcap F) = 0.305$, 30.5% of American adults have a four-year college degree and earn $30,000 or more per year.

(2.4) **Challenging.** Assuming that 100% of American adults with an advanced degree earn $30,000 or more per year, what is the fraction of Americans who have a four-year college degree, no advanced degree and earn less than $30,000 per year.

2.4 Answer: First, note that the question asks to compute $P(C \bigcap L)$.
It is given that 50% of American adults earn $30,000 or more per year, which means $P(M) = 0.5$. 100% of American adults with an advanced degree earn $30,000 or more per year means $P(M|D) = 1$, so $P(M \bigcap D) = P(M|D) * P(D) = 1 * 0.13 = 0.13$. We also computed in 2.2 answer that the fraction of American adults with no four-year college degree and earn $30,000 or more per year is $P(M \bigcap E) = P(M \bigcap A) + P(M \bigcap B) = 0.195$. Since events A, B, C, D are disjoint and exhaustive, $P(M \bigcap C) = P(M) - (P(M \bigcap A) + P(M \bigcap B) + P(M \bigcap D)) = 0.175$. An American adult belonging to event C either belongs to M or L, so $P(C \bigcap L) = P(C) - P(C \bigcap M) = (0.35 - 0.13) - 0.175 = 0.045$. 4.5% of Americans have a four-year college degree, no advanced degree and earn less than $30,000 per year.

Problem 3 (**Conditional Probability, Independence**)**.** According to a survey, the fraction of Americans who like science-fiction movies is 0.2 and the fraction interested in role-playing games is 0.1. Also, the proportion of the population interested in neither science-fiction nor role-playing games is 0.75.

(3.1) **Easy.** What is the probability that an American is interested in both science-fiction movies and role-playing games?

3.1 Answer: Let S denote the event that an American likes science-fiction movies, R denote the event that an American likes role-playing games, $\overline{S}, \overline{R}$ denote the complement events. From problem statement, we know $P(S) = 0.2, P(R) = 0.1, P(\overline{S} \bigcap \overline{R}) = 0.75$.
Event $\overline{S} \bigcap \overline{R}$ is equivalent to $\overline{S \bigcup R}$, so $1 - P(\overline{S} \bigcap \overline{R}) = 1 - P(\overline{S \bigcup R}) = P(S \bigcup R) = 0.25$.

$$P(S \bigcup R) = P(S) + P(R) - P(S \bigcap R) \Rightarrow P(S \bigcap R) = 0.2 + 0.1 - 0.25 = 0.05$$

The probability that an American is interested in both science-fiction movies and role-playing games is $P(S \bigcap R) = 0.05$.

(3.2) **Easy.** What is the conditional probability that a person is interested in science-fiction given that (s)he is interested in role-playing?

3.2 Answer: $P(S|R) = \frac{P(S \bigcap R)}{P(R)} = \frac{0.05}{0.1} = 0.5$, so the conditional probability that a person is interested in science-fiction given that (s)he is interested in role-playing is 0.5.

(3.3) **Easy.** Are the events "being interested in science-fiction" and "being interested in role-playing games" independent?

3.3 Answer: They are not independent because $P(S|R) \neq P(S)$.

Problem 4 (**Portfolio of Assets**)**.** We are considering three assets A, B, C whose rates of return have the following probability distributions, depending on the state of the economy in the coming year.

		A	B	C
Economy	Proba	Rate of Return		
Strong	0.3	0.18	−0.04	0.07
Moderate	0.3	0.12	0.02	0.07
Weak	0.3	0.04	0.14	0.04
Recession	0.1	0.02	0.16	0.04

(4.1) **Easy.** What is the expected rate of return of asset A ? of asset B ? of asset C ?

4.1 Answer: Let X, Y, Z be random variables respectively denoting the rate of return of asset A, B, C. Then expected return of asset A, B, C is:

$$E(X) = 0.3 * 0.18 + 0.3 * 0.12 + 0.3 * 0.04 + 0.1 * 0.02 = 0.104$$
$$E(Y) = -0.3 * 0.04 + 0.3 * 0.02 + 0.3 * 0.14 + 0.1 * 0.16 = 0.052$$
$$E(Z) = 0.3 * 0.07 + 0.3 * 0.07 + 0.3 * 0.04 + 0.1 * 0.04 = 0.058$$

(4.2) **Moderate.** What are the corresponding standard deviations ?

4.2 Answer: We first compute variance using formula $Var(X) = E(X^2) - E(X)^2$.

$$E(X^2) = 0.3 * 0.18^2 + 0.3 * 0.12^2 + 0.3 * 0.04^2 + 0.1 * 0.02^2 = 0.01456$$
$$E(Y^2) = 0.3 * 0.04^2 + 0.3 * 0.02^2 + 0.3 * 0.14^2 + 0.1 * 0.16^2 = 0.00904$$
$$E(Z^2) = 0.3 * 0.07^2 + 0.3 * 0.07^2 + 0.3 * 0.04^2 + 0.1 * 0.04^2 = 0.00358$$

$$Var(X) = 0.01456 - 0.104^2 = 0.003744, Var(Y) = 0.006336, Var(Z) = 0.000216$$
$$StDev(X) = \sqrt{Var(X)} = 0.061188, \quad StDev(Y) = 0.079599, \quad StDev(Z) = 0.014697$$

(4.3) **Challenging.** Compute the covariance between the rates of return of assets A and B.

4.3 Answer: We compute the covariance between the rates of return of assets A, B using the definition: $cov(X,Y) = E[(X - E(X))(Y - E(Y))]$.

$$\begin{aligned} cov(X,Y) = E[(X - 0.104)(Y - 0.052)] \\ &= 0.3 * (0.18 - 0.104) * (-0.04 - 0.052) + 0.3 * (0.12 - 0.104) * (0.02 - 0.052) \\ &\quad + 0.3 * (0.04 - 0.104) * (0.14 - 0.052) + 0.1 * (0.02 - 0.104) * (0.16 - 0.052) \\ &= -0.00485 \end{aligned}$$

(4.4) **Moderate.** Portfolio P is invested 60% in asset A and 40% in asset B. What are the expected value and standard deviation of the rate of return of portfolio P ?

4.4 Answer: Let W be a random variable denoting the rate of return of portfolio P. That is $W = 60\%X + 40\%Y$. Therefore,

$$E(W) = 0.6 * E(X) + 0.4 * E(Y) = 0.0832$$

$$Var(W) = 0.6^2 * Var(X) + 0.4^2 * Var(Y) + 2 * 0.6 * 0.4 * cov(X,Y)$$

$$= 0.36 * 0.061188 + 0.16 * 0.079599 + 0.48 * (-0.00485) = 0.00003456$$

$$StDev(W) = \sqrt{Var(W)} = 0.005879$$

(4.5) **Moderate.** Rank the 4 investments A, B, C and P from least risky to most risky.

4.5 Answer: More risky investment has higher standard deviation in its rate of return, so we compare the standard deviations to X, Y, Z, W. $StDev(W) < StDev(Z) < StDev(X) < StDev(Y)$, ranking from least risky to most risky is: P, C, A, B.

Problem 5 (**Poisson Distribution**). Cars and trucks arrive at the entrance of the turnpike according to independent Poisson processes. The rate of arrival of cars is 10 per minute and that of trucks is 2 per minute. The entry toll for cars is \$4 and for trucks it is \$10.

(5.1) **Easy.** What is the expected number of vehicles arriving at the entrance of the turnpike between 6:00PM and 7:00PM ?

5.1 Answer: Cars arrive with rate 10 cars per minute, and trucks arrive with rate 2 trucks per minute. In total, vehicles arrive at the entrance with rate 12 vehicles per minute. The expected number of vehicles arrived between 6:00PM and 7:00PM, which is 60 minutes, is $12 * 60 = 720$.

(5.2) **Easy.** What is the expected value of the tolls collected between 6:00PM and 7:00PM ?

5.2 Answer: Each car's entry toll is \$4, each truck's entry toll is \$10, so the expected value of tolls collected in 1 minute is $4*10+10*2 = 60$. The expected value of tolls collected between 6:00PM and 7:00PM (60 minutes) is $60 * 60 = 3600$.

(5.3) **Moderate.** What is the probability that at least 12 vehicles will arrive at the entrance of the turnpike during the next minute given that 10 cars but no truck arrived at the entrance of the turnpike during the previous minute.

5.3 Answer: In Poisson processes the probabilities of events happening in disjoint intervals are independent of each other. Hence 10 cars but no truck arriving at the entrance of the turnpike during the previous minute has no effect on the probabilities of having at least 12 vehicles arrive at the entrance of the turnpike during the next minute.
Let X be a random variable denoting the number of vehicles arriving at the entrance in a minute. Vehicles arrive in a poisson process with rate 12 per minute. Then the probability is

$$P(X \geq 12) = 1 - P(X \leq 11) = 1 - POISSON.DIST(11, 12, 1) = 0.5384$$

Problem 6 (**Binomial Distribution**). The manager of a restaurant is considering the possibility of overbooking its reservations. The manager knows from experience that there is a 20% probability that a party who reserved a table will not show up (and these "no show" events are independent across parties who reserved a table). The restaurant has 25 tables, and each table can be used once during lunch time. The manager is considering accepting 30 reservations each day for lunch.

(6.1) **Moderate.** What is the probability that all parties who made a reservation and show up at the restaurant can be assigned a table?

6.1 Answer: Let S be the number of parties who show up. The random variable S has a binomial distribution with $n = 30$ and $p = 0.8$. Then

$$\begin{aligned} P(\text{All parties who reserved and show up get a table}) &= P(S \leq 25) \\ &= BINOM.DIST(25, 30, 0.8, 1) \\ &= 0.7448 \end{aligned}$$

(6.2) **Moderate.** What is the probability that no table is empty?

6.2 Answer: For no table to be empty, at least 25 parties should show up. That is

$$\begin{aligned} P(\text{No table is empty}) &= P(S \geq 25) = 1 - BINOM.DIST(S \leq 24) \\ &= 1 - BINOM.DIST(24, 30, 0.8, 1) \\ &= 0.4275 \end{aligned}$$

(6.3) **Challenging.** The manager knows the negative impact to the restaurant's reputation of customers who show up but can't be assigned a table even though they have a reservation. The manager estimates that she can't afford more than a 5% probability that a customer with a reservation can't find a table. What is the maximum number of reservations that the restaurant should accept?

6.3 Answer: The restaurant wants to keep accepting reservations as long as the probability that a customer with a reservation gets a table remains above .95. Let us consider various reservation numbers below 30.

$$\begin{aligned} BINOM.DIST(25, 29, .8, 1) &= 0.8596 \\ BINOM.DIST(25, 28, .8, 1) &= 0.9388 \\ BINOM.DIST(25, 27, .8, 1) &= 0.9813 \end{aligned}$$

Therefore, the restaurant should take at most 27 reservations.