Math 3081 (Probability and Statistics)

Lecture #8 of 27 $\sim$ July 19th, 2021

Discrete Random Variables (Part 3)

- Properties of Independence
- Covariance and Correlation
- Continuous Random Variables

This material represents §2.1.5-2.2.2 from the course notes and problems 6-10 on WeBWorK 3.

Recall, I

We defined expected value, variance, and standard deviation:

Definition

If X is a discrete random variable, the expected value of X, written $E(X)$, is the sum $E(X) = \sum_{s_i \in S} P(s_i)X(s_i)$ over all outcomes s_i in the sample space S.

Definition

If X is a discrete random variable whose expected value $E(X) = \mu$ exists and is finite, we define the variance of X to be $\mathrm{var}(X) = E[(X - \mu)^2] = E(X^2) - E(X)^2$. The standard deviation is the square root of the variance: $\sigma(X) = \sqrt{\mathrm{var}(X)}$.

Recall, II

We established a few properties of expected value and variance:

Proposition (Linearity and Additivity of Expected Value)

If X and Y are discrete random variables defined on the same sample space whose expected values exist, and a and b are any real numbers, then $E(aX + b) = a \cdot E(X) + b$ and $E(X + Y) = E(X) + E(Y)$.

Proposition (Properties of Variance)

If X is a discrete random variable and a and b are any real numbers, then $\mathrm{var}(aX + b) = a^2 \mathrm{var}(X)$ and $\sigma(aX + b) = |a|\, \sigma(X)$.

Recall, III

We also introduced joint distributions:

Definition

If $X_1, X_2, \ldots, X_n$ are discrete random variables on the sample space S, then the function $p_{X_1,X_2,\ldots,X_n}$ defined on ordered n-tuples of events $(a_1, \ldots, a_n) \in S$ such that $p_{X_1,X_2,\ldots,X_n}(a_1, a_2, \ldots, a_n) = P(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n)$ is called the joint probability density function of $X_1, X_2, \ldots, X_n$.

With 2 variables, we usually display the joint distribution as a table.

Proposition (Marginal Densities)

If $p_{X,Y}(a, b)$ is the joint probability density function for the discrete random variables X and Y, then for any a and b we may compute the marginal probability density functions for X and Y as $p_X(a) = \sum_y p_{X,Y}(a, y)$ and $p_Y(b) = \sum_x p_{X,Y}(x, b)$.

Recall, IV

We also defined independence of two random variables:

Definition

Two discrete random variables X and Y with respective probability density functions $p_X(x)$ and $p_Y(y)$ are <u>independent</u> if their joint distribution $p_{X,Y}(x, y)$ satisfies $p_{X,Y}(a, b) = p_X(a) \cdot p_Y(b)$ for all real numbers a and b.

This extends to more variables as follows:

Definition

We say that the discrete random variables $X_1, X_2, \dots, X_n$ are <u>collectively independent</u> if the joint distribution $p_{X_1,X_2,\dots,X_n}(a_1, a_2, \dots, a_n) = p_{X_1}(a_1) \cdot p_{X_2}(a_2) \cdot \dots \cdot p_{X_n}(a_n)$ for all real numbers $a_1, a_2, \dots, a_n$.

Variance and Independence, I

Under the assumption of independence, we can give a few additional algebraic properties of expected value and variance:

Proposition (Variance and Independence)

If X and Y are independent discrete random variables whose expected values exist, then $E(XY) = E(X) \cdot E(Y)$, and $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Variance and Independence, I

Under the assumption of independence, we can give a few additional algebraic properties of expected value and variance:

Proposition (Variance and Independence)

If X and Y are independent discrete random variables whose expected values exist, then $E(XY) = E(X) \cdot E(Y)$, and $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

- Note that we do require the hypothesis that X and Y be independent in order for the variance to be additive.
- The result can be false for non-independent random variables: an easy counterexample occurs for $X = Y$, in which case $\mathrm{var}(X + Y) = \mathrm{var}(2X) = 4\mathrm{var}(X)$ which is not equal to $\mathrm{var}(X) + \mathrm{var}(Y) = 2\mathrm{var}(X)$.
- Do note, however, that expected value is always additive, whether or not the variables are independent.

Variance and Independence, II

1. If X and Y are independent, then $E(XY) = E(X) \cdot E(Y)$.

Proof:

- Suppose X takes the values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$ and Y takes the values $y_1, y_2, \ldots$ with probabilities $q_1, q_2, \ldots$.
- By the assumption of independence, X takes the value x_i and Y takes the value y_j, so that XY takes the value $x_i y_j$, with probability $p_i q_j$.
- We therefore have $E(XY) = \sum_{i,j} p_i q_j x_i y_j = \left[\sum_i p_i x_i\right] \cdot \left[\sum_j q_j y_j\right] = E(X) \cdot E(Y)$, as claimed.

Variance and Independence, III

2. If X and Y are independent, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Proof:

- We just showed that $E(XY) = E(X)E(Y)$.

Variance and Independence, III

2. If X and Y are independent, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Proof:

- We just showed that $E(XY) = E(X)E(Y)$.
- Then, by expanding, we see

$$\begin{aligned}
\mathrm{var}(X+Y) &= E[(X+Y)^2] - [E(X+Y)]^2 \\
&= E(X^2 + 2XY + Y^2) - [E(X) + E(Y)]^2 \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) \\
&\quad -[E(X)^2 + 2E(X)E(Y) + E(Y)^2] \\
&= [E(X^2) - E(X)^2] + [E(Y^2) - E(Y)^2] \\
&= \mathrm{var}(X) + \mathrm{var}(Y)
\end{aligned}$$

as claimed.

Variance and Independence, IV

Using these properties of variance, we can calculate the variance and standard deviation of a binomially-distributed random variable:

Corollary (Binomial Variance)

Let X be the binomially-distributed random variable representing the total number of successes obtained by performing n independent Bernoulli trials each of which has a success probability p. Then $E(X) = np$, $\mathrm{var}(X) = np(1-p)$, and $\sigma(X) = \sqrt{np(1-p)}$.

Variance and Independence, IV

Using these properties of variance, we can calculate the variance and standard deviation of a binomially-distributed random variable:

Corollary (Binomial Variance)

Let X be the binomially-distributed random variable representing the total number of successes obtained by performing n independent Bernoulli trials each of which has a success probability p. Then $E(X) = np$, $\mathrm{var}(X) = np(1-p)$, and $\sigma(X) = \sqrt{np(1-p)}$.

Example: For a binomial distribution with $n = 100$ and $p = 1/5$, the expected value is $np = 20$, the variance is $np(1-p) = 16$, and the standard deviation is $\sqrt{np(1-p)} = 4$.

Variance and Independence, V

Proof:

- Since X is obtained by summing over the individual trials, we can write $X = X_1 + X_2 + \cdots + X_n$ where X_i is the random variable representing success on the ith trial.
- Then $E(X_i) = (1-p) \cdot 0 + p \cdot 1 = p$, and so $E(X) = E(X_1) + \cdots + E(X_n) = np$ as we found earlier.
- Also, $E(X_i^2) = (1-p) \cdot 0^2 + p \cdot 1^2 = p$, so $\mathrm{var}(X_i) = E(X_i^2) - E(X_i)^2 = p(1-p)$.
- Each of the X_i is a single Bernoulli trial and they are all collectively independent by assumption.
- That means the variance is additive here, so $\mathrm{var}(X) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n) = np(1-p)$ and $\sigma(X) = \sqrt{\mathrm{var}(X)} = \sqrt{np(1-p)}$, as claimed.

Variance and Independence, VI

Example: An unfair coin with a probability $2/3$ of landing heads is flipped 450 times. Find the expected number and the standard deviation in the number of tails obtained.

Variance and Independence, VI

<u>Example</u>: An unfair coin with a probability $2/3$ of landing heads is flipped 450 times. Find the expected number and the standard deviation in the number of tails obtained.

- Each individual flip can be thought of as a Bernoulli trial, with success corresponding to obtaining tails with probability $p = 1/3$, with a total of $n = 450$ trials.
- Thus, from our results on the binomial distribution, the expected number of tails is $np = 450 \cdot 1/3 = \boxed{150}$ and the standard deviation is $\sqrt{np(1-p)} = \sqrt{450 \cdot 1/3 \cdot 2/3} = \boxed{10}$.

Variance and Independence, VII

Example: A car dealer has a probability 0.36 of selling a car to any individual customer, independently. If 25 customers patronize the dealership, determine the expected number and the standard deviation in the total number of cars sold.

Variance and Independence, VII

Example: A car dealer has a probability 0.36 of selling a car to any individual customer, independently. If 25 customers patronize the dealership, determine the expected number and the standard deviation in the total number of cars sold.

- Each individual customer can be thought of as a Bernoulli trial, with success corresponding to selling a car with probability $p = 0.36$, with a total of $n = 25$ trials.
- Thus, from our results on the binomial distribution, the expected number of cars sold is $np = 25 \cdot 0.36 = \boxed{9}$ and the standard deviation is $\sqrt{np(1-p)} = \sqrt{25 \cdot 0.36 \cdot 0.64} = \boxed{2.4}$.

Variance and Independence, VIII

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

1. The expected number of goals allowed per game.
2. The standard deviation of the number of goals allowed per game.
3. The probability of getting a shutout (saving every shot).
4. The probability of allowing at least 3 goals.
5. The most likely number of goals to allow.

Variance and Independence, VIII

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

1. The expected number of goals allowed per game.
2. The standard deviation of the number of goals allowed per game.
3. The probability of getting a shutout (saving every shot).
4. The probability of allowing at least 3 goals.
5. The most likely number of goals to allow.

- The total number of goals will be binomially distributed with parameters $n = 31$ and $p = 1 - 0.911 = 0.089$.
- Thus, the probability of allowing k goals is $\binom{31}{k}(0.089)^k(0.911)^{31-k}$.

Variance and Independence, IX

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

1. The expected number of goals allowed per game.

Variance and Independence, IX

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

1. The expected number of goals allowed per game.

- The expected number of goals is $np \approx 2.759$.

2. The standard deviation of the number of goals allowed per game.

Variance and Independence, IX

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

1. The expected number of goals allowed per game.
   - The expected number of goals is $np \approx 2.759$.
2. The standard deviation of the number of goals allowed per game.
   - The standard deviation in the number of goals is $\sqrt{np(1-p)} \approx 1.585$.
3. The probability of getting a shutout (saving every shot).

Variance and Independence, IX

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

1. The expected number of goals allowed per game.
   - The expected number of goals is $np \approx 2.759$.
2. The standard deviation of the number of goals allowed per game.
   - The standard deviation in the number of goals is $\sqrt{np(1-p)} \approx 1.585$.
3. The probability of getting a shutout (saving every shot).
   - From the formula, this probability is $(0.911)^{31} \approx 5.56\%$.
   - Remark: In the last two seasons, the actual shutout percentages were 6.10% and 5.08%.

Variance and Independence, X

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

4. The probability of allowing at least 3 goals.

Variance and Independence, X

<u>Example</u>: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

4. The probability of allowing at least 3 goals.

- The probability of allowing at least 3 goals requires summing many terms. We can instead find the probability of the complement, allowing at most 2 goals.
- That probability is $\binom{31}{0}(0.089)^0(0.911)^{31} + \binom{31}{1}(0.089)^1(0.911)^{30} + \binom{31}{2}(0.089)^2(0.911)^{29} \approx 0.4707$.
- Thus, the desired probability is $1 - 0.4707 \approx 52.93\%$.

Variance and Independence, XI

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

5. The most likely number of goals to allow.

Variance and Independence, XI

Example: During an average NHL game, a team's goaltender will face 31 shots, and will independently make a save on each shot with probability 0.911. Each non-saved shot results in a goal. Find the following:

5. The most likely number of goals to allow.

- It seems reasonable that the most likely number should probably be near the expected value.
- Evaluating the probabilities for 0, 1, 2, 3, 4, 5 goals yields 5.56%, 16.84%, 24.68%, 23.30%, 15.94%, 8.41%.
- So the most likely number of goals is 2 goals, followed closely by 3 goals.
- Notice that these two values sandwich the expected value, which is 2.759. (This is in fact the case for all binomial distributions.)

Covariance, I

Definition

If X and Y are random variables whose expected values exist and are μ_X and μ_Y respectively, then the covariance of X and Y is defined as $\operatorname{cov}(X, Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)]$.

- The covariance measures how well a change in the value of X (relative to its average) correlates with a change in the value of Y (relative to its average).
- If the covariance is large and positive, then when X increases, Y will tend also to increase, and inversely when X decreases, Y will tend also to decrease.
- The inverse occurs for a large negative covariance.
- When the covariance is near zero, then a change in the value of X does not tend to correspond to any particular type of change in the value of Y.

Covariance, II

Example: Find the covariance of the random variables X and Y with joint distribution below.

$X \setminus Y$	0	10	Sum
0	0.4	0.1	0.5
10	0.2	0.3	0.5
Sum	0.6	0.4	

Covariance, II

<u>Example</u>: Find the covariance of the random variables X and Y with joint distribution below.

$X \setminus Y$	0	10	Sum
0	0.4	0.1	0.5
10	0.2	0.3	0.5
Sum	0.6	0.4	

- We can compute $\mu_X = 5$, $\mu_Y = 4$.
- Then $\text{cov}(X, Y) =$ $0.4 \cdot (-5) \cdot (-4) + 0.1 \cdot (-5) \cdot (6) + 0.2 \cdot (5) \cdot (-4) + 0.3 \cdot (5) \cdot (6) = 10$.
- We can see that when X is 0, Y is more likely to be 0 than 10, and when X is 10, Y is more likely to be 10 than 0.

Covariance, III

Even in the simple example we just did, computing the covariance was a bit complicated. Here is an easier formula for practical calculation:

Proposition (Covariance Formula)

For any discrete random variables X and Y whose expected values exist, we have $\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y)$.

Covariance, III

Even in the simple example we just did, computing the covariance was a bit complicated. Here is an easier formula for practical calculation:

Proposition (Covariance Formula)

For any discrete random variables X and Y whose expected values exist, we have $\operatorname{cov}(X, Y) = E(XY) - E(X)E(Y)$.

Proof:

- If the expected values of X and Y are μ_X and μ_Y respectively, then by linearity of expectation we have
  $\operatorname{cov}(X, Y) = E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y)$
  $= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y$
  $= E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y)$, as claimed.

Covariance, IV

We have an immediate corollary of the covariance formula:

Corollary (Covariance and Independence)

If X and Y are independent, then $\mathrm{cov}(X, Y) = 0$.

Proof:

- If X and Y are independent, then $E(XY) = E(X)E(Y)$, so $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) = 0$.

We remark here that the converse of this statement is *not* true: if the covariance is zero, it does not imply that X and Y are independent.

Covariance, V

Example: Find the covariance of the random variables X and Y with joint distribution below.

$X \ \backslash \ Y$	0	5	10	Sum
0	0.3	0.2	0	0.5
10	0.4	0	0.1	0.5
Sum	0.7	0.2	0.1	

Covariance, V

<u>Example</u>: Find the covariance of the random variables X and Y with joint distribution below.

$X \setminus Y$	0	5	10	Sum
0	0.3	0.2	0	0.5
10	0.4	0	0.1	0.5
Sum	0.7	0.2	0.1	

- We compute $E(X) = 0.5 \cdot 0 + 0.5 \cdot 10 = 5$, $E(Y) = 0.7 \cdot 0 + 0.2 \cdot 5 + 0.1 \cdot 10 = 2$, and $E(XY) = 0.3 \cdot 0 + 0.2 \cdot 0 + 0 \cdot 0 + 0.4 \cdot 0 + 0 \cdot 50 + 0.1 \cdot 100 = 10$.
- Therefore, we see $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) = 10 - 5 \cdot 2 = \boxed{0}$.

Note here that X and Y have covariance zero, but are not independent.

Covariance, VI

We have various algebraic properties involving the covariance:

Proposition (Properties of Covariance)

If X and Y are discrete random variables whose expected values exist, then for any a and b we have the following:

1. $\mathrm{cov}(X, X) = \mathrm{var}(X)$.
2. $\mathrm{cov}(Y, X) = \mathrm{cov}(X, Y)$.
3. $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$.
4. $\mathrm{cov}(aX + b, cY + d) = ac \cdot \mathrm{cov}(X, Y)$.
5. $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)$.

Remark (for linear algebra students): Properties 1-3 imply that the covariance is an inner product on the vector space of discrete random variables on a fixed sample space S. As such, many tools of linear algebra (e.g., least-squares estimation) can be used here.

Covariance, VII

Proofs:

1. $\text{cov}(X, X) = \text{var}(X)$.

- By definitions, we have $\text{cov}(X, X) = E[(X - \mu_X)^2] = \text{var}(X)$.

2. $\text{cov}(Y, X) = \text{cov}(X, Y)$.

- Since $(X - \mu_X)(Y - \mu_Y) = (Y - \mu_Y)(X - \mu_X)$, the expected values of the two sides are also equal.

3. $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$.

- Note that $E(X + Y) = E(X) + E(Y) = \mu_X + \mu_Y$.
- Then, since $(X + Y - \mu_X - \mu_Y)(Z - \mu_Z) = (X - \mu_X)(Z - \mu_Z) + (Y - \mu_Y)(Z - \mu_Z)$, the expected values of the two sides are also equal.

Covariance, VIII

Proofs:

4. $\text{cov}(aX + b, cY + d) = ac \cdot \text{cov}(X, Y)$.
   - Notice $E(aX + b) = a\mu_X + b$ and $E(cY + d) = c\mu_Y + d$, so $\text{cov}(aX+b, cY+d) = E[(aX+b-a\mu_X-b)(cY+d-c\mu_Y-d)]$ $= ac \cdot E[(X - \mu_X)(Y - \mu_Y)] = ac \cdot \text{cov}(X, Y)$.
5. $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$.
   - Note that $E(X + Y) = \mu_X + \mu_Y$, so $\text{var}(X + Y)$ $= E[(X + Y - \mu_X - \mu_Y)^2] = E[(X - \mu_X)^2 + 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2] = \text{var}(X) + 2\text{cov}(X, Y) + \text{var}(Y)$.

Correlation, I

From the properties we identified, we can see that covariance scales linearly with both of the random variables X and Y.

In some situations, we prefer to have a "normalized" measure of covariance, which we can obtain by dividing the covariance by the product of the standard deviations:

Definition

If X and Y are discrete random variables whose variances exist and are nonzero, the (Pearson) correlation between X and Y is defined as $\operatorname{corr}(X, Y) = \dfrac{\operatorname{cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

- It is easy to see that the correlation (unlike the covariance) remains unchanged upon scaling and translating the variables: $\operatorname{corr}(aX + b, cY + d) = \operatorname{corr}(X, Y)$.

Correlation, II

For any X and Y, we always have $-1 \leq \text{corr}(X, Y) \leq 1$.

- A correlation near 1 indicates that the variables tend to increase linearly together and decrease linearly together (which agrees with the intuitive notion of two variables being strongly positively correlated).
- A correlation near -1 indicates that the variables tend to increase linearly as the other decreases linearly (which agrees with the intuitive notion of two variables being strongly negatively correlated).
- A correlation of zero is the same as a covariance of zero: it indicates that an increase in one variable does not tend to cause an increase or decrease in the other.
- Do note, however (as we saw before) that a correlation of zero is not equivalent to the variables being independent!

Correlation, III

In other contexts, the correlation is also known as the linear regression correlation coefficient, since it represents the closeness by which a linear function can describe the relationship between X and Y.

- A correlation coefficient near 1 indicates that there is a linear function with a positive slope that models the relationship closely, while a correlation coefficient near -1 indicates that there is a linear function with a negative slope that models the relationship closely.
- A correlation coefficient near 0 indicates that there is no linear function that models the relationship closely: but of course, this need not mean that the variables are unrelated, merely that any relationship is not linear.

Correlation, IV

Example: A fair coin is flipped 3 times. If X is the total number of heads in the first two flips and Y is the total number of heads in the last two flips, find the covariance and correlation between X and Y.

Correlation, IV

<u>Example</u>: A fair coin is flipped 3 times. If X is the total number of heads in the first two flips and Y is the total number of heads in the last two flips, find the covariance and correlation between X and Y.

- Last lecture, we found the joint distribution for X and Y:

$X \setminus Y$	0	1	2
0	1/8	1/8	0
1	1/8	2/8	1/8
2	0	1/8	1/8

- We can use this table to compute the covariance and correlation.

Correlation, V

Example: Find the covariance and correlation between X and Y:

$X \setminus Y$	0	1	2
0	1/8	1/8	0
1	1/8	2/8	1/8
2	0	1/8	1/8

Correlation, V

Example: Find the covariance and correlation between X and Y:

$X \setminus Y$	0	1	2
0	1/8	1/8	0
1	1/8	2/8	1/8
2	0	1/8	1/8

- We can compute $E(X) = E(Y) = 1$, $\sigma(X) = \sigma(Y) = \sqrt{1/2}$, and $E(XY) = \frac{3}{8} \cdot 0 + \frac{2}{8} \cdot 1 + \frac{2}{8} \cdot 2 + \frac{1}{8} \cdot 4 = \frac{5}{4}$, and so $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) = 1/4$.
- Then $\mathrm{corr}(X, Y) = \dfrac{\mathrm{cov}(X, Y)}{\sigma(X)\sigma(Y)} = 1/2$.
- We can see that there is a positive correlation of moderate size between X and Y, which is intuitively reasonable because X and Y each count the number of heads from one independent coin flip and one shared coin flip.

Motivation For Continuous Random Variables, I

We have just examined joint distributions, independence, covariance, and correlation for discrete random variables. This marks the end of the major items we will discuss about discrete random variables for the moment.

We now start the second major portion of this chapter by starting our discussion of continuous random variables, which essentially involves doing everything we just did in the last 2.5 lectures over again, except slightly differently, and with calculus!

Motivation For Continuous Random Variables, II

Our new object of study is the class of random variables whose underlying sample space is the entire real line.

- Because there are uncountably many possible outcomes, to evaluate probabilities of events we cannot simply sum over the outcomes that make them up.
- Instead, we must use the continuous analogue of summation, namely, integration.
- All of the results we will discuss are very similar to the corresponding ones for discrete random variables, with the only added complexity[1] being the requirement to evaluate integrals.

[1] This is almost a true statement, in the sense that I think it is true but you will probably think it is false.

Continuous Probability Density Functions, I

Definition

A continuous probability density function is a piecewise-continuous, nonnegative real-valued function $p(x)$ such that $\int_{-\infty}^{\infty} p(x)\,dx = 1$.

- Note that the integral $\int_{-\infty}^{\infty} p(x)\,dx$ is in general improper. (Usually this won't be an issue for us.)

Here is the purpose of this definition:

- In the discrete case, we compute probabilities of events by adding up values of the probability density function over the outcomes making up that event.
- In the continuous case, we will compute probabilities of events by integrating the value of the probability density function over the outcomes making up that event.

Continuous Probability Density Functions, II

Examples:

- The function $p(x) = \begin{cases} 1/4 & \text{for } 0 \leq x \leq 4 \\ 0 & \text{for other } x \end{cases}$ is a continuous probability density function, since the two components of $p(x)$ are both continuous and nonnegative, and $\int_{-\infty}^{\infty} p(x)\,dx = \int_0^4 \frac{1}{4}\,dx = \frac{1}{4}x\Big|_{x=0}^{4} = 1.$

Continuous Probability Density Functions, II

Examples:

- The function $p(x) = \begin{cases} 1/4 & \text{for } 0 \leq x \leq 4 \\ 0 & \text{for other } x \end{cases}$ is a continuous probability density function, since the two components of $p(x)$ are both continuous and nonnegative, and $\int_{-\infty}^{\infty} p(x)\,dx = \int_0^4 \frac{1}{4}\,dx = \frac{1}{4}x\Big|_{x=0}^{4} = 1.$

- The function $q(x) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1 \\ 0 & \text{for other } x \end{cases}$ is a continuous probability density function, since the two components of $q(x)$ are both continuous and nonnegative, and $\int_{-\infty}^{\infty} q(x)\,dx = \int_0^1 2x\,dx = x^2\big|_{x=0}^{1} = 1.$

Continuous Probability Density Functions, III

Examples (continued):

- The function $e(x) = \begin{cases} e^{-x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$ is a continuous probability density function, since the two components of $e(x)$ are both continuous and nonnegative, and $\int_{-\infty}^{\infty} e(x)\,dx = \int_0^{\infty} e^{-x}\,dx = -e^{-x}\big|_{x=0}^{\infty} = 1$.

Continuous Probability Density Functions, III

Examples (continued):

- The function $e(x) = \begin{cases} e^{-x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$ is a continuous probability density function, since the two components of $e(x)$ are both continuous and nonnegative, and $\int_{-\infty}^{\infty} e(x)\,dx = \int_0^{\infty} e^{-x}\,dx = -e^{-x}\big|_{x=0}^{\infty} = 1$.

- The function $f(x) = \dfrac{1}{\pi(1+x^2)}$ is a continuous probability density function since $f(x)$ is continuous, nonnegative, and $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \dfrac{1}{\pi(1+x^2)}\,dx = \dfrac{\tan^{-1}(x)}{\pi}\bigg|_{x=-\infty}^{\infty} = 1$.

Continuous Random Variables, I

Definition

We say that X is a <u>continuous random variable</u> if there exists a continuous probability density function $p(x)$ such that for any interval I on the real line, we have $P(X \in I) = \int_I p(x)\,dx$.

- In other words, probabilities for continuous random variables are computed via integrating the probability density function on the appropriate interval.
- Note that if X is a continuous random variable, then (no matter what the probability density function is), the value of $P(X = a)$ is zero for any value of a, since $P(X = a) = \int_a^a p(x)\,dx = 0$. This means that the probability that X will attain any specific value a is equal to zero.
- One may verify axioms [P1]-[P3] of a probability distribution will all hold for such a random variable.

Continuous Random Variables, II

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \leq x \leq 4 \\ 0 & \text{for other } x \end{cases}$, do:

1. Verify that p is a probability density function.
2. Find $P(1 \leq X \leq 3)$.
3. Find $P(X \leq 2)$.
4. Find $P(X \geq 5)$.
5. Find $P(-2 \leq X \leq 3)$.
6. Find $P(X = 2)$.

Continuous Random Variables, II

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

1. Verify that p is a probability density function.
2. Find $P(1 \le X \le 3)$.
3. Find $P(X \le 2)$.
4. Find $P(X \ge 5)$.
5. Find $P(-2 \le X \le 3)$.
6. Find $P(X = 2)$.

- For each of these, we need to evaluate the appropriate integral.
- When we set up the integrals, we must remember to break the range of integration up (if needed) so that we are integrating the correct component of $p(x)$ on the correct interval.

Continuous Random Variables, III

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

1. Verify that p is a probability density function.

Continuous Random Variables, III

<u>Example</u>: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

1. Verify that p is a probability density function.

- Here, we need to check that p integrates to 1.
- To do this, we compute $\int_{-\infty}^{\infty} p(x)\,dx =$ $\int_{-\infty}^{0} 0\,dx + \int_{0}^{4} \frac{x}{8}\,dx + \int_{4}^{\infty} 0\,dx = 0 + \frac{1}{16}x^2\Big|_{x=0}^{4} + 0 = 1$, as required.

Above, we could have simply discarded the parts of the integral where p is zero, since they will never contribute anything. We will do this from now on.

Continuous Random Variables, IV

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

2. Find $P(1 \le X \le 3)$.

Continuous Random Variables, IV

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

2. Find $P(1 \le X \le 3)$.

- We have
$$P(1 \le X \le 3) = \int_1^3 p(x)\,dx = \int_1^3 \frac{x}{8}\,dx = \frac{1}{16}x^2 \Big|_{x=1}^{3} = \frac{1}{2}.$$

3. Find $P(X \le 2)$.

Continuous Random Variables, IV

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \leq x \leq 4 \\ 0 & \text{for other } x \end{cases}$, do:

2. Find $P(1 \leq X \leq 3)$.
   - We have
     $$P(1 \leq X \leq 3) = \int_1^3 p(x)\,dx = \int_1^3 \frac{x}{8}\,dx = \frac{1}{16}x^2\Big|_{x=1}^{3} = \frac{1}{2}.$$
3. Find $P(X \leq 2)$.
   - We have
     $$P(X \leq 2) = \int_{-\infty}^2 p(x)\,dx = \int_0^2 \frac{x}{8}\,dx = \frac{1}{16}x^2\Big|_{x=0}^{2} = \frac{1}{4}.$$
   - Note that we discarded the part of the integral from $-\infty$ to 0 because p is zero there.

Continuous Random Variables, V

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

4. Find $P(X \ge 5)$.

Continuous Random Variables, V

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \leq x \leq 4 \\ 0 & \text{for other } x \end{cases}$, do:

4. Find $P(X \geq 5)$.
   - Here, $P(X \geq 5) = \int_5^\infty p(x)\,dx = \int_5^\infty 0\,dx = 0$.
5. Find $P(-2 \leq X \leq 3)$.

Continuous Random Variables, V

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

4. Find $P(X \ge 5)$.
   - Here, $P(X \ge 5) = \int_5^\infty p(x)\,dx = \int_5^\infty 0\,dx = 0$.
5. Find $P(-2 \le X \le 3)$.
   - Here,

$$P(-2 \le X \le 3) = \int_{-2}^{3} p(x)\,dx = \int_0^3 \frac{x}{8}\,dx = \frac{1}{16}x^2\Big|_{x=0}^{3} = \frac{9}{16}.$$

6. Find $P(X = 2)$.

Continuous Random Variables, V

Example: If X is the continuous random variable whose probability density function is $p(x) = \begin{cases} x/8 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, do:

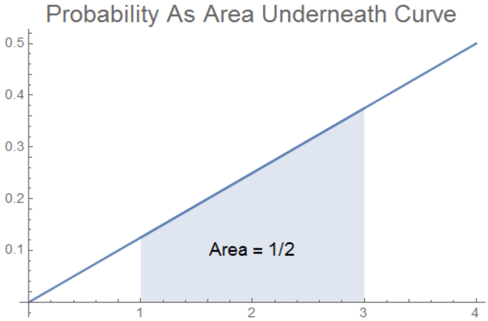
4. Find $P(X \ge 5)$.
   - Here, $P(X \ge 5) = \int_5^\infty p(x)\,dx = \int_5^\infty 0\,dx = 0$.
5. Find $P(-2 \le X \le 3)$.
   - Here,

     $$P(-2 \le X \le 3) = \int_{-2}^3 p(x)\,dx = \int_0^3 \frac{x}{8}\,dx = \frac{1}{16}x^2\Big|_{x=0}^3 = \frac{9}{16}.$$
6. Find $P(X = 2)$.
   - This is $P(X = 2) = \int_2^2 p(x)\,dx = 0$, since the limits of integration are equal.

Continuous Random Variables, VI

Since integrals represent areas, we can also view all of these calculations as computing the area under the graph of $y = p(x)$.

- The area under the curve from $x = a$ to $x = b$ then represents the probability $P(a < x < b)$.
- Here is a picture of the area corresponding to $P(1 \leq X \leq 3)$:



Cumulative Distribution Functions, I

Definition

If X is a continuous random variable with probability density function $p(x)$, its <u>cumulative distribution function</u> (<u>cdf</u>) $c(x)$ is defined as $c(x) = \int_{-\infty}^{x} p(t)\,dt$ for each real value of x.

- The cumulative distribution function $c(x)$ measures the total probability that the continuous random variable X takes a value $\leq x$: thus, $P(X \leq a) = c(a)$.

Here are some properties that follow from the definition:

- $P(X \geq a) = 1 - c(a)$ for all a, since $\int_{a}^{\infty} p(t)\,dt = \int_{-\infty}^{\infty} p(t)\,dt - \int_{-\infty}^{a} p(t)\,dt$.
- $P(a \leq X \leq b) = c(b) - c(a)$ for every a and b, since $\int_{a}^{b} p(t)\,dt = \int_{-\infty}^{b} p(t)\,dt - \int_{-\infty}^{a} p(t)\,dt$.

Cumulative Distribution Functions, II

Here are some more properties:

- By the fundamental theorem of calculus, we have $c'(x) = p(x)$ for every x.
- Thus, we may freely convert back and forth between the probability density function and the cumulative distribution function via differentiation and integration.
- Also, since $p(x)$ is nonnegative, if $a \leq b$ then $c(a) \leq c(b)$: this means that the cumulative function is increasing (technically, nondecreasing).
- Since $\int_{-\infty}^{\infty} p(x)\,dx = 1$, we also see $\lim_{x\to\infty} c(x) = 1$ and $\lim_{x\to-\infty} c(x) = 0$.

Cumulative Distribution Functions, III

Examples:

- For the random variable with probability density function $p(x) = \begin{cases} 1/4 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, the cumulative distribution function is $c(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x/4 & \text{for } 0 \le x \le 4 \\ 1 & \text{for } x \ge 4 \end{cases}$.

Cumulative Distribution Functions, III

Examples:

- For the random variable with probability density function $p(x) = \begin{cases} 1/4 & \text{for } 0 \le x \le 4 \\ 0 & \text{for other } x \end{cases}$, the cumulative distribution function is $c(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x/4 & \text{for } 0 \le x \le 4 \\ 1 & \text{for } x \ge 4 \end{cases}$.
- For the random variable with probability density function $q(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1/x^2 & \text{for } x \ge 1 \end{cases}$, the cumulative distribution function is $c(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1 - 1/x & \text{for } x \ge 1 \end{cases}$.

Cumulative Distribution Functions, IV

We could also have defined the cumulative distribution function for a discrete random variable.

- The typical definition is $c(x) = \sum_{n \leq x} p(n)$, which is the probability $P(X \leq x)$.
- However, for various reasons in certain cases one may prefer to sum only over all values less than x, rather than less than or equal to x, and it can be easy to mix up these situations.
- For example, $P(X \geq x) = 1 - P(X < x) \neq 1 - c(x)$ in the discrete case, since in fact $1 - c(x) = 1 - P(X \leq x)$.
- This is only an issue with discrete random variables: when X is a continuous random variable, since the probability of obtaining exactly the value x is always zero, we do have $P(X \leq x) = P(X < x)$.
- For this reason, we will work with cumulative distribution functions only in the context of continuous random variables.

Summary

We discussed independence of discrete random variables.

We introduced covariance and correlation of discrete random variables and discussed some of their properties.

We introduced continuous random variables and their associated PDFs and CDFs.

Next lecture: Continuous random variables (part 1.5)