Econ 21020 - Problem Set 1 SOLUTIONS

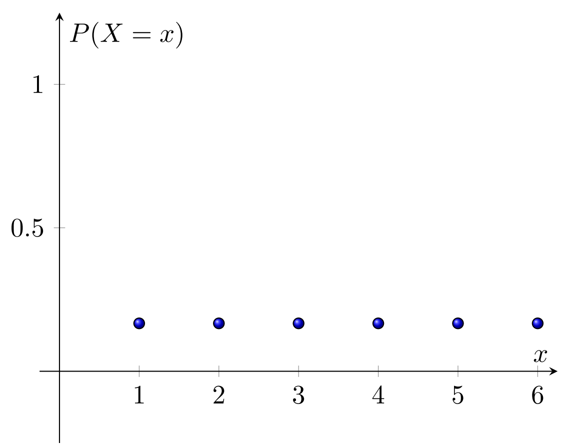
Due 10/11 by Start of Class

Problem 1

(a) Express the probability mass function and cumulative density function for a balanced 6-sided die (numbered 1-6) and draw images representing them (images need not be extremely precise - they just need to capture the important attributes of the functions).

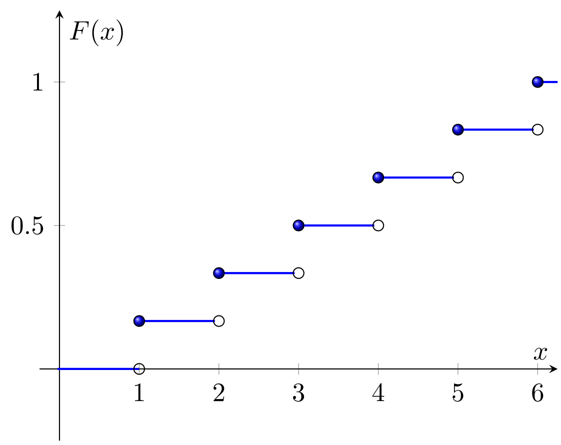
SOLUTION: Because the die is "balanced," there will be an equal likelihood of each of the 6 sides coming up on a given roll. Because the sum of the pmf across all values must equal 1, we can thus infer that each side has a $\frac{1}{6}$ likelihood. Calling the outcome of the roll X:

$$p(x) = \begin{cases} \frac{1}{6} \; if \; x \in \{1,2,3,4,5,6\} \\ 0 \; otherwise \end{cases}$$



We know that the CDF is equal to $F(x) = P(X \leq x)$. From the above this means the CDF will be zero for any value less than 1, will bump up by 1/6 at each potential value of a die roll, and is 1 after 6:

$$F(x) = \begin{cases} 0 \; if \; x < 1 \\ \frac{1}{6} \; if \; x \in [1,2) \\ \frac{2}{6} \; if \; x \in [2,3) \\ \frac{3}{6} \; if \; x \in [3,4) \\ \frac{4}{6} \; if \; x \in [4,5) \\ \frac{5}{6} \; if \; x \in [5,6) \\ 1 \; if \; x \geq 6 \end{cases}$$



(b) Calculate the expectation and variance of the outcome of a balanced 6-sided die roll. Calculate the expectation of the square of the outcome of a balanced 6-sided die roll (i.e. $E[X^2]$ if X represents the outcome of the roll).

SOLUTION: The expectation can be calculated directly from the given definition, now that we know the pmf:

$$\begin{aligned} E[X] &= \sum_{i=1}^{6} x_i P(X = x_i) \\ &= \frac{1}{6}(1+2+3+4+5+6) \\ &= 3.5 \end{aligned}$$

For the variance, we can calculate directly from the definition $Var(X) = E[(X - E[X])^2]$ in a manner similar to what we did for the expectation, or we can solve for $E[X^2]$ first, and use that in the alternative formulation. Opting for the latter, we use the given formula for the expectation of a function of a random variable:

$$
\begin{aligned}
E[X^2] &= \sum_{i=1}^{6} x_i^2 P(X = x_i) \\
&= \frac{1}{6}(1+4+9+16+25+36) \\
&= 15.167
\end{aligned}
$$

Finally we can plug in the above with the regular expectation we got above to solve for the variance:

$$
\begin{aligned}
Var(X) &= E[X^2] - E[X]^2 \\
&= 15.167 - 3.5^2 \\
&= 15.167 - 12.25 \\
&= 2.917
\end{aligned}
$$

(c) In class, we saw the CDF of a Uniform[a,b] variable. Explain and demonstrate how this CDF could be derived from the PDF we saw (recreated below):

$$
f(x) = \begin{cases} \frac{1}{b-a} & if\ a \leq x \leq b \\ 0 & if\ otherwise \end{cases}
$$

SOLUTION: We know that the CDF is defined $F(x) = P(X \leq x)$. This can be computed by taking the integral of the pdf over the range from $-\infty$ to x for each $x \in \mathbb{R}$. However, we can see that, as there is 0 probability of drawing values less than a or above b, this problem simplifies, as we will have $F(x)$ equal to 0 for any value below a and equal to 1 for any value above b. For any point $x \in [a, b]$, we can simply consider the integral over the range $[a, x]$, as the integral over the range $[-\infty, a)$ contributes 0 to the total integral.

$$
\begin{aligned}
\int_a^x \frac{1}{b-a} dx &= \frac{1}{b-a} \int_a^x dx \\
&= \frac{1}{b-a} [x]_a^x \\
&= \frac{x-a}{b-a}
\end{aligned}
$$

Thus, we can express the full CDF as:

$$
F(x) = \begin{cases} 0 & if\ x < a \\ \frac{x-a}{b-a} & if\ a \leq x \leq b \\ 1 & if\ x > b \end{cases}
$$

as seen in class.

Problem 2

Say that n is some counting number (1,2,3,...) and that there are a collection of independent Bernoulli random variables, $(X_1, X_2, X_3, ..., X_n)$, each of which is equal to 1 with probability p and equal to 0 with probability $1-p$ (this is identical across all of them). Define $X^* = \sum_{i=1}^{n} X_i$.

(a) Show that $E[X^*] = np$. Justify all steps.

SOLUTION: Let's first see what the expectation of any X_i is. Using our formula for expectations:

$$\begin{aligned} E[X_i] &= 0P(X=0) + 1P(X=1) \\ &= 0(1-p) + 1p \\ &= p \end{aligned}$$

We also note that the above is true $\forall\, i \in \{1, ..., n\}$, as we are told that all of the random variables have an identical probability distribution. Then, we can say:

$$\begin{aligned} E[X^*] &= E[\sum_{i=1}^{n} X_i] \\ &= \sum_{i=1}^{n} E[X_i] && \text{(Props of Expecs)} \\ &= \sum_{i=1}^{n} p \\ &= np \end{aligned}$$

(b) Show that $Var(X^*) = np(1-p)$. Justify all steps.

SOLUTION: Similarly, let's start by seeing the variance of any X_i is. In order to do so, let's caculate the expectation of X_i^2, using the formula for expectation of a function of a random variable:

$$\begin{aligned} E[X_i^2] &= 0^2P(X=0) + 1^2P(X=1) \\ &= 0(1-p) + 1p \\ &= p \end{aligned}$$

We note that this is $\forall\, i \in \{1, ..., n\}$, again because of the identical distributions. We then plug this into a formula for variance,

$$\begin{aligned} Var(X_i) &= E[X_i^2] - E[X_i]^2 \\ &= p - p^2 \\ &= p(1-p) \end{aligned}$$

which again is true for all i. Finally, we can say:

$$
\begin{aligned}
Var(X^*) &= Var(\sum_{i=1}^{n} X_i) \\
&= \sum_{i=1}^{n} Var(X_i) && \text{(Indep of } X_i\text{)} \\
&= \sum_{i=1}^{n} p(1-p) \\
&= np(1-p)
\end{aligned}
$$

Problem 3

The following table represents the joint probability mass function of college graduation status and employment status among the working-age population of South Africa.

	Unemployed (Y=0)	Employed (Y=1)
Non-college grads (X=0)	0.078	0.673
College grads (X=1)	0.042	0.207

(a) Explain in words what the number 0.078 in the top-left cell means.

SOLUTION: The 0.078 is the joint probability of being unemployed and being a non-college grad. This means that if we were to select a person from the working-age population of South Africa at random, then there would be 0.078 chance of selecting a person who has the characteristics of both being unemployed and not being a college graduate.

(b) Calculate the marginal probability of being unemployed ($P(Y = 0)$) and the marginal probability of being a college graduate ($P(X = 1)$).

SOLUTION: We know that the marginal probability of one jointly distributed variable taking on a given value is equal to the sum of the joint probabilities of all random vectors in which our variable of interest takes on the specified value. Thus, we need to sum the two joint probabilities that involve being unemployed to get the marginal probability of unemployment:

$$\begin{aligned}P(Y=0) &= P(Y=0, X=0) + P(Y=0, X=1)\\ &= 0.078 + 0.042\\ &= 0.12\end{aligned}$$

Similarly, there are two joint probabilities that involve being a college graduate:

$$\begin{aligned}P(X=1) &= P(Y=0, X=1) + P(Y=1, X=1)\\ &= 0.042 + 0.207\\ &= 0.249\end{aligned}$$

(c) Calculate the likelihood of unemployment both for non-college grads and for college grads ($P(Y=0|X=0)$ and $P(Y=0|X=1)$).

SOLUTION: We know that the conditional probability of an event is equal to a ratio with the joint probability of the event of interest and the conditioning event in the numerator and the marginal probability of the conditioning event in the denominator:

$$\begin{aligned}P(Y=0|X=0) &= \frac{P(Y=0, X=0)}{P(X=0)}\\ &= \frac{0.078}{0.078+0.673}\\ &= \frac{0.078}{0.751}\\ &= 0.104\end{aligned}$$

where we used the fact that marginal probabilities are equal to sums of joint probabilities in the second line. Similarly, the conditional probability of unemployment for college grads is:

$$\begin{aligned}P(Y=0|X=1) &= \frac{P(Y=0, X=1)}{P(X=1)}\\ &= \frac{0.042}{0.249}\\ &= 0.169\end{aligned}$$

where we borrowed a result from part b) of this question in the second line.

(d) Are college graduation status and employment status independent? Are they mean independent? Demonstrate that they are or are not.

SOLUTION: The definition of independence is that

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all sets A and B. If we can find even one counterexample to the above, then independence does not hold. For instance, using probabilities that were given or that we have already calculated, and setting $A = \{0\}$ and $B = \{0\}$, by the above notation, we can say

$$\begin{aligned} P(X = 0, Y = 0) &= 0.078 \\ P(X = 0)P(Y = 0) &= 0.751 * 0.12 \\ &= 0.090 \\ 0.078 &\neq 0.090 \\ \Rightarrow P(X = 0, Y = 0) &\neq P(X = 0)P(Y = 0) \end{aligned}$$

Thus, college graduation status and employment status are not independent.

Mean independence implies that the conditional expectation is equal to the unconditional expectation. Again, if we can find even one counterexample to this, we can demonstrate that there is no mean independence. Also, note that because mean independence is not symmetric, we should show that Y is not mean independent of X and that X is not mean independent of Y separately. Going in that order:

$$P(Y = 0|X = 0) = 0.169 \neq 0.12 = P(Y = 0)$$

Thus, Y is not mean independent of X.

$$\begin{aligned} P(X = 1|Y = 0) &= \frac{P(X = 1, Y = 0)}{P(Y = 0)} \\ &= \frac{0.042}{0.12} \\ &= 0.35 \end{aligned}$$

$$P(X = 1|Y = 0) = 0.35 \neq 0.249 = P(X = 1)$$

so X is also not mean independent of Y.

As a final note, the fact that independence implies mean independence means that we could have merely showed that we do not have mean independence, and then said that this implies no independence.

Problem 4

Prove each of the following statements using the results on the slide "Properties of Expectations" and the two given definitions of covariance and two given definitions of variance. Justify all steps.

(a) $\text{Cov}(X, a) = 0$

SOLUTION:

$$\begin{aligned} Cov(X, a) &= E[Xa] - E[X]E[a] \\ &= aE[X] - aE[X] && \text{(Prop of expecs)} \\ &= 0 \end{aligned}$$

(b) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

SOLUTION:

$$\begin{aligned} \text{Cov}(X + Y, Z) &= E[(X + Y)Z] - E[X + Y]E[Z] \\ &= E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z] \\ & \qquad \text{(Prop of expecs)} \\ &= (E[XZ] - E[X]E[Z]) + (E[YZ] - E[Y]E[Z]) \\ &= Cov(X, Z) + Cov(Y, Z) \end{aligned}$$

(c) $\text{Cov}(a + bX, Y) = b\text{Cov}(X, Y)$

SOLUTION:

$$\begin{aligned} \text{Cov}(a + bX, Y) &= E[(a + bX)Y] - E[a + bX]E[Y] \\ &= aE[Y] + bE[XY] - (aE[Y] + bE[X]E[Y]) \\ & \qquad \text{(Prop of expecs)} \\ &= b(E[XY] - E[X]E[Y]) \\ &= bCov(X, Y) \end{aligned}$$

(d) $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

SOLUTION:

$$\begin{aligned} Var(X + Y) &= E[(X + Y)^2] - E[X + Y]^2 \\ &= E[X^2 + 2XY + Y^2] - (E[X]^2 + 2E[X]E[Y] + E[Y]^2) \\ & \qquad \text{(Prop of expecs)} \\ &= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2) + 2(E[XY] - E[X]E[Y]) \\ & \qquad \text{(Prop of expecs)} \\ &= Var(X) + Var(Y) + 2Cov(X, Y) \end{aligned}$$

Problem 5

Prove that, for conditional variance, it is true that,

$$\text{Var}[g(X) + h(X)Y|X] = h^2(X)\text{Var}[Y|X]$$

for functions $g()$ and $h()$ and random variables X and Y using the definition of conditional variance and properties of conditional expectations. Justify all steps.

SOLUTION:

$$\begin{aligned}
\text{Var}[g(X)+h(X)Y|X] &= E[(g(X)+h(X)Y-E[g(X)+h(X)Y|X])^2|X] && \text{(Def of cond var)}\\
&= E[(g(X)+h(X)Y-g(X)-h(X)E[Y|X])^2|X] && \text{(Prop i of CE)}\\
&= E[(h(X)(Y-E[Y|X]))^2|X] \\
&= h(X)^2E[(Y-E[Y|X])^2|X] && \text{(Prop i of CE)}\\
&= h(X)^2Var(Y|X)
\end{aligned}$$

Problem 6

(a) Give an example of two random variables that are uncorrelated but not mean independent (different from any examples given in the notes) and show that the former property holds and the latter does not.

SOLUTION: There is an infinite range of examples that will fulfill this. One way to cook up an example that will fill this is to realize that uncorrelatedness follows from having a zero covariance, and we can induce a zero covariance by, for instance, having it be the case that

$$E[X] = E[XY] = 0$$

(can sub in $E[Y]$ above instead). This can be achieved by having both X and XY be symmetric around 0. However, we can allow for a lack of mean independence by having Y depend on X in a way that does not break the symmetry of XY around 0. For instance, we can vary Y only with the magnitude of X. This description characterizes the example given in class or the following alternative:

$$X \sim U[-1,1]$$

$$Y = \begin{cases} 0 \; if \; |X| > 0.5 \\ 1 \; if \; |X| \leq 0.5 \end{cases}$$

$$E[X] = \frac{-1+1}{2} = 0$$

$$\begin{aligned}
E[XY] &= E[XY||X|>0.5]P(|X|>0.5) + E[XY||X|\leq 0.5]P(|X|\leq 0.5) && \text{(LIE)}\\
&= 0 + E[X||X|\leq 0.5]\frac{1}{2}
\end{aligned}$$

The expectation in the above is the same as the expectation of $Z \sim U[-0.5, 0.5]$, which is also 0. Thus,

$$E[X] = E[XY] = 0 \Rightarrow Cov(X, Y) = 0 \Rightarrow Corr(X, Y) = 0$$

However,

$$\begin{aligned} E[Y] &= 0 * P(Y = 0) + 1 * P(Y = 1) \\ &= \frac{1}{2} \\ E[Y|X > 0.5] &= 0 \\ \Rightarrow E[Y] &\neq E[Y|X] \end{aligned}$$

so Y is not mean independent of X (one example of a conditional expectation not equalling an unconditional one is sufficient to show the general statement).

(b) Give an example of two random variables that are mean independent but not independent (different from any examples given in the notes) and show that the former property holds and the latter does not.

SOLUTION: Again, there is an infinite range of possible examples here. The general task will be to cook up an X, Y, A, and B so that we can have

$$P(X \in A, Y \in B) \neq P(X \in A)P(Y \in B)$$

while maintaining mean independence. This is, perhaps, easiest done by using an induced distribution - a distribution where one variable's distribution depends on the other's. Then, if we have either variable be symmetric around 0 in all cases, we can have the unconditional and all conditional expectations equal one another at 0. The example from class did this by having the induced distribution be symmetric around 0. We can also do this by having the non-induced variable be symmetric around 0. Interestingly, we can use the example from part a) again here, as X is mean independent of Y, but independence does not hold. We already stated that $E[X||X| \leq 0.5] = E[X|Y = 0] = 0$. Similarly,

$$\begin{aligned} E[X|Y = 1] &= E[X||X| > 0.5] \\ &= E[X|X > 0.5]P(X > 0.5||X| > 0.5) + E[X|X < -0.5]P(X < -0.5||X| > 0.5) \\ &= (0.75 - 0.75)\frac{1}{2} \\ &= 0 \end{aligned}$$

In the third line we used the fact that $E[X|X > 0.5] = E[A]$ for $A \sim U[0.5, 1]$ and the fact that $E[X|X < -0.5] = E[B]$ for $B \sim U[-1, -0.5]$

Thus, if

$$
\begin{aligned}
E[X|Y=1] = E[X|Y=0] &= 0 \\
&\Rightarrow E[X|Y] = 0 \\
&\Rightarrow E[X|Y] = E[X]
\end{aligned}
$$

so X is mean independent of Y. However, independence will not hold (which we already know because Y is not mean independent of X):

$$
\begin{aligned}
P(X > 0.5, Y = 1) &= 0 \\
P(X > 0.5)P(Y = 1) &= \frac{1}{4} * \frac{1}{2} \neq 0 \\
\Rightarrow P(X > 0.5, Y = 1) &\neq P(X > 0.5)P(Y = 1)
\end{aligned}
$$

Thus, the general definition of independence will not hold (again one counterexample is sufficient).

Problem 7

Perform the following using R or another statistical software of your choice (provided that you have cleared any alternative option with the TA).

(a) Generate 1,000 draws from a standard normal distribution ($N(0, 1)$) and plot the simulated data in a histogram.

(b) Generate 1,000 draws from a Uniform$[-1, 1]$ distribution and plot the simulated data in a histogram.