Ch. 4 - Discrete Random Variables and Probability Distributions

Ercan Karadas

New York University

ercan@nyu.edu

Summer 2016

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.
- We know the sample space of this experiment

$$S = \{HHHHH, HHHHT, \ldots, \}$$

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.
- We know the sample space of this experiment

$$S = \{HHHHH, HHHHT, \ldots, \}$$

- But may be we are not interested in just listing the elements in S, but something else...
- For example, we could gamble on the # of Heads in this experiment
  - if # of H > # T, I will give you \$8

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.
- We know the sample space of this experiment

$$S = \{HHHHH, HHHHT, \ldots, \}$$

- But may be we are not interested in just listing the elements in S, but something else...
- For example, we could gamble on the # of Heads in this experiment
  - if # of H $>$ # T, I will give you \$8
  - if # of H $<$ # T, you will give you \$8

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.
- We know the sample space of this experiment

$$S = \{HHHHH, HHHHT, \ldots, \}$$

- But may be we are not interested in just listing the elements in S, but something else...
- For example, we could gamble on the # of Heads in this experiment
  - if # of H $>$ # T, I will give you \$8
  - if # of H $<$ # T, you will give you \$8
- We are only interested in the number of Heads generated. Therefore, let us define a new variable to denote this:

$$X = \# \text{ of Heads}$$

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.
- We know the sample space of this experiment

$$S = \{HHHHH, HHHHT, \ldots, \}$$

- But may be we are not interested in just listing the elements in S, but something else...
- For example, we could gamble on the # of Heads in this experiment
  - if # of H $>$ # T, I will give you \$8
  - if # of H $<$ # T, you will give you \$8
- We are only interested in the number of Heads generated. Therefore, let us define a new variable to denote this:

$$X = \# \text{ of Heads}$$

  - What are possible values of X?

$$X \in \{0, 1, 2, 3, 4, 5\}$$

The Leading Example

- Consider the following experiment: Tossing a coin 5 times.
- We know the sample space of this experiment

$$S = \{HHHHH, HHHHT, \dots, \}$$

- But may be we are not interested in just listing the elements in S, but something else...
- For example, we could gamble on the # of Heads in this experiment
  - if # of H > # T, I will give you \$8
  - if # of H < # T, you will give you \$8
- We are only interested in the number of Heads generated. Therefore, let us define a new variable to denote this:

$$X = \#\text{ of Heads}$$

  - What are possible values of X?

$$X \in \{0, 1, 2, 3, 4, 5\}$$

  - Now we can rewrite the previous conditions as

$$\text{Experiment } = \begin{cases} \text{you win if} & X \geq 3 \\ \text{I win if} & X \leq 2 \end{cases}$$

- What about which bet you would take, H or T?

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.
- But now suppose that I am really confident that I am a lucky person (or know that the coin is biased!) so I offer you the following gamble

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.
- But now suppose that I am really confident that I am a lucky person (or know that the coin is biased!) so I offer you the following gamble

$$\text{Lucky Experiment} = \begin{cases} \text{You win if} & X \geq 2 \\ \text{I win if} & X \leq 1 \end{cases}$$

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.
- But now suppose that I am really confident that I am a lucky person (or know that the coin is biased!) so I offer you the following gamble

$$\text{Lucky Experiment} = \begin{cases} \text{You win if} & X \geq 2 \\ \text{I win if} & X \leq 1 \end{cases}$$

- Of course, if the coin is unbiased you would accept the gamble, right?

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.
- But now suppose that I am really confident that I am a lucky person (or know that the coin is biased!) so I offer you the following gamble

$$\text{Lucky Experiment } = \begin{cases} \text{You win if} & X \geq 2 \\ \text{I win if} & X \leq 1 \end{cases}$$

- Of course, if the coin is unbiased you would accept the gamble, right?
- What about the following question:

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.
- But now suppose that I am really confident that I am a lucky person (or know that the coin is biased!) so I offer you the following gamble

$$\text{Lucky Experiment } = \begin{cases} \text{You win if} & X \geq 2 \\ \text{I win if} & X \leq 1 \end{cases}$$

- Of course, if the coin is unbiased you would accept the gamble, right?
- What about the following question:
- "I am going to offer this gamble to the person who offers the highest participation fee".
  (i.e. first everyone decides how much to offer take the gamble. Then, if you are the one with the highest offer, I claim the participation fee from you first and then the result of Lucky Experiment determines who is going to receive \$8.)

- What about which bet you would take, H or T?
  - Depends on the winning probabilities, i.e. $P(X \geq 3)$
  - Both probabilities are the same, so you would be indifferent between H and T, so this is not interesting.
- But now suppose that I am really confident that I am a lucky person (or know that the coin is biased!) so I offer you the following gamble

$$\text{Lucky Experiment } = \begin{cases} \text{You win if} & X \geq 2 \\ \text{I win if} & X \leq 1 \end{cases}$$

- Of course, if the coin is unbiased you would accept the gamble, right?
- What about the following question:
- "I am going to offer this gamble to the person who offers the highest participation fee".
  (i.e. first everyone decides how much to offer take the gamble. Then, if you are the one with the highest offer, I claim the participation fee from you first and then the result of Lucky Experiment determines who is going to receive \$8.)
- How much you would offer at maximum to take the gamble?

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make

$$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make
$$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$
  - Let us find these probabilities fist
$$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& (\frac{1}{2})^5 + 5 \cdot (\frac{1}{2})^5 \\ =& 6 \cdot (\frac{1}{2})^5 = \frac{6}{32} \end{aligned}$$
  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$
  - Then **average return**
$$\begin{aligned} P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8) &= \left[P(X \geq 2) - P(X \leq 1)\right] \cdot 8 \\ &= (\frac{26}{32} - \frac{6}{32}) \cdot 8 = 5 \end{aligned}$$
  - You would **expect** to make \$5 if you had chance to participate the gamble.
  - Therefore, you would be willing to pay at most \$5 participation fee.

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make
  $$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$
  - Let us find these probabilities fist
  $$P(X \leq 1) = P(X = 0) + P(X = 1)$$
  $$= 6 \cdot \left(\frac{1}{2}\right)^5 = \frac{6}{32}$$
  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$
  - Then **average return**
  $$\begin{aligned} P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8) &= \left[P(X \geq 2) - P(X \leq 1)\right] \cdot 8 \\ &= \left(\frac{26}{32} - \frac{6}{32}\right) \cdot 8 = 5 \end{aligned}$$
  - You would **expect** to make \$5 if you had chance to participate the gamble.
  - Therefore, you would be willing to pay at most \$5 participation fee.

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make

$$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$

  - Let us find these probabilities fist

$$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& \left(\frac{1}{2}\right)^5 + 5 \cdot \left(\frac{1}{2}\right)^5 \\ =& 6 \cdot \left(\frac{1}{2}\right)^5 = \frac{6}{32} \end{aligned}$$

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make

$$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$

  - Let us find these probabilities fist

$$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& \left(\frac{1}{2}\right)^5 + 5 \cdot \left(\frac{1}{2}\right)^5 \\ =& 6 \cdot \left(\frac{1}{2}\right)^5 = \frac{6}{32} \end{aligned}$$

  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make
$$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$
  - Let us find these probabilities fist
$$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& (\frac{1}{2})^5 + 5 \cdot (\frac{1}{2})^5 \\ =& 6 \cdot (\frac{1}{2})^5 = \frac{6}{32} \end{aligned}$$
  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$
  - Then **average return**
$$\begin{aligned} P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8) &= \left[P(X \geq 2) - P(X \leq 1)\right] \cdot 8 \\ &= (\frac{26}{32} - \frac{6}{32}) \cdot 8 = 5 \end{aligned}$$
  - You would **expect** to make \$5 if you had chance to participate the gamble.
  - Therefore, you would be willing to pay at most \$5 participation fee.

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make

$$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$

  - Let us find these probabilities fist

$$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& \left(\frac{1}{2}\right)^5 + 5 \cdot \left(\frac{1}{2}\right)^5 \\ =& 6 \cdot \left(\frac{1}{2}\right)^5 = \frac{6}{32} \end{aligned}$$

  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$
  - Then **average return**

$$\begin{aligned} P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8) &= \left[P(X \geq 2) - P(X \leq 1)\right] \cdot 8 \\ &= \left(\frac{26}{32} - \frac{6}{32}\right) \cdot 8 = 5 \end{aligned}$$

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make
  $$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$
  - Let us find these probabilities fist
  $$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& (\frac{1}{2})^5 + 5 \cdot (\frac{1}{2})^5 \\ =& 6 \cdot (\frac{1}{2})^5 = \frac{6}{32} \end{aligned}$$
  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$
  - Then **average return**
  $$\begin{aligned} P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8) &= \left[P(X \geq 2) - P(X \leq 1)\right] \cdot 8 \\ &= (\frac{26}{32} - \frac{6}{32}) \cdot 8 = 5 \end{aligned}$$
  - You would **expect** to make \$5 if you had chance to participate the gamble.

- Finally, we got an interesting question to think about!
- In many contexts these type of problems are abundant, for example auctions, lotteries.
- Let us see how we can proceed
  - If you were already offered the gamble, **on average** you would expect to make
  $$P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8)$$
  - Let us find these probabilities fist
  $$\begin{aligned} P(X \leq 1) =& P(X = 0) + P(X = 1) \\ =& P(TTTTT) + P(4T \text{ and } 1H) \\ =& (\frac{1}{2})^5 + 5 \cdot (\frac{1}{2})^5 \\ =& 6 \cdot (\frac{1}{2})^5 = \frac{6}{32} \end{aligned}$$
  - Since $P(X \leq 1) + P(X \geq 2) = 1$, we also have $P(X \geq 2) = \frac{26}{32}$
  - Then **average return**
  $$\begin{aligned} P(X \leq 1) \cdot (-8) + P(X \geq 2) \cdot (8) &= \big[P(X \geq 2) - P(X \leq 1)\big] \cdot 8 \\ &= (\frac{26}{32} - \frac{6}{32}) \cdot 8 = 5 \end{aligned}$$
  - You would **expect** to make \$5 if you had chance to participate the gamble.
  - Therefore, you would be willing to pay at most \$5 participation fee.

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

can take a different value in each trial of the experiment.

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

can take a different value in each trial of the experiment. Since we cannot tell which value will be assumed in the next trial, we call it a **random variable**. In general,

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

  can take a different value in each trial of the experiment. Since we cannot tell which value will be assumed in the next trial, we call it a **random variable**. In general,

Random Variable:

represents a possible numerical value from a (random) experiment.

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

  can take a different value in each trial of the experiment. Since we cannot tell which value will be assumed in the next trial, we call it a **random variable**. In general,

Random Variable:

represents a possible numerical value from a (random) experiment.

- Random variables (r.v.) are defined on their respective sample spaces. In another words, for each element of the sample space r.v. associates a numerical value.

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

  can take a different value in each trial of the experiment. Since we cannot tell which value will be assumed in the next trial, we call it a **random variable**. In general,

Random Variable:

represents a possible numerical value from a (random) experiment.

- Random variables (r.v.) are defined on their respective sample spaces. In another words, for each element of the sample space r.v. associates a numerical value.
- If a random variable can take on no more than a countable number of values, we call it a **Discrete Random Variable**.

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

  can take a different value in each trial of the experiment. Since we cannot tell which value will be assumed in the next trial, we call it a **random variable**. In general,

Random Variable:

represents a possible numerical value from a (random) experiment.

- Random variables (r.v.) are defined on their respective sample spaces. In another words, for each element of the sample space r.v. associates a numerical value.
- If a random variable can take on no more than a countable number of values, we call it a **Discrete Random Variable**.
- For example in our leading example,

- As we have seen in the previous example, a variable like

$$X = \# \text{ of Heads in 5 coin tossing}$$

  can take a different value in each trial of the experiment. Since we cannot tell which value will be assumed in the next trial, we call it a **random variable**. In general,

Random Variable:

represents a possible numerical value from a (random) experiment.

- Random variables (r.v.) are defined on their respective sample spaces. In another words, for each element of the sample space r.v. associates a numerical value.
- If a random variable can take on no more than a countable number of values, we call it a **Discrete Random Variable**.
- For example in our leading example,

S(sample space)	X
TTTTT	0
TTTTH	1
TTHTT	1
HHTTH	3
⋮	⋮

- Let us add another column to write corresponding probabilities

- Let us add another column to write corresponding probabilities

S(sample space)	X	Prob
TTTTT	0	$(1/2)^5$
TTTTH	1	$(1/2)^5$
TTHTT	1	$(1/2)^5$
HHTTH	3	$(1/2)^5$
$\vdots$	$\vdots$	$\vdots$

- Let us add another column to write corresponding probabilities

S(sample space)	X	Prob
TTTTT	0	$(1/2)^5$
TTTTH	1	$(1/2)^5$
TTHTT	1	$(1/2)^5$
HHTTH	3	$(1/2)^5$
$\vdots$	$\vdots$	$\vdots$

- We can think of the table above as a concise mathematical representation of the experiment. However, for our purposes, as in the leading example, this representation is not the most useful one.
- Since all we are interested in is the possible values that X can assume and their corresponding probabilities, it is a good idea to sum up probabilities over the rows with the same X value:

- Let us add another column to write corresponding probabilities

S(sample space)	X	Prob
TTTTT	0	$(1/2)^5$
TTTTH	1	$(1/2)^5$
TTHTT	1	$(1/2)^5$
HHTTH	3	$(1/2)^5$
$\vdots$	$\vdots$	$\vdots$

- We can think of the table above as a concise mathematical representation of the experiment. However, for our purposes, as in the leading example, this representation is not the most useful one.
- Since all we are interested in is the possible values that X can assume and their corresponding probabilities, it is a good idea to sum up probabilities over the rows with the same X value:

X	$P(X)$
0	$1 \cdot (1/2)^5$
1	$5 \cdot (1/2)^5$
2	$10 \cdot (1/2)^5$
3	$10 \cdot (1/2)^5$
4	$5 \cdot (1/2)^5$
5	$1 \cdot (1/2)^5$

- Let us add another column to write corresponding probabilities

S(sample space)	X	Prob
TTTTT	0	$(1/2)^5$
TTTTH	1	$(1/2)^5$
TTHTT	1	$(1/2)^5$
HHTTH	3	$(1/2)^5$
$\vdots$	$\vdots$	$\vdots$

- We can think of the table above as a concise mathematical representation of the experiment. However, for our purposes, as in the leading example, this representation is not the most useful one.
- Since all we are interested in is the possible values that X can assume and their corresponding probabilities, it is a good idea to sum up probabilities over the rows with the same X value:

X	$P(X)$
0	$1 \cdot (1/2)^5$
1	$5 \cdot (1/2)^5$
2	$10 \cdot (1/2)^5$
3	$10 \cdot (1/2)^5$
4	$5 \cdot (1/2)^5$
5	$1 \cdot (1/2)^5$

- This table is called **Probability Distribution of X**

- We will use the following notation

- We will use the following notation

$$P(X = x) = \text{ probability that r.v. } X \text{ will take the value x}$$

- We will use the following notation

$$P(X = x) = \text{ probability that r.v. } X \text{ will take the value x}$$

- Here P is called the **Probability Distribution Function** of X, i.e. assigning probabilities for each possible value that X can assume.

- We will use the following notation

$$P(X = x) = \text{ probability that r.v. } X \text{ will take the value x}$$

- Here P is called the **Probability Distribution Function** of X, i.e. assigning probabilities for each possible value that X can assume.

Properties of P(x)

- We will use the following notation

$$P(X = x) = \text{ probability that r.v. } X \text{ will take the value x}$$

- Here P is called the **Probability Distribution Function** of X, i.e. assigning probabilities for each possible value that X can assume.

Properties of P(x)

1. $0 \leq P(X = x) \leq 1$, for each x.

- We will use the following notation

$$P(X = x) = \text{ probability that r.v. } X \text{ will take the value x}$$

- Here P is called the **Probability Distribution Function** of X, i.e. assigning probabilities for each possible value that X can assume.

Properties of P(x)

1. $0 \leq P(X = x) \leq 1$, for each x.
2. $\sum_x P(X = x) = 1$,

Leading Example again

- Let us go back to the leading example but now assume that the coin is actually biased

Leading Example again

- Let us go back to the leading example but now assume that the coin is actually biased

$$P(H) = 0.4 \text{ and } P(T) = 0.6$$

Leading Example again

- Let us go back to the leading example but now assume that the coin is actually biased
$$P(H) = 0.4 \text{ and } P(T) = 0.6$$
- Construct the probability distribution table again and calculate how much you would be willing to pay in this case?

Leading Example again

- Let us go back to the leading example but now assume that the coin is actually biased
$$P(H) = 0.4 \text{ and } P(T) = 0.6$$
- Construct the probability distribution table again and calculate how much you would be willing to pay in this case?
- Let us start with this

S(sample space)	X	Prob
TTTTT	0	$(0.6)^5$
TTTTH	1	$(0.4)(0.6)^4$
TTHTT	1	$(0.4)(0.6)^4$
HHTTH	3	$(0.4)^3(0.6)^2$
$\vdots$	$\vdots$	$\vdots$

Leading Example again

- Let us go back to the leading example but now assume that the coin is actually biased
$$P(H) = 0.4 \text{ and } P(T) = 0.6$$
- Construct the probability distribution table again and calculate how much you would be willing to pay in this case?
- Let us start with this

S(sample space)	X	Prob
TTTTT	0	$(0.6)^5$
TTTTH	1	$(0.4)(0.6)^4$
TTHTT	1	$(0.4)(0.6)^4$
HHTTH	3	$(0.4)^3(0.6)^2$
$\vdots$	$\vdots$	$\vdots$

- As before, we can also construct the probability distribution function

X	$P(X = x)$
0	$(0.6)^5$
1	$C_1^5 \cdot (0.4)^1(0.6)^4$
2	$C_2^5 \cdot (0.4)^2(0.6)^3$
3	$C_3^5 \cdot (0.4)^3(0.6)^2$
4	$C_4^5 \cdot (0.4)^4(0.6)^1$
5	$C_5^5 \cdot (0.4)^5(0.6)^0$

Leading Example again

- Let us generalize the previous set up further by assuming

Leading Example again

- Let us generalize the previous set up further by assuming

$$P(H) = p \text{ and } P(T) = 1 - p$$

Leading Example again

- Let us generalize the previous set up further by assuming

$$P(H) = p \text{ and } P(T) = 1 - p$$

- Then we can revise the table as follows

Leading Example again

- Let us generalize the previous set up further by assuming

$$P(H) = p \text{ and } P(T) = 1 - p$$

- Then we can revise the table as follows

X	$P(X = x)$
0	$C_0^5 \cdot p^0(1-p)^5$
1	$C_1^5 \cdot p^1(1-p)^4$
2	$C_2^5 \cdot p^2(1-p)^3$
3	$C_3^5 \cdot p^3(1-p)^2$
4	$C_4^5 \cdot p^4(1-p)^1$
5	$C_5^5 \cdot p^5(1-p)^0$

- Finally, note that if the experiment consisted of n coin tossing instead of 5, we would have

X	$P(X = x)$
0	$C_0^n \cdot p^0(1-p)^n$
1	$C_1^n \cdot p^1(1-p)^{n-1}$
2	$C_2^n \cdot p^2(1-p)^{n-2}$
3	$C_3^n \cdot p^3(1-p)^{n-3}$
$\vdots$	$\vdots$
n-1	$C_{n-1}^n \cdot p^{n-1}(1-p)^1$
n	$C_n^n \cdot p^n(1-p)^0$

- For example, suppose that we would like to calculate the probability of observing 3 Heads out of n coin tossing, where $P(H) = p$.
- We have already calculated this probability:

$$P(X = 3) = C_3^n \cdot p^3(1 - p)^{n-3}$$

- We can still answer even a more general question:
  "probability of observing x Heads out of n coin tossing when $P(H) = p$"

$$P(X = x) = C_x^n \cdot p^x(1 - p)^{n-x}$$

Example

Find the following probabilities:

1. Probability of observing 7 Heads out of 12 tossing with $P(H) = 0.7$.
   - Let X : # of Heads in 12 tossing, then

   $$P(X = 7) = C_7^{12} \cdot (0.7)^7(0.3)^5$$

   Alternatively, if we let Y : # of Tails, then we would have to calculate

   $$P(Y = 5) = C_5^{12} \cdot (0.3)^5(0.7)^7$$

2. In a classroom with 35 females and 65 males, what is the probability of having 4 females in a randomly selected group of 10 (with replacement)?
   - Let F : # of females selected in the experiment.
   - Also if we denote the probability of choosing a female by p for each member of the group, then $p = 0.35$ and because of "with replacement" assumption, this probability does not change throughout the experiment.
   - Then we can apply the previous results to answer the question as

   $$P(F = 4) = C_4^{10} \cdot (0.35)^4(0.65)^6$$

   - If the experiment was *without replacement*, would that answer still be valid?

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.
- Finally, **Probability Distribution Function** of the Binomial r.v. X

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.
- Finally, **Probability Distribution Function** of the Binomial r.v. X

$$P(X = x) = C_x^n \cdot p^x (1 - p)^{n-x}$$

  - X : # of successes in n trials.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.
- Finally, **Probability Distribution Function** of the Binomial r.v. X

$$P(X = x) = C_x^n \cdot p^x (1 - p)^{n-x}$$

  - X : # of successes in n trials.
  - P(X=x): probability of x successes in n trials.

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.
- Finally, **Probability Distribution Function** of the Binomial r.v. X

$$P(X = x) = C_x^n \cdot p^x(1-p)^{n-x}$$

  - X : # of successes in n trials.
  - P(X=x): probability of x successes in n trials.
  - x : number of "successes" in the sample, $(x \in \{0, 1, 2, ..., n\})$

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.
- Finally, **Probability Distribution Function** of the Binomial r.v. X

$$P(X = x) = C_x^n \cdot p^x (1 - p)^{n-x}$$

  - X : # of successes in n trials.
  - P(X=x): probability of x successes in n trials.
  - x : number of "successes" in the sample, $(x \in \{0, 1, 2, ..., n\})$
  - n : sample size (number of trials or observations)

Binomial Model

- Let us try to determine the assumptions that needed to be satisfied so that we can simply adapt the results we have obtained so far to different contexts.

Assumptions

Our leading example satisfies the following assumptions:

1. A fixed number of observations in each experiment, denoted by n.
2. Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
3. Constant probability of success for each observation in a given experiment, denoted by p.
4. Observations are independent throughout the experiment.

- In general, we will call an experiment satisfying these four assumptions a **Binomial Model**, and the random variable that is generated by this experiment a **Binomial Random Variable**.
- Binomial model is usually denoted by B(n,p), where n and p are **parameters** of the model.
- Finally, **Probability Distribution Function** of the Binomial r.v. X

$$P(X = x) = C_x^n \cdot p^x (1 - p)^{n-x}$$

  - X : # of successes in n trials.
  - P(X=x): probability of x successes in n trials.
  - x : number of "successes" in the sample, $(x \in \{0, 1, 2, ..., n\})$
  - n : sample size (number of trials or observations)
  - p : probability of "success" on each trial.

Possible Binomial Distribution Settings

- A manufacturing plant labels items as either defective or acceptable

Possible Binomial Distribution Settings

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for contracts will either get a contract or not

Possible Binomial Distribution Settings

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for contracts will either get a contract or not
- A marketing research firm receives survey responses of "yes I will buy" or "no I will not"

Possible Binomial Distribution Settings

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for contracts will either get a contract or not
- A marketing research firm receives survey responses of "yes I will buy" or "no I will not"
- New job applicants either accept the offer or reject it

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
   - Constant probability of success for each observation in a given experiment, denoted by p?

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
    - A fixed number of observations?
      Yes. $n = 15$
    - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
      Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
    - Constant probability of success for each observation in a given experiment, denoted by p?
      Yes. $p = 0.8$

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
   - Constant probability of success for each observation in a given experiment, denoted by p?
     Yes. $p = 0.8$
   - Observations are independent throughout the experiment.

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
   - Constant probability of success for each observation in a given experiment, denoted by p?
     Yes. $p = 0.8$
   - Observations are independent throughout the experiment.
     This is equivalent to ask: If you observe that one airline is late, would you think that it is more likely that other airlines will be late as well?

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
   - Constant probability of success for each observation in a given experiment, denoted by p?
     Yes. $p = 0.8$
   - Observations are independent throughout the experiment.
     This is equivalent to ask: If you observe that one airline is late, would you think that it is more likely that other airlines will be late as well? We will assume that this assumption is also satisfied. Therefore, we have a Binomial model, $B(15, 0.8)$.

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
   - Constant probability of success for each observation in a given experiment, denoted by p?
     Yes. $p = 0.8$
   - Observations are independent throughout the experiment.
     This is equivalent to ask: If you observe that one airline is late, would you think that it is more likely that other airlines will be late as well? We will assume that this assumption is also satisfied. Therefore, we have a Binomial model, $B(15, 0.8)$. But in reality sometimes, it does not have to be! For example, what airlines are late because of the bad weather conditions?

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
   - A fixed number of observations?
     Yes. $n = 15$
   - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
     Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
   - Constant probability of success for each observation in a given experiment, denoted by p?
     Yes. $p = 0.8$
   - Observations are independent throughout the experiment.
     This is equivalent to ask: If you observe that one airline is late, would you think that it is more likely that other airlines will be late as well? We will assume that this assumption is also satisfied. Therefore, we have a Binomial model, $B(15, 0.8)$. But in reality sometimes, it does not have to be! For example, what airlines are late because of the bad weather conditions?

2. What is the probability that three airlines will be late?

Example

You are investigating the punctuality of the airlines in Asia. Your survey tells you that, out of 15 airlines, 80% of them are likely to be late at least once a month.

1. Check whether assumptions of the binomial model are satisfied?
    - A fixed number of observations?
      Yes. $n = 15$
    - Two mutually exclusive and collectively exhaustive categories, customarily called *success* and *failure*.
      Yes. Let us say S = Late at least once, F = Not late. Note that in this case we should define our r.v. as X : # of airlines that are late at least once a month.
    - Constant probability of success for each observation in a given experiment, denoted by p?
      Yes. $p = 0.8$
    - Observations are independent throughout the experiment.
      This is equivalent to ask: If you observe that one airline is late, would you think that it is more likely that other airlines will be late as well? We will assume that this assumption is also satisfied. Therefore, we have a Binomial model, $B(15, 0.8)$. But in reality sometimes, it does not have to be! For example, what airlines are late because of the bad weather conditions?
2. What is the probability that three airlines will be late?
   $P(X = 3) = C_3^{15}(0.8)^3(0.2)^{12}$

3. What is the probability that at most two airlines will be late, $P(X \leq 2) = ?$

3 What is the probability that at most two airlines will be late, $P(X \leq 2) = ?$
$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

3 What is the probability that at most two airlines will be late, $P(X \leq 2) = ?$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$

3. What is the probability that at most two airlines will be late, $P(X \leq 2) =?$
$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$
4. What is the probability that at least three airlines will be late, $P(X \geq 3)$?

3 What is the probability that at most two airlines will be late, $P(X \leq 2) = ?$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$

4 What is the probability that at least three airlines will be late, $P(X \geq 3)$?

Since $\sum_{x \in \{0,1,2,\ldots,15\}} P(X = x) = 1$

3 What is the probability that at most two airlines will be late, $P(X \leq 2) =?$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$

4 What is the probability that at least three airlines will be late, $P(X \geq 3)$?

Since $\sum_{x \in \{0,1,2,\ldots,15\}} P(X = x) = 1$

$$\Longrightarrow P(X \geq 3) = 1 - P(X \leq 2)$$

3 What is the probability that at most two airlines will be late, $P(X \leq 2) =?$

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$

4 What is the probability that at least three airlines will be late, $P(X \geq 3)$?

Since $\sum_{x \in \{0,1,2,\ldots,15\}} P(X=x) = 1$

$$\Longrightarrow P(X \geq 3) = 1 - P(X \leq 2)$$

5 How many airlines will be late in one month on average?

3. What is the probability that at most two airlines will be late, $P(X \leq 2) = ?$
$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$
4. What is the probability that at least three airlines will be late, $P(X \geq 3)$?
Since $\sum_{x \in \{0,1,2,\ldots,15\}} P(X=x) = 1$
$$\implies P(X \geq 3) = 1 - P(X \leq 2)$$
5. How many airlines will be late in one month on average?
6. Can you calculate the variation in the number of airlines that are late?

3. What is the probability that at most two airlines will be late, $P(X \leq 2) =?$
$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$
4. What is the probability that at least three airlines will be late, $P(X \geq 3)$?
Since $\sum_{x \in \{0,1,2,\dots,15\}} P(X=x) = 1$
$$\implies P(X \geq 3) = 1 - P(X \leq 2)$$
5. How many airlines will be late in one month on average?
6. Can you calculate the variation in the number of airlines that are late?

The last two are interesting questions but we need two more concept in order to think about these type of questions:

- Expected Value of a r.v.

3. What is the probability that at most two airlines will be late, $P(X \leq 2) = ?$
$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$
$$= C_0^{15}(0.8)^0(0.2)^{15} + C_1^{15}(0.8)^1(0.2)^{14} + C_2^{15}(0.8)^2(0.2)^{13}$$
4. What is the probability that at least three airlines will be late, $P(X \geq 3)$?
Since $\sum_{x \in \{0,1,2,\ldots,15\}} P(X=x) = 1$
$$\Longrightarrow P(X \geq 3) = 1 - P(X \leq 2)$$
5. How many airlines will be late in one month on average?
6. Can you calculate the variation in the number of airlines that are late?

The last two are interesting questions but we need two more concept in order to think about these type of questions:

- Expected Value of a r.v.
- Variance (or std) of a r.v.

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Notation: $\mu = E(X)$

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Notation: $\mu = E(X)$

Variance of a discrete random variable X, with probability distribution function $P(x)$ is given by

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Notation: $\mu = E(X)$

Variance of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$V(X) = \sum_{i=1}^{n} \left(x_i - \mu\right)^2 \cdot P(X = x_i)$$

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Notation: $\mu = E(X)$

Variance of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(X = x_i)$$

Notation: $\sigma^2 = V(X)$

Expected Value and Variance of a Discrete Random Variable

Expected Value of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Notation: $\mu = E(X)$

Variance of a discrete random variable X, with probability distribution function $P(x)$ is given by

$$V(X) = \sum_{i=1}^{n} \left(x_i - \mu\right)^2 \cdot P(X = x_i)$$

Notation: $\sigma^2 = V(X)$

Standard Deviation of X is defined as the square root of the variance, therefore with this notation it will be simply σ.

Example

Toss 2 coins, and let

$$X = \#ofHeads$$

compute μ and σ?

Example

Toss 2 coins, and let

$$X = \#ofHeads$$

compute μ and σ?

- Let us first construct the probability distribution table

Example

Toss 2 coins, and let

$$X = \#ofHeads$$

compute μ and σ?

- Let us first construct the probability distribution table

X	$P(X)$
0	.25
1	0.50
2	0.25

Example

Toss 2 coins, and let

$$X = \#ofHeads$$

compute μ and σ?

- Let us first construct the probability distribution table

X	$P(X)$
0	.25
1	0.50
2	0.25

- Expected Value of X

$$\mu = E(X) = \sum_{i=1}^{3} x_i \cdot P(X = x_i) = 0 \cdot (0.25) + 1 \cdot (0.50) + 2 \cdot (0.25) = 1$$

Example

Toss 2 coins, and let

$$X = \#ofHeads$$

compute μ and σ?

- Let us first construct the probability distribution table

X	$P(X)$
0	.25
1	0.50
2	0.25

- Expected Value of X

$$\mu = E(X) = \sum_{i=1}^{3} x_i \cdot P(X = x_i) = 0 \cdot (0.25) + 1 \cdot (0.50) + 2 \cdot (0.25) = 1$$

- Variance of X

$$V(X) = \sum_{i=1}^{3} (x_i - \mu)^2 \cdot P(X = x_i)$$
$$= (0-1)^2 \cdot (0.25) + (1-1)^2 \cdot (0.50) + (2-1)^2 \cdot (0.25) = 0.5$$

Therefore, standard deviation of X is

$$\sigma = \sqrt{V(X)} = \sqrt{0.5} = 0.707$$

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1 - p)$$

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n,p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1-p)$$

- Sdv of X

$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1-p)}$$

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

- Expected Value of X
$$\mu = E(X) = n \cdot p$$
- Variance of X
$$\sigma^2 = V(X) = n \cdot p \cdot (1 - p)$$
- Sdv of X
$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1 - p)}$$

Now, we revisit our airline example, which was a Binomial model with parameters, $n = 15$, and $p = 0.8$, i.e. $B(15, 0.8)$.

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1 - p)$$

- Sdv of X

$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1 - p)}$$

Now, we revisit our airline example, which was a Binomial model with parameters, $n = 15$, and $p = 0.8$, i.e. $B(15, 0.8)$.

5. How many airlines will be late in one month on average?

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1 - p)$$

- Sdv of X

$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1 - p)}$$

Now, we revisit our airline example, which was a Binomial model with parameters, $n = 15$, and $p = 0.8$, i.e. $B(15, 0.8)$.

5 How many airlines will be late in one month on average?

$$\mu = E(X) = n \cdot p = 15 \cdot (0.8) = 12$$

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n,p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1-p)$$

- Sdv of X

$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1-p)}$$

Now, we revisit our airline example, which was a Binomial model with parameters, $n = 15$, and $p = 0.8$, i.e. $B(15, 0.8)$.

5. How many airlines will be late in one month on average?

$$\mu = E(X) = n \cdot p = 15 \cdot (0.8) = 12$$

6. Can you calculate the variation in the number of airlines that are late?

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n, p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1 - p)$$

- Sdv of X

$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1 - p)}$$

Now, we revisit our airline example, which was a Binomial model with parameters, $n = 15$, and $p = 0.8$, i.e. $B(15, 0.8)$.

5. How many airlines will be late in one month on average?

$$\mu = E(X) = n \cdot p = 15 \cdot (0.8) = 12$$

6. Can you calculate the variation in the number of airlines that are late?

$$\sigma^2 = V(X) = n \cdot p \cdot (1 - p) = 15 \cdot (0.8) \cdot (0.2) = 2.4$$

Expected Value and Variance of a Binomial Random Variable

If X is a Binomial r.v. with parameters n and p (i.e. $B(n,p)$)

- Expected Value of X

$$\mu = E(X) = n \cdot p$$

- Variance of X

$$\sigma^2 = V(X) = n \cdot p \cdot (1-p)$$

- Sdv of X

$$\sigma = \sqrt{V(X)} = \sqrt{n \cdot p \cdot (1-p)}$$

Now, we revisit our airline example, which was a Binomial model with parameters, $n = 15$, and $p = 0.8$, i.e. $B(15, 0.8)$.

5. How many airlines will be late in one month on average?

$$\mu = E(X) = n \cdot p = 15 \cdot (0.8) = 12$$

6. Can you calculate the variation in the number of airlines that are late?

$$\sigma^2 = V(X) = n \cdot p \cdot (1-p) = 15 \cdot (0.8) \cdot (0.2) = 2.4$$

Therefore, $\sigma = \sqrt{2.4}$

A contractor estimates the probabilities for the number of days required to complete a certain type of construction project as follows:

Time (days)	1	2	3	4	5
Probability	0.05	0.2	0.35	0.3	0.1

1. What is the probability that a randomly chosen project takes less than 3 days to complete?
2. Find the expected time to complete the project.
3. Find the standard deviation of time required to complete the project.
4. Suppose the contractor's project cost is made of two parts: a fixed cost of $20,000 and $2,000 for each day taken to complete the project. Find the mean and standard deviation of total project cost.
5. If three projects are undertaken, what is the probability that at least two of them will take at least 4 days to complete? Assume that project completion times are independent.

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.

The Practice of Modeling

- So far we have studied two problems
  - $\#$ Heads in 5 coin toss
  - $\#$ of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.
  - Study each class separately and see what kind of results are based on just these distinguishing characteristics $\rightarrow$ call them **Models**.

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.
  - Study each class separately and see what kind of results are based on just these distinguishing characteristics $\rightarrow$ call them **Models**.
  - Then, each time you encounter a new problem, check whether you can fit the problem into one of the Models you have already studied.

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.
  - Study each class separately and see what kind of results are based on just these distinguishing characteristics $\rightarrow$ call them **Models**.
  - Then, each time you encounter a new problem, check whether you can fit the problem into one of the Models you have already studied.
  - If yes, then try to apply the results obtained for that Model to the problem at hand.

The Practice of Modeling

- So far we have studied two problems
  - $\#$ Heads in 5 coin toss
  - $\#$ of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.
  - Study each class separately and see what kind of results are based on just these distinguishing characteristics $\rightarrow$ call them **Models**.
  - Then, each time you encounter a new problem, check whether you can fit the problem into one of the Models you have already studied.
  - If yes, then try to apply the results obtained for that Model to the problem at hand.
- So far we have seen only one of these classes and summarized our results under the name Binomial Model.

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.
  - Study each class separately and see what kind of results are based on just these distinguishing characteristics $\rightarrow$ call them **Models**.
  - Then, each time you encounter a new problem, check whether you can fit the problem into one of the Models you have already studied.
  - If yes, then try to apply the results obtained for that Model to the problem at hand.
- So far we have seen only one of these classes and summarized our results under the name Binomial Model.
- However, there are some important problems that do not fit into the Binomial Model.

The Practice of Modeling

- So far we have studied two problems
  - # Heads in 5 coin toss
  - # of airlines out of 15 airlines that are late at least once a month
- And we were able to analyze these two seemingly different problems in the same framework .
- Because in terms of the essential characteristics, they were of the same type, i.e. both satisfied the four assumptions above, so we have studied both of them under the same Model.
- In general we,
  - Classify most frequently encountered problems in terms of their essential characteristics.
  - Study each class separately and see what kind of results are based on just these distinguishing characteristics $\rightarrow$ call them **Models**.
  - Then, each time you encounter a new problem, check whether you can fit the problem into one of the Models you have already studied.
  - If yes, then try to apply the results obtained for that Model to the problem at hand.
- So far we have seen only one of these classes and summarized our results under the name Binomial Model.
- However, there are some important problems that do not fit into the Binomial Model.
- Next, we are going to study another important class of problems.

The Poisson Model

- Consider the following problem:

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  \- **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  \- **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  $\rightarrow$ We can answer this question by using the Binomial model.

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  \- **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  → We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  \- **Q:** *What is the probability that 5 females will enter the store in the next second?*

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  - **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  $\rightarrow$ We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  - **Q:** *What is the probability that 5 females will enter the store in the next second?*

  $\rightarrow$ This can not be answered by the Binomial model. Why?

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  \- **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  → We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  \- **Q:** *What is the probability that 5 females will enter the store in the next second?*

  → This can not be answered by the Binomial model. Why?
- This is just one example from the type of problems that we are going to study under the **Poisson Model**. Some other examples:

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  - **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  $\rightarrow$ We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  - **Q:** *What is the probability that 5 females will enter the store in the next second?*

  $\rightarrow$ This can not be answered by the Binomial model. Why?
- This is just one example from the type of problems that we are going to study under the **Poisson Model**. Some other examples:
  - The number of failures in a large computer system during a given day.

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  - **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  $\rightarrow$ We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  - **Q:** *What is the probability that 5 females will enter the store in the next second?*

  $\rightarrow$ This can not be answered by the Binomial model. Why?
- This is just one example from the type of problems that we are going to study under the **Poisson Model**. Some other examples:
  - The number of failures in a large computer system during a given day.
  - The number of customers to arrive for flights during each 10-minute time interval.

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  - **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  $\rightarrow$ We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  - **Q:** *What is the probability that 5 females will enter the store in the next second?*

  $\rightarrow$ This can not be answered by the Binomial model. Why?
- This is just one example from the type of problems that we are going to study under the **Poisson Model**. Some other examples:
  - The number of failures in a large computer system during a given day.
  - The number of customers to arrive for flights during each 10-minute time interval.
  - The number of delivery trucks to arrive at a central warehouse in an hour.

The Poisson Model

- Consider the following problem:
  From the past data provided by Whole Foods (Union Square, NY), you know that 60% of shoppers are female.
  - **Q:** *What is the probability that 3 of the next 10 shoppers will be female.*

  $\rightarrow$ We can answer this question by using the Binomial model.
- Now, consider the following version:
  From the past data provided by Whole Foods (Union Square, NY), you know that 3 females enter the store per second, on average.
  - **Q:** *What is the probability that 5 females will enter the store in the next second?*

  $\rightarrow$ This can not be answered by the Binomial model. Why?
- This is just one example from the type of problems that we are going to study under the **Poisson Model**. Some other examples:
  - The number of failures in a large computer system during a given day.
  - The number of customers to arrive for flights during each 10-minute time interval.
  - The number of delivery trucks to arrive at a central warehouse in an hour.
- Sometimes a problem can be analyzed by both Binomial and Poisson Models, but in this course we will only see the problems that just one of them fits the situation best.

The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences (or successes) of a certain event of interest in a given continuous interval (such as time, surface area, or length).

The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences (or successes) of a certain event of interest in a given continuous interval (such as time, surface area, or length).

The most important **distinguishing characteristics** of the Poisson model *(and how they appeared in Whole Foods example are in italics)*:

The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences (or successes) of a certain event of interest in a given continuous interval (such as time, surface area, or length).

The most important **distinguishing characteristics** of the Poisson model *(and how they appeared in Whole Foods example are in italics)*:

- Data consists of a very large number of observations over a continuous variable like time or area. You can imagine the data as a continuously moving/expanding sequence of observations over time *(i.e. shoppers arriving at Whole Foods forms a continuously expanding sequence)*.

The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences (or successes) of a certain event of interest in a given continuous interval (such as time, surface area, or length).

The most important **distinguishing characteristics** of the Poisson model *(and how they appeared in Whole Foods example are in italics)*:

- Data consists of a very large number of observations over a continuous variable like time or area. You can imagine the data as a continuously moving/expanding sequence of observations over time *(i.e. shoppers arriving at Whole Foods forms a continuously expanding sequence)*.
- You can count how many times the event of interest occurs on average in a certain time interval *(i.e. may be from the data consisting of the number of females per day, we concluded that on average 3 females arrive per second on average)*.

The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences (or successes) of a certain event of interest in a given continuous interval (such as time, surface area, or length).

The most important **distinguishing characteristics** of the Poisson model *(and how they appeared in Whole Foods example are in italics)*:

- Data consists of a very large number of observations over a continuous variable like time or area. You can imagine the data as a continuously moving/expanding sequence of observations over time *(i.e. shoppers arriving at Whole Foods forms a continuously expanding sequence)*.
- You can count how many times the event of interest occurs on average in a certain time interval *(i.e. may be from the data consisting of the number of females per day, we concluded that on average 3 females arrive per second on average)*.
  $\rightarrow$ Alternatively, imagine that you take a snapshot from this long sequence at random *(i.e. wait at the door one second that randomly determined during the day)* and count number of occurrences of the event of interest in the snapshot *(i.e. 3 females counted)*.

The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences (or successes) of a certain event of interest in a given continuous interval (such as time, surface area, or length).

The most important **distinguishing characteristics** of the Poisson model *(and how they appeared in Whole Foods example are in italics)*:

- Data consists of a very large number of observations over a continuous variable like time or area. You can imagine the data as a continuously moving/expanding sequence of observations over time *(i.e. shoppers arriving at Whole Foods forms a continuously expanding sequence)*.
- You can count how many times the event of interest occurs on average in a certain time interval *(i.e. may be from the data consisting of the number of females per day, we concluded that on average 3 females arrive per second on average)*.
  $\rightarrow$ Alternatively, imagine that you take a snapshot from this long sequence at random *(i.e. wait at the door one second that randomly determined during the day)* and count number of occurrences of the event of interest in the snapshot *(i.e. 3 females counted)*.
- Based on this snapshot data, then you would like to calculate various probabilities regarding the number of occurrences of the event of interest *(i.e. the probability that 5 females will enter in the very next second)*.

The Poisson Distribution Function

To introduce the Poisson model, let us

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

- And denote the average number of occurrences of the event of interest per unit by λ.

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

- And denote the average number of occurrences of the event of interest per unit by λ.
- Then the Poisson model enables us to calculate the probability of observing x number of occurrences per unit, $P(X = x)$, for any possible x through the following **Poisson Distribution Function**

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

- And denote the average number of occurrences of the event of interest per unit by λ.
- Then the Poisson model enables us to calculate the probability of observing x number of occurrences per unit, $P(X = x)$, for any possible x through the following **Poisson Distribution Function**

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad (e = 2.71828...)$$

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

- And denote the average number of occurrences of the event of interest per unit by λ.
- Then the Poisson model enables us to calculate the probability of observing x number of occurrences per unit, $P(X = x)$, for any possible x through the following **Poisson Distribution Function**

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad (e = 2.71828...)$$

- In the Whole Foods example, we know that on average 3 females arrive per second ($\lambda = 3$ per second) and we want to calculate the probability of observing 4 females in the next second (i.e. $X =$ number of females arrive per second, and want to compute P(X=4)=?).

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

- And denote the average number of occurrences of the event of interest per unit by λ.
- Then the Poisson model enables us to calculate the probability of observing x number of occurrences per unit, $P(X = x)$, for any possible x through the following **Poisson Distribution Function**

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad (e = 2.71828...)$$

- In the Whole Foods example, we know that on average 3 females arrive per second ($\lambda = 3$ per second) and we want to calculate the probability of observing 4 females in the next second (i.e. $X =$ number of females arrive per second, and want to compute P(X=4)=?).
- This is a very typical property of the Poisson problems: we know the average number of occurrences of a particular event per unit, and we would like to compute probabilities for specific number of occurrences per unit.

The Poisson Distribution Function

To introduce the Poisson model, let us

- Define a random variable as

$$X = \text{ number of occurrences per unit}$$

- And denote the average number of occurrences of the event of interest per unit by λ.
- Then the Poisson model enables us to calculate the probability of observing x number of occurrences per unit, $P(X = x)$, for any possible x through the following **Poisson Distribution Function**

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad (e = 2.71828...)$$

- In the Whole Foods example, we know that on average 3 females arrive per second ($\lambda = 3$ per second) and we want to calculate the probability of observing 4 females in the next second (i.e. $X =$ number of females arrive per second, and want to compute P(X=4)=?).
- This is a very typical property of the Poisson problems: we know the average number of occurrences of a particular event per unit, and we would like to compute probabilities for specific number of occurrences per unit.
- Note that X and λ should be expressed in term of the same unit. Next problem will clarify the importance of this point further...

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day.

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then

$$P(X = 0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then
   $$P(X=0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

2. What is the probability of two or more failures in a given day?

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then

$$P(X=0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

2. What is the probability of two or more failures in a given day?

$$\begin{aligned} P(X \geq 2) =& 1 - P(X=0) - P(X=1) \\ =& 1 - \frac{e^{-0.03}(0.03)^0}{0!} - \frac{e^{-0.03}(0.03)^1}{1!} = 1 - 1.03e^{-0.03} \end{aligned}$$

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then

   $$P(X=0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

2. What is the probability of two or more failures in a given day?

   $$\begin{aligned} P(X \geq 2) =& 1 - P(X=0) - P(X=1) \\ =& 1 - \frac{e^{-0.03}(0.03)^0}{0!} - \frac{e^{-0.03}(0.03)^1}{1!} = 1 - 1.03e^{-0.03} \end{aligned}$$

3. What is the probability of two failures in a 3-day period?

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then
   $$P(X=0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

2. What is the probability of two or more failures in a given day?
   $$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - \frac{e^{-0.03}(0.03)^0}{0!} - \frac{e^{-0.03}(0.03)^1}{1!} = 1 - 1.03e^{-0.03} \end{aligned}$$

3. What is the probability of two failures in a 3-day period?

   Note: $P(X=2) = \frac{e^{-0.09}(0.09)^2}{2!}$ would not be the correct answer. Because this is the probability of observing 2 failures **per day** but we would like to compute the probability of observing 2 failures **per 3-days**. Therefore, we need a new pair of λ and X.

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then
   $$P(X=0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

2. What is the probability of two or more failures in a given day?
   $$\begin{aligned} P(X \geq 2) =& 1 - P(X=0) - P(X=1) \\ =& 1 - \frac{e^{-0.03}(0.03)^0}{0!} - \frac{e^{-0.03}(0.03)^1}{1!} = 1 - 1.03e^{-0.03} \end{aligned}$$

3. What is the probability of two failures in a 3-day period?

   Note: $P(X=2) = \frac{e^{-0.09}(0.09)^2}{2!}$ would not be the correct answer. Because this is the probability of observing 2 failures **per day** but we would like to compute the probability of observing 2 failures **per 3-days**. Therefore, we need a new pair of λ and X.

   First calculate $\lambda = 0.09$ per 3-days and define a new r.v. as $X = \#$ of failures per 3-days

Example

A computer manager reports that his computer system experienced three component failures during the past 100 days.

1. What is the probability of no failures in a given day?

   $X = \#$ of failures per day
   $\lambda = 0.03$ per one day. Then
   $$P(X=0) = \frac{e^{-0.03}(0.03)^0}{0!} = e^{-0.03}$$

2. What is the probability of two or more failures in a given day?
   $$\begin{aligned} P(X \geq 2) =& 1 - P(X=0) - P(X=1) \\ =& 1 - \frac{e^{-0.03}(0.03)^0}{0!} - \frac{e^{-0.03}(0.03)^1}{1!} = 1 - 1.03e^{-0.03} \end{aligned}$$

3. What is the probability of two failures in a 3-day period?

   Note: $P(X=2) = \frac{e^{-0.09}(0.09)^2}{2!}$ would not be the correct answer. Because this is the probability of observing 2 failures **per day** but we would like to compute the probability of observing 2 failures **per 3-days**. Therefore, we need a new pair of λ and X.

   First calculate $\lambda = 0.09$ per 3-days and define a new r.v. as $X = \#$ of failures per 3-days then $P(X=2) = \frac{e^{-0.09}(0.09)^2}{2!}$

Linear Functions of A Random Variable

- Suppose X is a discrete random variable with some prob. dist. function $P(x)$.

Linear Functions of A Random Variable

- Suppose X is a discrete random variable with some prob. dist. function $P(x)$.
- In this section, we will see how we can generate new r.v.'s from X and how the expected value and variance of these new r.v.'s are related to μ_X and σ_X^2.

Linear Functions of A Random Variable

- Suppose X is a discrete random variable with some prob. dist. function $P(x)$.
- In this section, we will see how we can generate new r.v.'s from X and how the expected value and variance of these new r.v.'s are related to μ_X and σ_X^2.
- Let us begin with the following

$$Y = a + bX$$

where a and b are constants.

Linear Functions of A Random Variable

- Suppose X is a discrete random variable with some prob. dist. function $P(x)$.
- In this section, we will see how we can generate new r.v.'s from X and how the expected value and variance of these new r.v.'s are related to μ_X and σ_X^2.
- Let us begin with the following

$$Y = a + bX$$

  where a and b are constants.
- The most important observation: Y is a r.v. too and $P(y)$ is completely determined by $P(x)$.

Example

- Suppose our original rv X has the following probability distribution

X	-2	1	2
P(X)	.5	.3	.2

Example

- Suppose our original rv X has the following probability distribution

X	-2	1	2
P(X)	.5	.3	.2

- And let us define a new random variable based on X as

$$Y = 200 + 30X$$

Example

- Suppose our original rv X has the following probability distribution

X	-2	1	2
P(X)	.5	.3	.2

- And let us define a new random variable based on X as

$$Y = 200 + 30X$$

- In order to construct $P(y)$, let us start from $X = -2$.

Example

- Suppose our original rv X has the following probability distribution

X	-2	1	2
P(X)	.5	.3	.2

- And let us define a new random variable based on X as

$$Y = 200 + 30X$$

- In order to construct $P(y)$, let us start from $X = -2$. For $X = -2$
  - $Y = 200 + 30(-2) = 140$
  - and $P(Y = 140)$ is exactly equal to $P(X = -2)$ because there is no other x value for which $Y = 140$, so $P(Y = 140) = 0.5$.

Example

- Suppose our original rv X has the following probability distribution

X	-2	1	2
P(X)	.5	.3	.2

- And let us define a new random variable based on X as

$$Y = 200 + 30X$$

- In order to construct $P(y)$, let us start from $X = -2$. For $X = -2$
  - $Y = 200 + 30(-2) = 140$
  - and $P(Y = 140)$ is exactly equal to $P(X = -2)$ because there is no other x value for which $Y = 140$, so $P(Y = 140) = 0.5$.
- By doing the same thing for $X = 1$ and $X = 2$, we obtain $P(y)$ as

Y	140	230	260
P(Y)	.5	.3	.2

- To really appreciate the role of linearity, consider the same r.v. X

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- But now define another r.v., **a nonlinear r.v.**, as

$$Y_{nl} = X^2 \qquad (nl \text{ for nonlinear})$$

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- But now define another r.v., **a nonlinear r.v.**, as

$$Y_{nl} = X^2 \qquad (nl \text{ for nonlinear})$$

- Note that both for $X = -2$ and $X = 2$, we have the same Y_{nl} value, i.e. $Y_{nl} = 4$, so

$$P(Y_{nl} = 4) = P(X = -2 \text{ or } X = 2) = 0.7$$

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- But now define another r.v., **a nonlinear r.v.**, as

$$Y_{nl} = X^2 \qquad (nl \text{ for nonlinear})$$

- Note that both for $X = -2$ and $X = 2$, we have the same Y_{nl} value, i.e. $Y_{nl} = 4$, so

$$P(Y_{nl} = 4) = P(X = -2 \text{ or } X = 2) = 0.7$$

- Therefore, the prob. dist. function of Y_{nl} is

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- But now define another r.v., **a nonlinear r.v.**, as

$$Y_{nl} = X^2 \qquad (nl \text{ for nonlinear})$$

- Note that both for $X = -2$ and $X = 2$, we have the same Y_{nl} value, i.e. $Y_{nl} = 4$, so

$$P(Y_{nl} = 4) = P(X = -2 \text{ or } X = 2) = 0.7$$

- Therefore, the prob. dist. function of Y_{nl} is

Y_{nl}	1	4
$P(Y_{nl})$	.3	.7

- Compare the probability distribution of Y and Y_{nl} and make sure that you have understood the role of linearity and how we compute the probability distribution of linear or nonlinear random variables.

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- But now define another r.v., **a nonlinear r.v.**, as

$$Y_{nl} = X^2 \qquad (nl \text{ for nonlinear})$$

- Note that both for $X = -2$ and $X = 2$, we have the same Y_{nl} value, i.e. $Y_{nl} = 4$, so

$$P(Y_{nl} = 4) = P(X = -2 \text{ or } X = 2) = 0.7$$

- Therefore, the prob. dist. function of Y_{nl} is

Y_{nl}	1	4
$P(Y_{nl})$	.3	.7

- Compare the probability distribution of Y and Y_{nl} and make sure that you have understood the role of linearity and how we compute the probability distribution of linear or nonlinear random variables.
- As an exercise, construct the probability distribution tables for the following random variables.

- To really appreciate the role of linearity, consider the same r.v. X

X	-2	1	2
P(X)	.5	.3	.2

- But now define another r.v., **a nonlinear r.v.**, as

$$Y_{nl} = X^2 \qquad (nl \text{ for nonlinear})$$

- Note that both for $X = -2$ and $X = 2$, we have the same Y_{nl} value, i.e. $Y_{nl} = 4$, so

$$P(Y_{nl} = 4) = P(X = -2 \text{ or } X = 2) = 0.7$$

- Therefore, the prob. dist. function of Y_{nl} is

Y_{nl}	1	4
$P(Y_{nl})$	.3	.7

- Compare the probability distribution of Y and Y_{nl} and make sure that you have understood the role of linearity and how we compute the probability distribution of linear or nonlinear random variables.
- As an exercise, construct the probability distribution tables for the following random variables.
  1. $Y_1 = 3 - 7X$
  2. $Y_2 = X^2 + X - 2$

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s
  - $E(X) = \sum_{x \in \{-2,1,2\}} xP(X = x) = (-2)(.5) + (1)(.3) + (2)(.2) = -0.3$

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s
  - $E(X) = \sum_{x \in \{-2,1,2\}} xP(X = x) = (-2)(.5) + (1)(.3) + (2)(.2) = -0.3$
  - $E(Y) = \sum_{y \in \{140,230,260\}} yP(Y = y) = 140(.5) + 230(.3) + 260(.2) = 191$

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s
  - $E(X) = \sum_{x \in \{-2,1,2\}} xP(X = x) = (-2)(.5) + (1)(.3) + (2)(.2) = -0.3$
  - $E(Y) = \sum_{y \in \{140,230,260\}} yP(Y = y) = 140(.5) + 230(.3) + 260(.2) = 191$
  - $E(Y_{nl}) = \sum_{y_{nl} \in \{1,4\}} y_{nl} P(Y_{nl} = y_{nl}) = 1(.3) + 4(.7) = 3.1$

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s
  - $E(X) = \sum_{x \in \{-2,1,2\}} xP(X = x) = (-2)(.5) + (1)(.3) + (2)(.2) = -0.3$
  - $E(Y) = \sum_{y \in \{140,230,260\}} yP(Y = y) = 140(.5) + 230(.3) + 260(.2) = 191$
  - $E(Y_{nl}) = \sum_{y_{nl} \in \{1,4\}} y_{nl}P(Y_{nl} = y_{nl}) = 1(.3) + 4(.7) = 3.1$
- Now notice the relationship between $E(X)$ and $E(Y)$:

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s
  - $E(X) = \sum_{x \in \{-2,1,2\}} xP(X = x) = (-2)(.5) + (1)(.3) + (2)(.2) = -0.3$
  - $E(Y) = \sum_{y \in \{140,230,260\}} yP(Y = y) = 140(.5) + 230(.3) + 260(.2) = 191$
  - $E(Y_{nl}) = \sum_{y_{nl} \in \{1,4\}} y_{nl} P(Y_{nl} = y_{nl}) = 1(.3) + 4(.7) = 3.1$
- Now notice the relationship between $E(X)$ and $E(Y)$:

$$E(Y) = 200 + 30E(X) = 200 + 30(-0.3) = 191$$

- In words, once we know or have already calculated $E(X)$, we can use this to calculate $E(Y)$ **when** Y is a linear function of X. Let us express this important result for later reference.

Expected Values of Linear Random Variables

- Whenever we have a random variable we can calculate its expected value and variance, no matter how complicated the r.v. is.
- In this section, we have seen three r.v.'s so far: X, Y and Y_{nl}.
- Let us compute the expected value for each of these r.v.'s
  - $E(X) = \sum_{x \in \{-2,1,2\}} xP(X = x) = (-2)(.5) + (1)(.3) + (2)(.2) = -0.3$
  - $E(Y) = \sum_{y \in \{140,230,260\}} yP(Y = y) = 140(.5) + 230(.3) + 260(.2) = 191$
  - $E(Y_{nl}) = \sum_{y_{nl} \in \{1,4\}} y_{nl}P(Y_{nl} = y_{nl}) = 1(.3) + 4(.7) = 3.1$
- Now notice the relationship between $E(X)$ and $E(Y)$:

$$E(Y) = 200 + 30E(X) = 200 + 30(-0.3) = 191$$

- In words, once we know or have already calculated $E(X)$, we can use this to calculate $E(Y)$ **when** Y is a linear function of X. Let us express this important result for later reference.

Result 1

Let X be r.v. with $\mu_x = E(X)$, then for $Y = a + bX$ we have

$$E(Y) = a + bE(X) \quad \text{or} \quad \mu_y = a + b\mu_x$$

Variances of Linear Random Variables

- Now, let us calculate the variances for each of the three r.v.'s

$$\begin{aligned}\sigma_X^2 &= \sum_{x\in\{-2,1,2\}} (x-\mu_x)^2 P(X=x)\\ &= (-2+0.3)^2(.5)+(1+0.3)^2(0.3)+(2+0.3)^2(0.2)=3.01\\ \sigma_Y^2 &= \sum_{y\in\{140,230,260\}} (y-\mu_y)^2 P(Y=y)\\ &= (140-191)^2(.5)+(230-191)^2(0.3)+(260-191)^2(0.2)=2709\\ \sigma_{Y_{nl}}^2 &= \sum_{y_{nl}\in\{1,4\}} (y_{nl}-\mu_{y_{nl}})^2 P(Y_{nl}=y_{nl})\\ &= (1-3.1)^2(.3)+(4-3.1)^2(0.7)=1.89\end{aligned}$$

- Now notice the relationship between σ_X^2 and σ_Y^2:

$$\sigma_Y^2 = 30^2\sigma_X^2 = 900(3.01) = 2709$$

- In words, again once we know or have already calculated σ_X^2, we can use this to calculate σ_Y^2 **when** Y is a linear function of X. Also note that the constant term a drops when we talk about the variance.

Result 2

Let X be r.v. with $Var(X)=\sigma_x^2$, then for $Y=a+bX$ we have

$$Var(Y)=b^2Var(X) \quad \text{or} \quad \sigma_y^2=b^2\sigma_x^2 \quad \text{and also} \quad \sigma_y=|b|\sigma_x$$

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.
- But sometimes it is more useful and/or natural to study two random variables simultaneously. Consider the following example:

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.
- But sometimes it is more useful and/or natural to study two random variables simultaneously. Consider the following example:
- Suppose you are part of a team studying the relationship between smoking habits and cancer.

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.
- But sometimes it is more useful and/or natural to study two random variables simultaneously. Consider the following example:
- Suppose you are part of a team studying the relationship between smoking habits and cancer.
- As a part of the study you surveyed a randomly chosen 1000 males between ages 60-65 across the country.

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.
- But sometimes it is more useful and/or natural to study two random variables simultaneously. Consider the following example:
- Suppose you are part of a team studying the relationship between smoking habits and cancer.
- As a part of the study you surveyed a randomly chosen 1000 males between ages 60-65 across the country.
- In the survey, first you ask whether they smoke and if they smoke, the follow up question is to ask they are light or heavy smokers. Finally, you ask whether they have been diagnosed with any type of cancer.

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.
- But sometimes it is more useful and/or natural to study two random variables simultaneously. Consider the following example:
- Suppose you are part of a team studying the relationship between smoking habits and cancer.
- As a part of the study you surveyed a randomly chosen 1000 males between ages 60-65 across the country.
- In the survey, first you ask whether they smoke and if they smoke, the follow up question is to ask they are light or heavy smokers. Finally, you ask whether they have been diagnosed with any type of cancer.
- Note that each observation will fall into only one of the cells in the following table

Jointly Distributed Discrete Random Variables

- So far we have dealt only with a single r.v. at a time. At this point it is useful to recognize that even when we generate a new r.v. Y from X, we still we work with a single r.v. at a time.
- But sometimes it is more useful and/or natural to study two random variables simultaneously. Consider the following example:
- Suppose you are part of a team studying the relationship between smoking habits and cancer.
- As a part of the study you surveyed a randomly chosen 1000 males between ages 60-65 across the country.
- In the survey, first you ask whether they smoke and if they smoke, the follow up question is to ask they are light or heavy smokers. Finally, you ask whether they have been diagnosed with any type of cancer.
- Note that each observation will fall into only one of the cells in the following table

Smoking / Cancer	Yes	No
Non-smoker	.	.
Light-smoker	.	.
Heavy-smoker	.	.

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...
- How can we think of this problem in terms of random variables?

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...
- How can we think of this problem in terms of random variables?
- To do this let us define the following two random variables

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...
- How can we think of this problem in terms of random variables?
- To do this let us define the following two random variables
  - C= Cancer history, coded with $C = 0$ for Yes and $C = 1$ for No.

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...
- How can we think of this problem in terms of random variables?
- To do this let us define the following two random variables
  - C= Cancer history, coded with $C = 0$ for Yes and $C = 1$ for No.
  - S= Smoking habit, coded with $S = 0$ for non-smokers, $S = 1$ for light-smokers and $S = 2$ for heavy-smokers.

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...
- How can we think of this problem in terms of random variables?
- To do this let us define the following two random variables
  - C= Cancer history, coded with $C = 0$ for Yes and $C = 1$ for No.
  - S= Smoking habit, coded with $S = 0$ for non-smokers, $S = 1$ for light-smokers and $S = 2$ for heavy-smokers.
- Now we can represent the same table in terms of these r.v.'s and probabilities

Jointly Distributed Discrete Random Variables

- Suppose your survey have produced the following data

Smoking / Cancer	Yes	No
Non-smoker	20	580
Light-smoker	35	215
Heavy-smoker	45	105

- Of course, once we have this table we can answer many questions regarding smoking habits and cancer tendency in this age group. For example, we can compute the probability of randomly chosen male in that age range to be non-smoker, light-smoker and cancer, etc...
- How can we think of this problem in terms of random variables?
- To do this let us define the following two random variables
  - C= Cancer history, coded with $C = 0$ for Yes and $C = 1$ for No.
  - S= Smoking habit, coded with $S = 0$ for non-smokers, $S = 1$ for light-smokers and $S = 2$ for heavy-smokers.
- Now we can represent the same table in terms of these r.v.'s and probabilities

$S(row)/C$	0	1
0	.020	.580
1	.035	.215
2	.045	.105

$S(row)/C$	0	1
0	.020	.580
1	.035	.215
2	.045	.105

- At this point, make sure that you understand why S and C are r.v.'s.
- For example, $P(S = 1, C = 0) = 0.035$ is the probability that a randomly chosen male of that age group is a light-smoker **and** diagnosed with cancer.
- Instead of writing $P(S = 1, C = 0) = 0.035$, you can simply write $P(1, 0) = 0.035$ as long as it is clear that the first component is the value of S and the second component is that of C.
- With this data what else we can do? For example,
  - Probability of randomly selected male to be a non-smoker?

    $$P(S = 0) = P(S = 0, C = 0) + P(S = 0, C = 1) = 0.020 + 0.580 = 0.6$$

    i.e. at $S = 0$ sum over C. We are going to call this the **marginal probabilities** of S.
  - Similarly, we can compute $P(S = 1) = 0.25$ and $P(S = 2) = 0.15$, and then construct **marginal probability distribution** table for S as

S	0	1	2
P(S=s)	.60	.25	.15

  - Since we have already obtained $P(s)$, $E(S)$ and $V(S)$ can be easily computed as usual. [Verify that $E(S) = .55$ and $V(S) = .366$.]

$S(row)/C$	0	1
0	.020	.580
1	.035	.215
2	.045	.105

- With this data what else we can do? (Cont'd)
  - Probability of a randomly selected male having experienced cancer given that the person is a non-smoker?

$$P(C=0|S=0) = \frac{P(S=0, C=0)}{P(S=0)} = \frac{0.02}{0.60} = 0.033$$

  Similarly, we can also calculate $P(C=1|S=0)$. These two are called **conditional probabilities** of C given the person is a non-smoker
- This last point is usually not well-understood. The most important thing to observe is that $C|(S=0)$ is a r.v. because although we know that $S=0$, we do not know whether $C=0$ or $C=1$ yet. Each of these two possibilities have their respective probabilities, denoted by $P(C=0|S=0)$ and $P(C=1|S=0)$, respectively.
- If you wish you can give another name to $C|S=0$. For example, say you defined $Z=(C|S=0)$. Then Z is either 0 or 1, and we can calculate $P(Z=0)$ and $P(Z=1)$ from the table. With this notation $P(Z=0)=P(C=0|S=0)$ and $P(Z=1)=P(C=1|S=0)$.

- Since $C|(S=0)$ is a r.v. we should be able to calculate its expected value and variance, right?
- Before that, let us summarize what we have done so far under the probability distribution of $C|S=0$ as follows (in two alternative ways)

$C\|(S=0)$	0	1
$P(C\|S=0)$	.03	.97

Z	0	1
$P(Z)$	.03	.97

- Once you write down the probability distribution as above it is straightforward to compute its expected value and variance:

$$
\begin{aligned}
E(C|S=0) = \mu_{C|S} &= \sum_{c\in\{0,1\}} cP(C=c|S=0) \\
&= 0\times(.03) + 1\times(.97) = 0.97 \\
Var(C|S=0) = \sigma^2_{C|S} &= \sum_{c\in\{0,1\}} (\mu_{C|S}-c)^2 P(C=c|S=0) \\
&= (0-.97)^2(.03) + (1-.97)^2(0.97) = 0.0291
\end{aligned}
$$

- Through this example, we have seen most of the most important concepts that you should know in this section. Next, we are going to express these concepts in terms of two arbitrary r.v.'s X and Y.

Definitions

- A joint probability distribution is used to express the probability that simultaneously X takes the specific value x and Y takes the value y:

$$P(x, y) = P(X = x, Y = y)$$

- The marginal probability distributions are

$$P(x) = \sum_y P(x, y), \qquad P(y) = \sum_x P(x, y)$$

- The conditional probability distribution of the random variable Y expresses the probability that Y takes the value y when the value x is specified for X:

$$P(y|x) = \frac{P(x, y)}{P(x)}$$

- Similarly, the conditional probability function of X, given $Y = y$ is

$$P(x|y) = \frac{P(x, y)}{P(y)}$$

- The jointly distributed random variables X and Y are said to be independent if the following condition is satisfied

$$P(x, y) = P(x)P(y)$$

for all possible (x, y) pairs.

Definitions Cont'd

- The conditional mean of X given $Y = y^*$ is

$$\mu_{X|Y} = E[X|Y] = \sum_x x \cdot P(X = x|Y = y^*)$$

Notation: In order to clarify on which particular value of Y conditioned on we could have used the notation $\mu_{X|Y=y^*} = E[X|Y = y^*]$. But as long as it is clear from the content that it is conditioned on $Y = y^*$ we use the notation in the above equation. If you would like to simplify the notation further, you can denote $P(X = x|Y = y^*)$ by $P(x|y^*)$.

- Similarly, the conditional mean of Y given $X = x^*$ is

$$\mu_{Y|X} = E[Y|X] = \sum_y y \cdot P(Y = y|X = x^*)$$

- The conditional variance of X given $Y = y^*$ is

$$\sigma^2_{X|Y} = \sum_x (x - \mu_{X|Y})^2 P(X = x|Y = y^*)$$

- When conditional probabilities are involved notation becomes little cumbersome, so I recommend just to define another r.v. and work with it. For example, if you define a new r.v. as $Z = (X|Y = y^*)$ and compute its prob. dist., the conditional variance defined above equals actually just to the variance of Z, i.e. $\sigma^2_Z = \sum_z (z - \mu_z)^2 P(z)$.

Covariance

- Let X and Y be two discrete r.v.'s with means μ_x and μ_y.
- The expected value of $(X-\mu_x)(Y-\mu_y)$ is called covariance between X and Y:

$$Cov(X,Y) = E\left[(X-\mu_x)(Y-\mu_y)\right] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)P(x,y)$$

- Sometimes the following might be easier to compute the Cov(X,Y)

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

- If two random variables are statistically independent, the covariance between them is 0. However, the converse is not necessarily true.
- As an example, let us compute the covariance between smoking habits and cancer, i.e. $Cov(S,C) = ?$

$$\begin{aligned} Cov(S,C) &= E\left[(S-\mu_S)(C-\mu_C)\right] \\ &= \sum_s \sum_c (s-\mu_S)(c-\mu_C)P(s,c) \\ &= (0-.55)(0-.9)(.02) \\ &+ (0-.55)(1-.9)(.580) + \ldots \\ &+ \ldots (2-.55)(1-.9)(.105) = ? \end{aligned}$$

S/C	0	1
0	.020	.580
1	.035	.215
2	.045	.105

$\mu_S = .55$ and $\mu_C = .9$

Correlation

- Remember that in chapter 2, we discussed why the covariance might not be a good measure at revealing the relationship between two variables. (Basically because covariance is unit sensitive but review that lecture if you don't remember)
- The same applies here: Covariance only tells the direction of the relationship, not the strength of it. In order to measure both we use correlation coefficient...
- Correlation Coefficient measures the direction and relative strength of the linear relationship between two r.v. variables

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_Y}$$

Correlation

- Remember that in chapter 2, we discussed why the covariance might not be a good measure at revealing the relationship between two variables. (Basically because covariance is unit sensitive but review that lecture if you don't remember)
- The same applies here: Covariance only tells the direction of the relationship, not the strength of it. In order to measure both we use correlation coefficient...
- Correlation Coefficient measures the direction and relative strength of the linear relationship between two r.v. variables

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_Y}$$

Properties and interpretation of ρ_{XY}

- $-1 \leq \rho \leq 1$

Correlation

- Remember that in chapter 2, we discussed why the covariance might not be a good measure at revealing the relationship between two variables. (Basically because covariance is unit sensitive but review that lecture if you don't remember)
- The same applies here: Covariance only tells the direction of the relationship, not the strength of it. In order to measure both we use correlation coefficient...
- Correlation Coefficient measures the direction and relative strength of the linear relationship between two r.v. variables

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_Y}$$

Properties and interpretation of ρ_{XY}

- $-1 \leq \rho \leq 1$
- $\rho = 0$: no linear relationship between X and Y
- $\rho > 0$: positive linear relationship between X and Y, i.e. when X is high (low) then Y is likely to be high (low).
- $\rho < 0$: negative linear relationship between X and Y, i.e. when X is high (low) then Y is likely to be low (high)
- When $\rho = 1$ ($\rho = -1$) there is a perfect positive (negative) linear dependency between X and Y.

Linear Functions of Two Random Variables

- Just like in the single r.v. case, when we have two r.v.'s X and Y we can generate/define new r.v's.
- For example, $W_1 = 3X + Y$ and $W_2 = x^2 - 5Y$ are just two new r.v.'s that are generated from X and Y.
- In this section, we are going to be interested only in linear combinations of two r.v.'s:

$$W = aX \pm bY$$

- The Expected Value for the linear combination of two r.v.'s

$$E[W] = aE(X) \pm bE(Y) \quad \text{or}$$
$$\mu_w = a\mu_x \pm b\mu_y$$

- The Variance for the linear combination of two r.v.'s

$$Var[W] = a^2 Var(X) + b^2 Var(Y) \pm 2abCov(X, Y) \quad \text{or}$$
$$\sigma_w^2 = a^2\sigma_x^2 + b^2\sigma_y^2 \pm 2abCov(X, Y)$$

- As an example, let us look at the following portfolio example. We will also be able to gain some intuition why the term $2abCov(X, Y)$ appears in the formula above.

Portfolio Analysis

- Let random variable X be the price for stock A
- Let random variable Y be the price for stock B
- The market value, W, for the portfolio is given by the linear function

$$\boxed{W = aX + bY}$$

(a is the number of shares of stock A,
b is the number of shares of stock B)

Portfolio Analysis

(continued)

- The mean value for W is

$$\boxed{\begin{aligned}\mu_W &= E[W] = E[aX + bY] \\ &= a\mu_X + b\mu_Y\end{aligned}}$$

- The variance for W is

$$\boxed{\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\mathrm{Cov}(X,Y)}$$

or using the correlation formula

$$\boxed{\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\mathrm{Corr}(X,Y)\sigma_X\sigma_Y}$$

Example: Investment Returns

Return per $1,000 for two types of investments

		Investment	
$P(x_i y_i)$	Economic condition	Passive Fund X	Aggressive Fund Y
.2	Recession	- $ 25	- $200
.5	Stable Economy	+ 50	+ 60
.3	Expanding Economy	+ 100	+ 350

$$E(x) = \mu_x = (-25)(.2) + (50)(.5) + (100)(.3) = 50$$

$$E(y) = \mu_y = (-200)(.2) + (60)(.5) + (350)(.3) = 95$$

Computing the Standard Deviation for Investment Returns

$P(x_i y_i)$	Economic condition	Investment: Passive Fund X	Investment: Aggressive Fund Y
0.2	Recession	- $ 25	- $200
0.5	Stable Economy	+ 50	+ 60
0.3	Expanding Economy	+ 100	+ 350

$$\sigma_X = \sqrt{(-25-50)^2(0.2)+(50-50)^2(0.5)+(100-50)^2(0.3)}$$
$$= 43.30$$

$$\sigma_y = \sqrt{(-200-95)^2(0.2)+(60-95)^2(0.5)+(350-95)^2(0.3)}$$
$$= 193.71$$

Covariance for Investment Returns

		Investment	
$P(x_i y_i)$	Economic condition	Passive Fund X	Aggressive Fund Y
.2	Recession	- \$ 25	- \$200
.5	Stable Economy	+ 50	+ 60
.3	Expanding Economy	+ 100	+ 350

$$\begin{aligned} Cov(X,Y) &= (-25-50)(-200-95)(.2) + (50-50)(60-95)(.5) \\ &\quad + (100-50)(350-95)(.3) \\ &= 8250 \end{aligned}$$

Portfolio Example

Investment X: $\mu_x = 50$ $\sigma_x = 43.30$
Investment Y: $\mu_y = 95$ $\sigma_y = 193.21$
$\sigma_{xy} = 8250$

Suppose 40% of the portfolio (P) is in Investment X and 60% is in Investment Y:

$$E(P) = .4(50) + (.6)(95) = 77$$

$$\sigma_P = \sqrt{(.4)^2(43.30)^2 + (.6)^2(193.21)^2 + 2(.4)(.6)(8250)} = 133.04$$

The portfolio return and portfolio variability are between the values for investments X and Y considered individually

Interpreting the Results for Investment Returns

- The aggressive fund has a higher expected return, but much more risk

$$\mu_y = 95 \;>\; \mu_x = 50$$

but

$$\sigma_y = 193.21 \;>\; \sigma_x = 43.30$$

- The Covariance of 8250 indicates that the two investments are positively related and will vary in the same direction

Review Example 1

Suppose that X is the number of tennis rackets and Y is the number of golf clubs sold daily in a small sports store, and their joint probability distribution is as in the table below

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

1. Determine the marginal probability distributions of X and Y.

Review Example 1

Suppose that X is the number of tennis rackets and Y is the number of golf clubs sold daily in a small sports store, and their joint probability distribution is as in the table below

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

1. Determine the marginal probability distributions of X and Y.

X	1	2	3
P(X)	.6	.3	.1

Y	1	2	3
P(Y)	.5	.3	.2

Review Example 1

Suppose that X is the number of tennis rackets and Y is the number of golf clubs sold daily in a small sports store, and their joint probability distribution is as in the table below

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

1. Determine the marginal probability distributions of X and Y.

X	1	2	3
P(X)	.6	.3	.1

Y	1	2	3
P(Y)	.5	.3	.2

2. Are X and Y independent? Explain.

Review Example 1

Suppose that X is the number of tennis rackets and Y is the number of golf clubs sold daily in a small sports store, and their joint probability distribution is as in the table below

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

1. Determine the marginal probability distributions of X and Y.

X	1	2	3
P(X)	.6	.3	.1

Y	1	2	3
P(Y)	.5	.3	.2

2. Are X and Y independent? Explain.

   X and Y are independent since $P(x,y) = P(x) \cdot P(y)$ for all pairs (x,y). For example,

$$P(X=1, Y=1) = P(X=1) \cdot P(Y=1)$$
$$0.3 = (0.6) \cdot (0.5)$$

   Verify that this condition is satisfied for all the other pairs (x,y).

Example Cont'd

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

3 Calculate the conditional probability $P(Y = 3|X = 1)$.

Example Cont'd

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

3 Calculate the conditional probability $P(Y = 3|X = 1)$.

$$P(Y = 3|X = 1) = \frac{P(X=1 \text{ and } Y=3)}{P(X=1)} = \frac{0.12}{0.60} = 0.20$$

Example Cont'd

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

3 Calculate the conditional probability $P(Y = 3|X = 1)$.

$$P(Y = 3|X = 1) = \frac{P(X=1 \text{ and } Y=3)}{P(X=1)} = \frac{0.12}{0.60} = 0.20$$

4 Calculate the expected values of X and Y, i.e, E(X) and E(Y).

Example Cont'd

$X(row)/Y(col.)$	1	2	3
1	.30	.18	.12
2	.15	.09	.06
3	.05	.03	.02

3 Calculate the conditional probability $P(Y = 3|X = 1)$.

$$P(Y = 3|X = 1) = \frac{P(X=1 \text{ and } Y=3)}{P(X=1)} = \frac{0.12}{0.60} = 0.20$$

4 Calculate the expected values of X and Y, i.e, E(X) and E(Y).

To calculate E(X) and E(Y) we need the marginal probabilities from part (a).

$$\mu_x = \sum_{x\in\{1,2,3\}} x \cdot P(x) = 1.(0.6) + 2.(0.3) + 3.(0.1) = 1.5$$

$$\mu_y = \sum_{y\in\{1,2,3\}} y \cdot P(y) = 1.(0.5) + 2.(0.3) + 3.(0.2) = 1.7$$

5 Calculate the variance for X and Y.

5 Calculate the variance for X and Y.

$$\sigma_y^2 = \sum_{x \in \{1,2,3\}} (x - \mu_x)^2 P(X = x)$$
$$= (1 - 1.7)^2 (0..5) + (2 - 1.7)^2 (0.3) + (3 - 1.7)^2 (0.2) = 0.61$$

Similarly, verify that $\sigma_x^2 = .45$.

6 Calculate $E(XY)$.

5 Calculate the variance for X and Y.

$$\sigma_y^2 = \sum_{x \in \{1,2,3\}} (x - \mu_x)^2 P(X = x)$$
$$= (1 - 1.7)^2(0..5) + (2 - 1.7)^2(0.3) + (3 - 1.7)^2(0.2) = 0.61$$

Similarly, verify that $\sigma_x^2 = .45$.

6 Calculate $E(XY)$.

Let us give little more detail on this part. We can think of in two different ways:

(a) Define $Z = XY$, and construct the prob. dist. table for Z, and then calculate $E(Z)$:

Z	1	2	3	4	6	9
P(z)	.30	.33	0.17	.09	.09	.02

(Note that we get $Z = 2$ for two different (x, y) pairs, $(1, 2)$ and $(2, 1)$, so we need to add them up to get $P(Z = 2)$. In another words, $P(Z = 2) = P(1, 2) + P(2, 1) = 0.18 + 0.15 = 0.33$. We did the same thing to compute $P(Z = 3)$ and $P(Z = 6)$. Check!)
$E(Z) = \sum_z z \cdot P(z) = 1(.3) + 2(.33) + \ldots + 9(.02) = 2.55$

(b) Alternatively, for each (x, y) pair compute $x \cdot y \cdot P(x, y)$, and then add them up:
$E(XY) = (1)(1)(.3)+(1)(2)(.18)+(1)(3)(.12)+\ldots+(3)(3)(.02) = 2.55$

7 Calculate Cov(X,Y). Did you expect this answer? Why?

7 Calculate Cov(X,Y). Did you expect this answer? Why?

$Cov(X, Y) = E(XY) - E(X)E(Y) = 2.55 - (1.70)(1.50) = 0.$
Yes, this was expected since X and Y are independent, from part (c).

8 Find the probability distribution of the random variable $W = X + Y$.

7 Calculate Cov(X,Y). Did you expect this answer? Why?

$Cov(X, Y) = E(XY) - E(X)E(Y) = 2.55 - (1.70)(1.50) = 0.$
Yes, this was expected since X and Y are independent, from part (c).

8 Find the probability distribution of the random variable $W = X + Y$.

W	2	3	4	5	6
P(w)	.3	.33	.26	.09	.02

9 Compute $E(W)$ and $Var(W)$ directly by using the probability distribution of W.

7 Calculate Cov(X,Y). Did you expect this answer? Why?

$Cov(X, Y) = E(XY) - E(X)E(Y) = 2.55 - (1.70)(1.50) = 0.$
Yes, this was expected since X and Y are independent, from part (c).

8 Find the probability distribution of the random variable $W = X + Y$.

W	2	3	4	5	6
P(w)	.3	.33	.26	.09	.02

9 Compute $E(W)$ and $Var(W)$ directly by using the probability distribution of W.

$$\mu_w = \sum_w w \cdot P(W = w) = 2(.3) + \ldots + 6(.02) = 3.2$$

$$\sigma_w^2 = \sum_w (w - \mu_w)^2 \cdot P(W = w) = (2 - 3.2)^2(.3) + \ldots + (6 - 3.2)^2(= 1.06$$

10 Compute $E(W)$ and $Var(W)$ by using the formulas given for the r.v.'s are linear combinations of two r.v.'s.

7 Calculate Cov(X,Y). Did you expect this answer? Why?

$Cov(X, Y) = E(XY) - E(X)E(Y) = 2.55 - (1.70)(1.50) = 0.$
Yes, this was expected since X and Y are independent, from part (c).

8 Find the probability distribution of the random variable $W = X + Y$.

W	2	3	4	5	6
P(w)	.3	.33	.26	.09	.02

9 Compute $E(W)$ and $Var(W)$ directly by using the probability distribution of W.

$$\begin{aligned}\mu_w =& \sum_w w \cdot P(W = w)\\ =& 2(.3) + \ldots + 6(.02)\\ =& 3.2\end{aligned}$$

$$\begin{aligned}\sigma_w^2 =& \sum_w (w - \mu_w)^2 \cdot P(W = w)\\ =& (2 - 3.2)^2(.3) + \ldots + (6 - 3.2)^2(\\ =& 1.06\end{aligned}$$

10 Compute $E(W)$ and $Var(W)$ by using the formulas given for the r.v.'s are linear combinations of two r.v.'s.

$W = aX + bY$, so here $a = 1$ and $b = 1$. Therefore,

$$\begin{aligned}\mu_w =& 1.\mu_x + 1.\mu_y\\ =& 1.5 + 1.7\\ =& 3.2\end{aligned}$$

$$\begin{aligned}\sigma_w^2 =& (1)^2\sigma_x^2 + (1)^2\sigma_y^2 + 2(1)(1)Cov(X,\\ =& 0.61 + 0.45 + 2 \cdot 0\\ =& 1.06\end{aligned}$$

11 Show that $Var(X+Y) = Var(X) + Var(Y)$. Did you expect this result? Why?

11 Show that $Var(X+Y) = Var(X) + Var(Y)$. Did you expect this result? Why?

$Var(X) + Var(Y) = .61 + .45 = 1.06$ which is equal to $Var(W = X + Y) = 1.06$. Yes, since X and Y are independent random variables.

12 Compute $P(X = 2 \mid X + Y = 4)$=?

11 Show that $Var(X + Y) = Var(X) + Var(Y)$. Did you expect this result? Why?

$Var(X) + Var(Y) = .61 + .45 = 1.06$ which is equal to $Var(W = X + Y) = 1.06$. Yes, since X and Y are independent random variables.

12 Compute $P(X = 2 \mid X + Y = 4)$=?

Since we know that $X + Y = 4$, we can forget all (x,y) pairs that doesn't satisfy this condition. So the joint probability distribution reduces to the following

$X(row)/Y(col.)$	1	2	3
1	.30	.18	**.12**
2	.15	**.09**	.06
3	**.05**	.03	.02

Then,

$$P(X = 2 \mid X + Y = 4) = \frac{.09}{.12 + .09 + .05} = 9/26$$

Review Example 2

Suppose you are asked to analyze a stock portfolio that contains 5 shares of stock A and 10 shares of stock B. The joint probability distribution of the stock prices is presented in the table below. For example, $P(A = \$60, B = \$55) = 0.20$, and so on. Also, numbers at the end of each row and column denote the marginal probabilities, i.e., $P(B = \$55) = 0.50$ etc.

	Stock B			
Stock A	\$35	\$55	\$80	
\$50	0.10	?	?	0.30
\$60	0.10	0.20	?	0.40
\$70	?	0.15	0.10	?
	?	0.50	0.25	1.00

1. Complete the missing joint and marginal probabilities in the table.
2. Compute the expected value of the portfolio.
3. Compute the standard deviation of the portfolio.
4. What is the expected price of Stock A given that price of Stock B is 55?

1.

	Stock B			
Stock A	\$35	\$55	\$80	
\$50	0.10	**0.15**	**0.05**	0.30
\$60	0.10	0.20	**0.10**	0.40
\$70	**0.05**	0.15	0.10	**0.30**
	0.25	0.50	0.25	1.00

2. We represent this portfolio as $W = 5A + 10B$, then we know that $E(W) = 5\mu_A + 10\mu_B$. Therefore, first we need to find expected values of stock A and stock B.
$\mu_A = 50(.3) + 60(.4) + 70(.3) = 60$, and
$\mu_B = 35(.25) + 55(.5) + 80(.25) = 56.3$. Now we can compute the expected value of the portfolio as $\mu_w = 5(60) + 10(56.3) = 862.5$

3. $\sigma_w^2 = 5^2\sigma_B + 10^2\sigma_B + 2(5)(10)Cov(A, B)$.
Let us first find Var(A), Var(B), and Cov(A,B):
$\sigma_A^2 = (50-60)^2 0.3 + (60-60)^2 0.4 + (70-60)^2 0.3 = 60$, $\sigma_B^2 = 254.7$

$$
\begin{aligned}
Cov(A, B) =& (50-60)(35-56.3)(0.10) + (60-60)(35-56.3)(0.10) \\
& + (70-60)(35-56.3)(0.05) + \ldots + (70-60)(80-56.3)(0.10) \\
& = 22.5
\end{aligned}
$$

3. Finally,

$$\begin{aligned} Var(W) &= 5^2 Var(A) + 10^2 Var(B) + 2(5)(10) Cov(A, B) \\ &= 5^2(60) + 10^2(254.69) + 2(5)(10)(22.5) \\ &= 29,219 \end{aligned}$$

$$\sigma_w = \sqrt{Var(W)} = \sqrt{29,219} = 170.9$$

4. Let us first construct the probability distribution table for the random variable $A|B = 55$:

$A \mid B = 55$	50	60	70
$P(A \mid B = 55)$	$.15/.50$	$.20/.50$	$.15/.50$

Then, we can compute the expected value of this r.v. as usual:

$$E(A|B = 55) = 50(\frac{0.15}{0.50}) + 60(\frac{0.20}{0.50}) + 70(\frac{0.15}{0.50}) = 60$$