Math 3081 (Probability and Statistics)

Lecture #7 of 27 $\sim$ July 15th, 2021

Discrete Random Variables (Part 2)

- Expected Value (more)
- Variance and Standard Deviation
- Joint Distributions and Independence

This material represents §2.1.2-2.1.4 from the course notes, and problems 17-20 on WeBWorK 2 and problems 1-8 on WeBWorK 3.

Recall, I

Recall our definition of a random variable from last class:

Definition

A random variable is a (real-valued) function defined on the outcomes in a sample space. A discrete random variable is a random variable is one whose underlying sample space is finite or countably infinite.

The probability density function packages all of the information about a random variable:

Definition

If X is a random variable on the sample space S, then the function p_X such that $p_X(E) = P(X \in E)$ for any event E is called the probability density function (pdf) of X.

Recall, II

We also gave a definition of the "average value" for an arbitrary discrete random variable:

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.

The expected value $E(X)$ is the average of the values that X takes on the outcomes in the sample space, weighted by the probability of each outcome. A common application of expected value is to calculate the expected winnings from a game of chance.

More Expected Value, II

Example: In one version of a "Pick 3" lottery, a single entry ticket costs \$1. In this lottery, 3 single digits are drawn at random, and a ticket must match all 3 digits in the correct order to win the \$500 prize. What is the expected value of one ticket for this lottery?

More Expected Value, II

<u>Example</u>: In one version of a "Pick 3" lottery, a single entry ticket costs \$1. In this lottery, 3 single digits are drawn at random, and a ticket must match all 3 digits in the correct order to win the \$500 prize. What is the expected value of one ticket for this lottery?

- From the description, we can see that there is a $1/1000$ probability of winning the prize and a $999/1000$ probability of winning nothing.
- Since winning the prize nets a total of \$499 (the prize minus the \$1 entry fee), and winning nothing nets a total of $-\$1$, the expected value of the random variable giving the net winnings is equal to $\frac{1}{1000}(\$499) + \frac{999}{1000}(-\$1) = -\$0.50$.
- The expected value of $-\$0.50$, in this case, indicates that if one plays this lottery many times, on average one should expect to lose 50 cents on every ticket.

More Expected Value, II

Example: In one version of the game "Chuck-a-luck", three standard 6-sided dice are rolled. Prizes for a bet of $1 are awarded as follows: $66 for a roll of three 6s, $5 for a roll of two 6s, $1 for one 6, and $0 for any other roll.

1. If X represents the net win (or loss) from playing the game once, find the probability distribution for X.
2. Find the expected value of X.

More Expected Value, II

Example: In one version of the game "Chuck-a-luck", three standard 6-sided dice are rolled. Prizes for a bet of \$1 are awarded as follows: \$66 for a roll of three 6s, \$5 for a roll of two 6s, \$1 for one 6, and \$0 for any other roll.

1. If X represents the net win (or loss) from playing the game once, find the probability distribution for X.
2. Find the expected value of X.

- To find the probability distribution, we calculate the probabilities of the various outcomes and tabulate the winnings. Notice that these probabilities will be binomially distributed with $n = 3$ and $p = 1/6$.
- We can then compute the expected value directly from the probability distribution.

More Expected Value, III

Example: In one version of the game "Chuck-a-luck", three standard 6-sided dice are rolled. Prizes for a bet of \$1 are awarded as follows: \$66 for a roll of three 6s, \$5 for a roll of two 6s, \$1 for one 6, and \$0 for any other roll.

1. If X represents the net win (or loss) from playing the game once, find the probability distribution for X.

More Expected Value, III

Example: In one version of the game "Chuck-a-luck", three standard 6-sided dice are rolled. Prizes for a bet of \$1 are awarded as follows: \$66 for a roll of three 6s, \$5 for a roll of two 6s, \$1 for one 6, and \$0 for any other roll.

1. If X represents the net win (or loss) from playing the game once, find the probability distribution for X.

- We have $P(\text{three 6s}) = (1/6)^3 = 1/216$. Net winnings: \$65.
- We have $P(\text{two 6s}) = \binom{3}{2}(1/6)^2(5/6) = 15/216$. Net winnings: \$4.
- We have $P(\text{one 6}) = \binom{3}{1}(1/6)(5/6)^2 = 75/216$. Net winnings: \$0.
- Finally, $P(\text{no 6s}) = (5/6)^3 = 125/216$. Net winnings: $-\$1$.

n	\$65	\$4	\$0	$-\$1$
$P(X = n)$	1/216	15/216	75/216	125/216

More Expected Value, IV

<u>Example</u>: In one version of the game "Chuck-a-luck", three standard 6-sided dice are rolled. Prizes for a bet of $1 are awarded as follows: $66 for a roll of three 6s, $5 for a roll of two 6s, $1 for one 6, and $0 for any other roll.

2. Find the expected value of X.

More Expected Value, IV

<u>Example</u>: In one version of the game "Chuck-a-luck", three standard 6-sided dice are rolled. Prizes for a bet of \$1 are awarded as follows: \$66 for a roll of three 6s, \$5 for a roll of two 6s, \$1 for one 6, and \$0 for any other roll.

2. Find the expected value of X.

n	\$65	\$4	\$0	$-$\$1
$P(X = n)$	$1/216$	$15/216$	$75/216$	$125/216$

- We can now simply use the probability distribution to compute $E(X)$.
- $E(X) = \frac{1}{216} \cdot (\$65) + \frac{15}{216} \cdot (\$4) + \frac{75}{216} \cdot (\$0) + \frac{125}{216} \cdot (-\$1) = \$0.$

For this game, we can see that the expected winnings are \$0, meaning that the game is fair (in the sense that neither the player nor the person running the game should expect to win or lose money on average over the long term).

Properties of Expected Value, I

Expected value has several important algebraic properties:

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 $E(X + Y) = E(X) + E(Y)$.

Properties of Expected Value, I

Expected value has several important algebraic properties:

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 $E(X + Y) = E(X) + E(Y)$.

- Intuitively, if the expected value of X is 4, then it is reasonable to feel that the expected value of $X + 1$ should be 5, while the expected value of $2X$ should be 8. These two observations, taken together, form essentially the first part of the statement.
- Likewise, if the expected value of Y is 3, then it is also reasonable to feel that the expected value of $X + Y$ should be 7, the sum of the expected values of X and Y.

Properties of Expected Value, II

Proof:

- Suppose the outcomes in the sample space are $s_1, s_2, \ldots$ on which X takes on the values $x_1, x_2, \ldots$ and Y takes on the values $y_1, y_2, \ldots$ with probabilities $p_1, p_2, \ldots$, where $p_1 + p_2 + \cdots = 1$.
- Then $E(X) = p_1 x_1 + p_2 x_2 + \cdots$, $E(Y) = p_1 y_1 + p_2 y_2 + \cdots$.
- Since $aX + b$ takes on the values $ax_1 + b$, $ax_2 + b, \ldots$ on the events of the sample space, so
$E(aX + b) = p_1(ax_1 + b) + p_2(ax_2 + b) + \cdots$
$= a(p_1 x_1 + p_2 x_2 + \cdots) + b(p_1 + p_2 + \cdots) = a \cdot E(X) + b$.
- Also, $X + Y$ takes on the values $x_1 + y_1$, $x_2 + y_2, \ldots$ on the events of the sample space, so
$E(X + Y) = p_1(x_1 + y_1) + p_2(x_2 + y_2) + \cdots$
$= (p_1 x_1 + p_2 x_2 + \cdots) + (p_1 y_1 + p_2 y_2 + \cdots) = E(X) + E(Y)$.

Properties of Expected Value, III

Example: An unfair coin with probability p of landing heads is flipped n times. Find the expected number of heads obtained.

Properties of Expected Value, III

Example: An unfair coin with probability p of landing heads is flipped n times. Find the expected number of heads obtained.

- One approach would be to let X be the random variable giving the total number of heads, then compute the probability distribution of X and use the result to find the expected value.
- Since the number of heads is binomially distributed, the probability of obtaining k heads is $\binom{n}{k} p^k (1-p)^{n-k}$, the expected value is $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot k$.
- It is not so obvious how to evaluate this sum, but by incorporating the factor of k into the binomial coefficient, reindexing the sum, and using the binomial theorem, it can be shown that the value is np.

Properties of Expected Value, IV

<u>Example</u>: An unfair coin with probability p of landing heads is flipped n times. Find the expected number of heads obtained.

- The method on the previous slide is not so easy, because it requires us to do some algebraic manipulations of binomial coefficient identities.
- We can give a vastly simpler approach using properties of expected value

Properties of Expected Value, IV

Example: An unfair coin with probability p of landing heads is flipped n times. Find the expected number of heads obtained.

- The method on the previous slide is not so easy, because it requires us to do some algebraic manipulations of binomial coefficient identities.
- We can give a vastly simpler approach using properties of expected value: first, write X as the sum of random variables $X = X_1 + X_2 + \cdots + X_n$, where X_i is the number of heads obtained on the nth flip.
- Since $E(X_1) = E(X_2) = \cdots = E(X_n) = p$ since the flips each have a probability p of landing heads, we can apply the additivity of expectation to see that $E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$.

Variance and Standard Deviation, I

In addition to computing the expected value of a random variable, we also would like to be able to measure how much variation the values have relative to their expected value.

- For example, if one random variable X is always equal to 0, then there is no variation in its value.
- If the random variable Y is equal to -2 half the time and 2 the other half the time, then its expected value is also 0, but there is much more variation in the values of Y than in X.
- A third random variable Z has value randomly selected from $\{-100, -99, \ldots, 0, 1, \ldots, 100\}$. The expected value of this random variable is also 0, but it has even more variation in its values than Y does.

Variance and Standard Deviation, II

We can quantify this "amount of variation" as follows:

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 $\mathrm{var}(X) = E[(X - \mu)^2]$, the expected value of the square of the difference between X and its expectation.

If the random variable X has units, we often also want to have the variation quantified using the same type of units, which we can achieve by taking the square root of the variance:

Definition

The standard deviation of X, denoted $\sigma(X)$, is the square root of the variance: $\sigma(X) = \sqrt{\mathrm{var}(X)}$.

Variance and Standard Deviation, III

Roughly speaking, the standard deviation measures the "average distance" that a typical outcome of X will be from the expected outcome. We can also give another formula for the variance:

Proposition (Variance Formula)

For any discrete random variable X, $\mathrm{var}(X) = E(X^2) - E(X)^2$.

Variance and Standard Deviation, III

Roughly speaking, the standard deviation measures the "average distance" that a typical outcome of X will be from the expected outcome. We can also give another formula for the variance:

Proposition (Variance Formula)

For any discrete random variable X, $\mathrm{var}(X) = E(X^2) - E(X)^2$.

<u>Proof</u>:

- By the linearity of expectation, we can write $E[(X-\mu)^2] = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$.
- Then since $\mu = E(X)$, this formula simplifies to $\mathrm{var}(X) = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$, as claimed.

It is usually faster to evaluate the variance using this formula, rather than the definition.

Variance and Standard Deviation, IV

Example: If a coin with probability p of landing heads is flipped once, find the expected value, variance, and standard deviation of the random variable X giving the number of heads.

Variance and Standard Deviation, IV

Example: If a coin with probability p of landing heads is flipped once, find the expected value, variance, and standard deviation of the random variable X giving the number of heads.

- There are two possible outcomes: either $X = 0$ (probability $1 - p$), or $X = 1$ (probability p).
- The expected value is then $E(X) = (1 - p) \cdot 0 + p \cdot 1 = p$.
- For the variance we have $\mathrm{var}(X) = E(X^2) - [E(X)]^2$.
- Then $E(X^2) = (1 - p) \cdot 0^2 + p \cdot 1^2 = p$, and so using $E(X) = p$ from above, we get $\mathrm{var}(X) = p - p^2 = p(1 - p)$.
- Finally, we have $\sigma(X) = \sqrt{\mathrm{var}(X)} = \sqrt{p(1 - p)}$.

Variance and Standard Deviation, IV

Example: If a coin with probability p of landing heads is flipped once, find the expected value, variance, and standard deviation of the random variable X giving the number of heads.

- There are two possible outcomes: either $X = 0$ (probability $1 - p$), or $X = 1$ (probability p).
- The expected value is then $E(X) = (1 - p) \cdot 0 + p \cdot 1 = p$.
- For the variance we have $\mathrm{var}(X) = E(X^2) - [E(X)]^2$.
- Then $E(X^2) = (1 - p) \cdot 0^2 + p \cdot 1^2 = p$, and so using $E(X) = p$ from above, we get $\mathrm{var}(X) = p - p^2 = p(1 - p)$.
- Finally, we have $\sigma(X) = \sqrt{\mathrm{var}(X)} = \sqrt{p(1 - p)}$.

In particular, when $p = 1/2$ the standard deviation is $1/2$. This agrees with the natural idea that although the expected number of heads obtained when flipping a fair coin is $1/2$, the actual outcome is always a distance $1/2$ away from the expectation.

Variance and Standard Deviation, V

Example: A fair coin is flipped 4 times. Find the expected value, variance, and standard deviation of the random variable X giving the total number of heads obtained.

Variance and Standard Deviation, V

<u>Example</u>: A fair coin is flipped 4 times. Find the expected value, variance, and standard deviation of the random variable X giving the total number of heads obtained.

- Here is the probability distribution of X (it is binomial):

n	0	1	2	3	4
$P(X = n)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

- Since X is binomial, its expected value is $E(X) = 4 \cdot (1/2) = 2$ as we showed last time. (Alternatively, we could compute it using the table above.) Also, $E(X^2) = \frac{1}{16} \cdot 0^2 + \frac{4}{16} \cdot 1^2 + \frac{6}{16} \cdot 2^2 + \frac{4}{16} \cdot 3^2 + \frac{1}{16} \cdot 4^2 = 5.$
- Thus, $\mathrm{var}(X) = E(X^2) - [E(X)]^2 = 5 - 2^2 = 1$, and $\sigma(X) = \sqrt{\mathrm{var}(X)} = 1$ also.

Later, we will establish the general formula $\mathrm{var}(X) = np(1 - p)$ for the variance of a binomial distribution with parameters n and p.

Variance and Standard Deviation, VI

Example: If a standard 6-sided die is rolled once, find the variance and standard deviation of the random variable X giving the result of the die roll.

- Each of the possible outcomes $X = 1, 2, 3, 4, 5, 6$ occurs with probability $1/6$.
- We compute
$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{7}{2} = 3.5,$$
and also
$$E(X^2) = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 3^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 5^2 + \frac{1}{6} \cdot 6^2 = \frac{91}{6}.$$
- Thus, $\mathrm{var}(X) = E(X^2) - [E(X)]^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \approx 2.917$,

  and $\sigma(X) = \sqrt{\mathrm{var}(X)} = \sqrt{\frac{35}{12}} \approx 1.708$.

Variance and Standard Deviation, VII

Example: Compute the expected value, variance, and standard deviation of the random variable X whose probability distribution is given below:

n	1	3	4	8
$P(X = n)$	0.2	0.5	0.1	0.2

Variance and Standard Deviation, VII

Example: Compute the expected value, variance, and standard deviation of the random variable X whose probability distribution is given below:

n	1	3	4	8
$P(X = n)$	0.2	0.5	0.1	0.2

- We compute $E(X) = 0.2 \cdot 1 + 0.5 \cdot 3 + 0.1 \cdot 4 + 0.2 \cdot 8 = 3.7$.
- Also, $E(X^2) = 0.2 \cdot 1^2 + 0.5 \cdot 3^2 + 0.1 \cdot 4^2 + 0.2 \cdot 8^2 = 19.1$.
- Thus, $\mathrm{var}(X) = E(X^2) - E(X)^2 = 19.1 - 3.7^2 = 5.41$, and $\sigma(X) = \sqrt{5.41} \approx 2.326$.

Variance Properties, I

As we saw earlier, the expected value of a random variable does not always exist, and even when it does, it can be infinite. We might ask about similar pathologies for the variance.

- The discrete random variable X whose value is 2^n occurring with probability 2^{-n} for $n \geq 1$ has infinite expected value.
- The discrete random variable Y whose value is $(-2)^n$ occurring with probability 2^{-n} for $n \geq 1$ has undefined expected value.

As long as the expected value exists, however, the variance will also exist, because it is computed by summing nonnegative values.

- However, even when the expected value is finite, the variance can, peculiarly, be infinite.

Variance Properties, II

Example: Show that the random variable that takes the value 2^n with probability $2/3^n$, for integers $n \geq 1$, has a finite expected value but an infinite variance.

Variance Properties, II

Example: Show that the random variable that takes the value 2^n with probability $2/3^n$, for integers $n \geq 1$, has a finite expected value but an infinite variance.

- First recall the formula $a + ar + ar^2 + \cdots = a/(1-r)$ for the sum of a geometric series, which is valid for $|r| < 1$.
- Using the formula, the expected value is $E(X) = 2 \cdot \dfrac{2}{3} + 2^2 \cdot \dfrac{2}{9} + 2^3 \cdot \dfrac{2}{27} + \cdots = 4$.
- For the variance, we also must compute $E(X^2) = 4 \cdot \dfrac{2}{3} + 4^2 \cdot \dfrac{2}{9} + 4^3 \cdot \dfrac{2}{27} + \cdots$.
- But since the common ratio in this geometric series is $4/3 > 1$, the sum is infinite: thus, $E(X^2) = \infty$.
- But then the variance is $\mathrm{var}(X) = E(X^2) - E(X)^2 = \infty$, which is to say, the variance is infinite.

Variance Properties, III

Like expected value, variance also possesses some convenient algebraic properties:

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 $\sigma(aX + b) = |a|\,\sigma(X)$.

Variance Properties, III

Like expected value, variance also possesses some convenient algebraic properties:

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 $\sigma(aX + b) = |a|\, \sigma(X)$.

Proof:

- From the linearity of expectation, we know that $E(aX + b) = a \cdot E(X) + b$, and therefore we have $aX + b - E(aX + b) = aX + b - a\, E(X) - b = a \cdot [X - E(x)]$.
- Then $\mathrm{var}(aX + b) = E[(aX + b - E(aX + b))^2] = E[a^2 \cdot (X - E(x))^2] = a^2 \mathrm{var}(X)$, and by taking the square root of both sides we then get $\sigma(aX + b) = |a|\, \sigma(X)$.

Variance Properties, IV

Example: If X is a random variable with expected value 1 and standard deviation 3, what are the expected value and standard deviation of $2X + 4$?

Variance Properties, IV

Example: If X is a random variable with expected value 1 and standard deviation 3, what are the expected value and standard deviation of $2X + 4$?

- From the properties of expected value and variance, we have $E(2X + 4) = 2E(X) + 4 = \boxed{6}$, and $\sigma(2X + 4) = |2|\,\sigma(X) = \boxed{6}$.

Joint Distributions, I

We now discuss the situation of having several discrete random variables defined on the same sample space, to extend some of our analysis of conditional probability and independence to random variables.

- If we have a collection of random variables $X_1, X_2, \ldots, X_n$, we can summarize all of the possible information about the behavior of these random variables simultaneously using a joint probability density distribution, which simply lists all the possible collections of values of these random variables together with their probabilities.

[Note: The material from this slide onward is not on midterm 1, but is eligible for midterm 2.]

Joint Distributions, II

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 <u>joint probability density function</u> of $X_1, X_2, \ldots, X_n$.

- The joint probability density function simply measures the probability that the various random variables take particular values, for all combinations of possible values.
- For the situation of two random variables X and Y, we can display the joint probability density function by tabulating all of the possible values of X and Y in a grid.

Joint Distributions, III

<u>Example</u>: The joint distribution for the two discrete random variables X and Y is given below.

$X \setminus Y$	1	4	5
0	0.1	0.2	0
2	0	0.3	0.2
4	0.1	0.1	0

- We interpret the table as follows: each entry in the table gives the probability that X and Y take the indicated values (in the row and column headers) together.
- Thus, for example, the probability that $X = 0$ and $Y = 1$ at the same time is 0.1, while the probability that $X = 2$ and $Y = 5$ at the same time is 0.2, and the probability that $X = 4$ and $Y = 5$ at the same time is 0.

Joint Distributions, IV

Example: An unfair coin that comes up heads 2/3 of the time is flipped 3 times. Let X be the random variable counting the total number of heads and Y is the random variable counting the longest run of tails. Find the following:

1. The joint probability distribution table for X and Y.
2. $P(X = Y = 1)$.
3. $P(X = 3, Y = 1)$.
4. $P(X = Y)$.
5. $P(Y = 2)$.
6. $P(Y - X = 1)$.
7. $P(X + Y = 3)$.

Joint Distributions, IV

<u>Example</u>: An unfair coin that comes up heads 2/3 of the time is flipped 3 times. Let X be the random variable counting the total number of heads and Y is the random variable counting the longest run of tails. Find the following:

1. The joint probability distribution table for X and Y.
2. $P(X = Y = 1)$.
3. $P(X = 3, Y = 1)$.
4. $P(X = Y)$.
5. $P(Y = 2)$.
6. $P(Y - X = 1)$.
7. $P(X + Y = 3)$.

We first tabulate the various outcomes and the values of X and Y on each. Then we use these calculations to set up the joint distribution table.

Joint Distributions, V

Example: An unfair coin that comes up heads $2/3$ of the time is flipped 3 times. Let X be the random variable counting the total number of heads and Y is the random variable counting the longest run of tails. Find the following:

1. The joint probability distribution table for X and Y.

Joint Distributions, V

Example: An unfair coin that comes up heads $2/3$ of the time is flipped 3 times. Let X be the random variable counting the total number of heads and Y is the random variable counting the longest run of tails. Find the following:

1. The joint probability distribution table for X and Y.

Here is a table of all the possible outcomes, their probabilities, and the values of X and Y on each:

Outcome(s)	Probability	Value of X	Value of Y
HHH	$8/27$	3	0
HHT, HTH, THH	$4/27$	2	1
HTT, TTH	$2/27$	1	2
THT	$2/27$	1	1
TTT	$1/27$	0	3

Joint Distributions, VI

Example: An unfair coin that comes up heads 2/3 of the time is flipped 3 times. Let X be the random variable counting the total number of heads and Y is the random variable counting the longest run of tails. Find the following:

1. The joint probability distribution table for X and Y.

Joint Distributions, VI

Example: An unfair coin that comes up heads $2/3$ of the time is flipped 3 times. Let X be the random variable counting the total number of heads and Y is the random variable counting the longest run of tails. Find the following:

1. The joint probability distribution table for X and Y.

Using the table of outcomes, we can construct the joint probability distribution table:

$Y \setminus X$	0	1	2	3
0	0	0	0	$8/27$
1	0	$2/27$	$12/27$	0
2	0	$4/27$	0	0
3	$1/27$	0	0	0

Joint Distributions, VI

<u>Example</u>: Use the joint distribution table to find:

2. $P(X = Y = 1)$.
3. $P(X = 3, Y = 1)$.
4. $P(X = Y)$.
5. $P(Y = 2)$.
6. $P(Y - X = 1)$.
7. $P(X + Y = 3)$.

$Y \setminus X$	0	1	2	3
0	0	0	0	8/27
1	0	2/27	12/27	0
2	0	4/27	0	0
3	1/27	0	0	0

Joint Distributions, VI

Example: Use the joint distribution table to find:

2. $P(X = Y = 1)$.
3. $P(X = 3, Y = 1)$.
4. $P(X = Y)$.
5. $P(Y = 2)$.
6. $P(Y - X = 1)$.
7. $P(X + Y = 3)$.

$Y \setminus X$	0	1	2	3
0	0	0	0	$8/27$
1	0	$2/27$	$12/27$	0
2	0	$4/27$	0	0
3	$1/27$	0	0	0

- We see $P(X = Y = 1) = 2/27$, $P(Y = 2) = 4/27$, $P(X = 3, Y = 1) = 0$, $P(Y - X = 1) = 4/27$, $P(X = Y) = 2/27$, and $P(X + Y = 3) = (1/27) + (4/27) + (12/27) + (8/27) = 25/27$.

Joint Distributions, VII

<u>Example</u>: The joint probability distribution for the random variables I and O counting the number of diners sitting inside and outside (respectively) at a small cafe in August is given below. Find

1. $P(I = 2, O = 3)$
2. $P(I = 1)$
3. $P(O = 2)$
4. $P(I + O \geq 6)$
5. $P(I = O)$
6. $P(|I - O| > 2)$
7. The probability distribution for O.
8. The probability distribution for I.

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

Joint Distributions, VIII

1. $P(I = 2, O = 3)$
2. $P(I = 1)$
3. $P(O = 2)$
4. $P(I + O \geq 6)$
5. $P(I = O)$
6. $P(|I - O| > 2)$

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

Joint Distributions, VIII

1. $P(I = 2, O = 3)$
2. $P(I = 1)$
3. $P(O = 2)$
4. $P(I + O \geq 6)$
5. $P(I = O)$
6. $P(|I - O| > 2)$

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

- We see $P(I = 2, O = 3) = 0.01$,
  $P(I = 1) = 0.01 + 0.07 + 0.10 + 0.08 + 0.06 = 0.32$,
  $P(O = 2) = 0.12 + 0.07 = 0.19$,
  $P(I + O \geq 6) = 0.03 + 0.06 + 0.02 + 0.01 = 0.12$,
  $P(I = O) = 0.03 + 0.01 = 0.04$, and
  $P(|I - O| > 2) = 0.18 + 0.10 + 0.08 + 0.08 + 0.06 + 0.02 = 0.52$.

Joint Distributions, IX

7. The probability distribution for O.

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

Joint Distributions, IX

7. The probability distribution for O.

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

- To find the probability distribution for O by itself, we simply sum over all of the corresponding entries in the table having the same value for O (i.e., down the columns).

O	0	1	2	3	4	5
Probability	0.03	0.11	0.19	0.29	0.21	0.17

Joint Distributions, X

8. The probability distribution for I.

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

Joint Distributions, X

8. The probability distribution for I.

$I \setminus O$	0	1	2	3	4	5
0	0.03	0.10	0.12	0.18	0.10	0.08
1	0	0.01	0.07	0.10	0.08	0.06
2	0	0	0	0.01	0.03	0.02
3	0	0	0	0	0	0.01

- For I, we instead sum across the rows:

I	Probability
0	0.61
1	0.32
2	0.06
3	0.01

Marginal Distributions, I

We can recover individual distributions from the joint distribution:

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 single-variable probability density functions for X and Y as $p_X(a) = \sum_y p_{X,Y}(a,y)$ and $p_Y(b) = \sum_x p_{X,Y}(x,b)$.

Marginal Distributions, I

We can recover individual distributions from the joint distribution:

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 single-variable probability density functions for X and Y as $p_X(a) = \sum_y p_{X,Y}(a,y)$ and $p_Y(b) = \sum_x p_{X,Y}(x,b)$.

Proof:

- Observe that the event $\{E : X = a\}$ is the union over all real numbers y of the sets $\{E : X = a, Y = y\}$.
- Since each of these sets are disjoint (since the random variable Y can only take one value at a time), we can simply sum the corresponding probabilities by the probability axioms.
- This yields the first formula. The second follows similarly.

Marginal Distributions, II

A probability density function obtained by restricting a given probability distribution to a subset is called a <u>marginal probability distribution</u>.

- The proposition gives the procedure for computing the marginal probability distribution on the subsets $X = a$ (as a varies) and also the subsets of the form $Y = b$ (as b varies).
- The word "marginal" is used to evoke the idea of writing the row and column sums in the margins of the probability distribution table. When using an actual table, we often do write the sums this way.

This result also extends to more than two variables in a fairly natural way (in short: to find the joint distribution for a subset, simply sum over the other variables).

Marginal Distributions, III

Example: Here are the marginal distributions, displayed as row and column sums, for the joint I/O distribution in the last example:

$I \setminus O$	0	1	2	3	4	5	Sum (I)
0	0.03	0.10	0.12	0.18	0.10	0.08	0.61
1	0	0.01	0.07	0.10	0.08	0.06	0.32
2	0	0	0	0.01	0.03	0.02	0.06
3	0	0	0	0	0	0.01	0.01
Sum (O)	0.03	0.11	0.19	0.29	0.21	0.17	

Independence, I

We would now like to use joint distributions to describe when two random variables are independent.

- Like with independence of events in probability spaces, we would say that two random variables X and Y are independent when knowing the value of one gives no additional information about the value of the other.

Independence, I

We would now like to use joint distributions to describe when two random variables are independent.

- Like with independence of events in probability spaces, we would say that two random variables X and Y are independent when knowing the value of one gives no additional information about the value of the other.
- Explicitly, we want $P(X = a|Y = b) = P(X = a)$ for all a, b.
- By our results on conditional probability, this is the same as saying $P(X = a, Y = b) = P(X = a) \cdot P(Y = b)$.
- Recast in the language of probability density functions, this says $p_{X,Y}(a, b) = p_X(a) \cdot p_Y(b)$.
- In other words, the probability that both $X = a$ and $Y = b$ is the product of the probabilities of those two separate events (namely, that $X = a$ and that $Y = b$).

Independence, II

Per the discussion on the previous slide, we can define independence of two random variables as follows:

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.

Note the similarity to the condition for independence of events: $P(A \cap B) = P(A) \cdot P(B)$.

Independence, III

Example: If X and Y have the joint distribution displayed below, determine whether X and Y are independent.

$X \setminus Y$	0	1	2	3
1	0.12	0.18	0.24	0.06
3	0.02	0.03	0.04	0.01
5	0.06	0.09	0.12	0.03

Independence, III

Example: If X and Y have the joint distribution displayed below, determine whether X and Y are independent.

$X \setminus Y$	0	1	2	3
1	0.12	0.18	0.24	0.06
3	0.02	0.03	0.04	0.01
5	0.06	0.09	0.12	0.03

We must first compute the probability distributions for X and Y, which we may do by summing the rows and columns, and then we must check whether $p_{X,Y}(a,b) = p_X(a) \cdot p_Y(b)$ for each (a,b) in the table.

Independence, IV

Example: If X and Y have the joint distribution displayed below, determine whether X and Y are independent.

$X \setminus Y$	0	1	2	3	Sum
1	0.12	0.18	0.24	0.06	0.6
3	0.02	0.03	0.04	0.01	0.1
5	0.06	0.09	0.12	0.03	0.3
Sum	0.2	0.3	0.4	0.1	

Independence, IV

Example: If X and Y have the joint distribution displayed below, determine whether X and Y are independent.

$X \setminus Y$	0	1	2	3	Sum
1	0.12	0.18	0.24	0.06	0.6
3	0.02	0.03	0.04	0.01	0.1
5	0.06	0.09	0.12	0.03	0.3
Sum	0.2	0.3	0.4	0.1	

- The row and column sums have been added to the table.
- Now we just need to check whether each entry is the product of its row sum and column sum.
- For the top left entry, we see $0.12 = 0.6 \cdot 0.2$ as required.
- In fact, the calculation works out correctly for each entry, so X and Y are independent.

Independence, V

<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, determine whether X and Y are independent.

Independence, V

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, determine whether X and Y are independent.

- Intuitively, we would expect that these variables should not be independent, since both X and Y will be affected by the outcome of the second coin flip.
- Indeed, we have $P(X = 2, Y = 0) = 0$ since $X = 2$ requires the middle flip to be heads while $Y = 0$ requires the middle flip to be tails.
- However, $P(X = 2) = \dfrac{1}{4}$ and $P(Y = 0) = \dfrac{1}{4}$ also, and so $P(X = 2) \cdot P(Y = 0) = \dfrac{1}{16} \neq P(X = 2, Y = 0) = 0.$
- Thus, X and Y are not independent.

Independence, VI

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, determine whether X and Y are independent.

Independence, VI

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, determine whether X and Y are independent.

- Alternatively, here is the full joint distribution of X and Y:

$X \setminus Y$	0	1	2	Sum
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
Sum	1/4	1/2	1/4	

- We can see that there are four entries (the four corner entries) that are not equal to the product of the corresponding row and column sums, so any of these would yield an appropriate counterexample.

Independence, VII

Just like with probabilities, we may easily extend the notion of independence to more than two random variables.

- The analogous condition is that $X_1, X_2, \ldots, X_n$ are independent when, for any subset $Y_1, \ldots, Y_k$ of the X_i, the joint distribution $p_{Y_1,\ldots,Y_k}(a_1, \ldots, a_k)$ is equal to the product of the individual distributions $p_{Y_1}(a_1) \cdot \cdots \cdot p_{Y_k}(a_k)$.
- However, we may compute all of these joint distributions using the single joint distribution $p_{X_1,X_2,\ldots,X_n}(a_1, a_2, \ldots, a_n)$ (namely, by summing over all of the possible values of the random variables we are not considering).
- So in fact, in fact all of these conditions follow from the single condition that $p_{X_1,X_2,\ldots,X_n}(a_1, a_2, \ldots, a_n) = p_{X_1}(a_1) \cdot p_{X_2}(a_2) \cdot \cdots \cdot p_{X_n}(a_n)$.

Independence, VIII

Definition

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

As a practical matter, unless we have convenient formulas for the random variables, it is fairly cumbersome to work with joint distributions involving 3 or more variables at a time. Thus, we will primarily do calculations in the 2-variable case only.

Summary

We defined the expected value of a discrete random variable and examined some of its properties.

We defined the variance and standard deviation of a discrete random variable and examined some of their properties.

We introduced joint distributions and independence of discrete random variables.

Next lecture: More with independence, covariance, and correlation.