Lecture 4

- **Read:** Chapter 2.8-2.9, 3.1-3.4.
- Discrete RVs
  - Conditional Probability Mass Function
- Multiple Discrete RVs
  - Joint PMFs
  - Marginal PMFs
  - Functions of Two Random Variables
  - Expectations of Functions of Two Random Variables
  - Covariance and Correlation

Conditional PMF

- Recall from our previous discussion of conditional probability that the conditional probability $P[A|B]$ is a number that expresses our knowledge about the occurrence of event A, when we learn that another event B occurs.
- Here, we consider event A to be the observation of a particular value of a random variable ($A = \{X = x\}$).
- The conditioning event B contains information about X but not the precise value of X.
  - For example, we might learn that $X \leq 33$ or that $|X| > 100$
- In general, we learn of the occurrence of an event B which describes some property of X.

Conditional PMF: Example

- Let $N =$ number of bytes in a fax
- A conditioning event might be the event I that the fax contains an image.
- A second kind of conditioning would be the event $\{N > 10,000\}$ which tells us that the fax required more than 10,000 bytes.
- Both events I and $\{N > 10,000\}$ give us information that the fax is likely to have many bytes.

Conditional PMF

- The occurrence of the conditioning event B changes the probabilities of the event $\{\mathrm{X} = \mathrm{x}\}$.
- Given this information and a probability model for our experiment, we can use the definition of conditional probability to write

$$P[A|B] = P[X = x|B]$$

  for all real numbers x.
- This collection of probabilities is a function of x.
- **Definition:** Given an event B with $P[B] > 0$, conditional PMF of X is:

$$p_{X|B}(x) = P[X = x|B]$$

- Two kinds of conditioning...

Conditional PMF: Version 1

- $p_{X|B_i}(x)$ is a model for the PMF of X given some information B_i.
- **Example:** $B_i =$ the ith month of the year
  $X = \#$ of cars on the highway
- In this case, we are given an event space $B_1, B_2, \ldots, B_m$ that describes mutually exclusive possibilities for an experiment.
- Associated with each event B_i is a probability model for X in the form of the conditional PMF $p_{X|B_i}(x)$.
- We then use the law of total probability to find the PMF $p_X(x)$:

$$p_X(x) = \sum_{i=1}^{m} p_{X|B_i}(x) P[B_i]$$

Conditional PMF: Version 1 Example 1

- Let B_i denote the ith hour of the day
- $B_1 =$ from 0 to 1 AM
- Let $X = \#$ of packets that arrive in a given hour
- $p_{X|B_i}(x) =$ probability that $X = x$ during the ith hour of the day
- What is the PMF of X?

. . . .

$$
\begin{aligned}
p_X(x) &= \sum_{i=1}^{m} p_{X|B_i}(x) P[B_i] \\
&= \sum_{i=1}^{24} p_{X|B_i}(x) \times \frac{1}{24} \\
&= \text{ probability that regardless of time of day I see } x \text{ packets}
\end{aligned}
$$

Conditional PMF: Version 1 Example 2

- In the ith month of the year, the number of cars N crossing the Bosphorus Bridge is Poisson with parameter α_i.
- For a randomly chosen month, what is the PMF of X?

. . . .

$$p_{X|B_i}(x) = \begin{cases} \frac{{\alpha_i}^x e^{-\alpha_i}}{x!} & , \mathrm{x} = 0,1,2,\ldots \\ 0 & , \text{otherwise} \end{cases}$$

$$p_X(x) = \frac{1}{12} \sum_{i=1}^{12} p_{X|B_i}(x)$$

Conditional PMF: Version 1 Example 3

- Let X denote the number of additional years that a randomly chosen 70-year-old person will live.
- If the person has high blood pressure, denoted as event H, then X is a geometric RV with $p = 0.1$.
- Otherwise, if the person's blood pressure is regular, event R, then X has a geometric PMF with parameter $p = 0.05$.
- What is the conditional PMF of X given event H, $p_{X|H}(x)$?
- What is the conditional PMF of X given event R, $p_{X|R}(x)$?

. . . .

$$p_{X|H}(x) = \begin{cases} 0.1(0.9)^{x-1} & , \text{x} = 1,2,\dots \\ 0 & , \text{otherwise} \end{cases}$$

$$p_{X|R}(x) = \begin{cases} 0.05(0.95)^{x-1} & , \text{x} = 1,2,\dots \\ 0 & , \text{otherwise} \end{cases}$$

Conditional PMF: Version 1 Example 3 (cont.)

- If 40% of all seventy-year-olds have high blood pressure, what is the PMF of X?

. . . .

Since H, R is an event space, we can use the law of total probability to write

$$p_X(x) = p_{X|H}(x)P[H] + p_{X|R}(x)P[R]$$
$$= \begin{cases} (0.4)(0.1)(0.9)^{x-1} + (0.6)(0.05)(0.95)^{x-1} & \text{, x= 1,2,...} \\ 0 & \text{, otherwise} \end{cases}$$

Conditional PMF: Version 2

- The event B is defined as a subset of S_X such that for each $x \in S_X$, either $x \in B$ or $x \notin B$.
- In this case, the PMF $p_X(x)$ is enough to specify both the probability of B as well as the conditional PMF $p_{X|B}(x)$.
- When B is a subset of S_X, the definition of conditional probability permits us to write

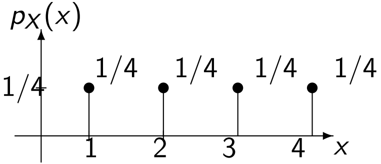
$$p_{X|B}(x) = \frac{P[X = x, B]}{P[B]} = \frac{P[\{X = x\} \cap B]}{P[B]}$$

- Now either event $X = x$ is contained in event B or it is not.
  - If $x \in B$, then $\{X = x\} \cap B = \{X = x\}$ and $P[X = x, B] = p_X(x)$.
  - Otherwise, if $x \notin B$, then $\{X = x\} \cap B = \emptyset$ and $P[X = x, B] = 0$.
- Thus, we can calculate conditional PMF

$$p_{X|B}(x) = P[X = x|B] = \begin{cases} \frac{p_X(x)}{P[B]} & \text{, if x} \in \text{B} \\ 0 & \text{, if x} \notin \text{B} \end{cases}$$

Conditional PMF Version 2: Example 1

- Consider X with PMF $p_X(x)$.



- What is $p_{X|B}(x)$ if $B = \{x \geq 3\}$?

. . . .

From the graph, we observe that $P[B] = 1/2$. So,

$$p_{X|B}(x) = \begin{cases} \frac{1/4}{1/2} = 1/2 & , \mathrm{x} = 4 \\ \frac{1/4}{1/2} = 1/2 & , \mathrm{x} = 3 \\ 0 & , \mathrm{x} = 2 \\ 0 & , \mathrm{x} = 1 \end{cases}$$

Conditional PMF Version 2: Example 2

- X is geometric with $p = 0.1$.
- What is the conditional PMF of X given $B = \{x > 9\}$?

. . . .

$$
\begin{aligned}
p_X(x) = P[X = x] &= (1-p)^{x-1}p, \qquad \text{x} = 1,\ 2,\ 3,\ \ldots \\
P[B] = P[X > 9] &= 1 - P[X \le 9] \\
&= 1 - \sum_{x=1}^{9} p_X(x) \\
&= 1 - \sum_{x=1}^{9} (1-p)^{x-1}p \\
&= 1 - [1 - (1-p)^9] \quad \text{sum of the first 9 terms for a geometric series} \\
&= (1-p)^9 \quad \text{(failed nine times)}
\end{aligned}
$$

Conditional PMF Version 2: Example 2 (cont.)

- X is geometric with $p = 0.1$.
- What is the conditional PMF of X given $B = \{x > 9\}$?

. . . .

$$p_X(x) = P[X = x] = (1 - p)^{x-1} p, \qquad \text{x} = 1, 2, 3, \ldots$$

$$P[B] = (1 - p)^9$$

$$p_{X|B}(x) = \begin{cases} \frac{(1-p)^{x-1} p}{(1-p)^9} & , \text{x} = 10, 11, 12, \ldots \\ 0 & , \text{otherwise} \end{cases}$$

Conditional PMF Version 2: Example 3

- In the probability model for the fax example, the length of a fax has PMF

$$p_X(x) = \begin{cases} 0.15 & , \text{x} = 1,\ 2,\ 3,\ 4 \\ 0.1 & , \text{x} = 5,\ 6,\ 7,\ 8 \\ 0 & , \text{otherwise} \end{cases}$$

- Suppose the company has two fax machines, one for faxes shorter than five pages and the other for faxes that have five or more pages.

- What is the PMF of fax durations in the second machine?

Conditional PMF Version 2: Example 3 (cont.)

- Relative to $p_X(x)$, we seek a conditional PMF.
- The condition is $X \in L$ where $L = \{5, 6, 7, 8\}$.

$$p_{X|L}(x) = \begin{cases} \frac{p_X(x)}{P[L]} & , \text{x}= 5,\ 6,\ 7,\ 8 \\ 0 & , \text{otherwise} \end{cases}$$

- From the definition of L, we have

$$P[L] = \sum_{5}^{8} p_X(x) = 0.4$$

- With $p_X(x) = 0.1$ for $x \in L$,

$$p_{X|L}(x) = \begin{cases} 0.25 & , \text{x}= 5,\ 6,\ 7,\ 8 \\ 0 & , \text{otherwise} \end{cases}$$

- Thus, the lengths of long faxes are equally likely. Among the long faxes, each length has probability 0.25.

Conditional PMF Version 2: Example 4

- Suppose X, the time in minutes that you wait for a bus, has the uniform PMF

$$p_X(x) = \begin{cases} 1/20 & \text{, x= 1, 2,..., 20} \\ 0 & \text{, otherwise} \end{cases}$$

- Suppose the bus has not arrived by the eighth minute, what is the conditional PMF of your waiting time X?

. . . .

- Let A denote the event $X > 8$.

- Observing that $P[A] = 12/20$, we can write the conditional PMF of X as

$$p_{X|X>8}(x) = \begin{cases} \frac{1/20}{12/20} = \frac{1}{12} & \text{, x= 9,10,...,20} \\ 0 & \text{, otherwise} \end{cases}$$

Conditional PMF Version 2 Summary

- Conditioning an RV X on an event B: remove samples that do not belong to B and normalize

  $$p_{X|B}(x) = P[X = x|B] = \frac{P[\{X = x\} \cap B]}{P[B]} = \begin{cases} \frac{p_X(x)}{P[B]} & \text{, if x} \in \text{B} \\ 0 & \text{, otherwise} \end{cases}$$

  This is called the conditional PMF of X given B, with all the nice properties of a PMF described earlier.

Conditional PMF Version 2: Example of Normalization

- Let $X =$ roll of a die, and $A = \{\text{outcome was even}\}$.
- Find $p_{X|A}(x)$.

. . . .

$$p_X(x) = \begin{cases} \frac{1/6}{1/2} = 1/3 & \text{, if x} = 2,4,6 \\ 0 & \text{, if x} = 1,3,5 \end{cases}$$

Conditional Probability Mass Function is Also a PMF (I)

- Note that $p_{X|B}(x)$ is also a PMF.
- Therefore, relative to the conditioning event B, it conforms to the three axioms of probability:
  1. For any $x \in B$, $p_{X|B}(x) \geq 0$.
  2. $\sum_{x \in B} p_{X|B}(x) = 1$.
  3. For $x_1, x_2 \in B$ and $x_1 \neq x_2$,
  $$P[\{x_1, x_2\}|B] = p_{X|B}(x_1) + p_{X|B}(x_2)$$
- Therefore, we can compute the statistics of the random variable $X|B$ in the same way that we compute statistics of X.
  - The only difference is that we use the conditional PMF $p_{X|B}(\cdot)$ in place of $p_X(\cdot)$.

Conditional Expected Value

- **Definition:(Conditional Expected Value)** Given the condition B, the conditional expected value of RV X is:

$$E[X|B] = \mu_{X|B} = \sum_{x \in B} x p_{X|B}(x)$$

- For a random variable X resulting from an experiment with event space $B_1, \ldots, B_m$, we can compute the expected value $E[X]$ in terms of the conditional expected values $E[X|B_i]$

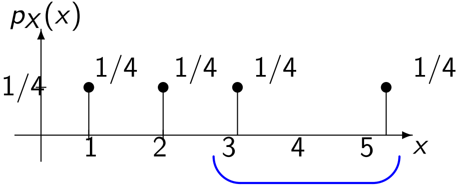
$$E[X] = \sum_{i=1}^{m} E[X|B_i] P[B_i]$$

- Given the condition B, the conditional expected value of a derived random variable $Y = g(X)$ is

$$E[g(X)|B] = \sum_{x \in B} g(x) p_{X|B}(x)$$

Conditional Expected Value: Example

- Suppose $X \sim \text{uniform}\{1, 2, 3, 5\}$.



- If $B = \{x \geq 3\}$, what is E[X|B]?

. . . .

From the graph, P[B]$= 1/2$.

$$p_{X|B}(x) = \begin{cases} \frac{p_X(3)}{P[B]} = \frac{1/4}{1/2} = \frac{1}{2} & , \text{x} = 3 \\ \frac{p_X(5)}{P[B]} = \frac{1}{2} & , \text{x} = 5 \\ 0 & , \text{otherwise} \end{cases}$$

$$E[X|B] = \sum_{x \in B} x p_{X|B}(x) = 3 \cdot \frac{1}{2} + 5 \cdot \frac{1}{2} = 4$$

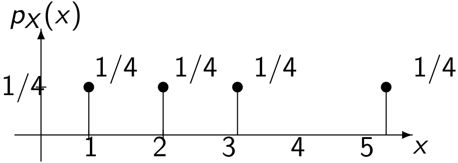
Conditional Variance

- **Definition:(Conditional Variance)** Given the condition B, the conditional variance of RV X is:

$$Var[X|B] = E[(X - E[X|B])^2|B] \quad = \sum_{x \in B} \underbrace{(x - E[X|B])^2}_{g(x)} p_{X|B}(x)$$

Conditional Variance: Example

- Suppose $X \sim \text{uniform}\{1, 2, 3, 5\}$.



- If $B = \{x \geq 3\}$, what is Var[X|B]?

. . . .

From the graph, $P[B] = 1/2$.

$$p_{X|B}(x) = \begin{cases} \frac{p_X(3)}{P[B]} = \frac{1/4}{1/2} = \frac{1}{2} & , \text{x} = 3 \\ \frac{p_X(5)}{P[B]} = \frac{1}{2} & , \text{x} = 5 \\ 0 & , \text{otherwise} \end{cases}$$

$$Var[X|B] = \sum_{x \in B} (x - \underbrace{4}_{E[X|B]})^2 p_{X|B}(x)$$

$$= (3-4)^2 \cdot \frac{1}{2} + (5-4)^2 \cdot \frac{1}{2} = 1$$

Conditional Mean, Conditional Variance, and Conditional Standard Deviation Example

- We had found that the conditional PMF for the long faxes was

$$p_{X|L}(x) = \begin{cases} 0.25 & , \text{x= 5, 6, 7, 8} \\ 0 & , \text{otherwise} \end{cases}$$

- Find the conditional mean, the conditional variance, and the conditional standard deviation for the long faxes.

. . . .

$$E[X|L] = \mu_{X|L} = \sum_{x=5}^{8} x p_{X|L}(x) = 0.25 \sum_{x=5}^{8} x = 6.5 \text{ pages}$$

$$E[X^2|L] = 0.25 \sum_{x=5}^{8} x^2 = 43.5 \text{ pages}^2$$

$$Var[X|L] = E[X^2|L] - \mu_{X|L}^2 = 1.25 \text{ pages}^2$$

$$\sigma_{X|L} = \sqrt{Var[X|L]} = 1.12 \text{ pages}$$

Conditional Variance and Conditional Standard Deviation

- What does conditional variance or conditional standard deviation mean?
- Does having some information A decrease the conditional variance $Var[X|A]$ of X (e.g., your earnings on the stock market)?

Variance Example: Recall

- Let X be the outcome of the roll of a die.
- Find its mean and variance.

. . . .

- We can use the definitions of expectation and variance.
- $E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 21/6 = 3.5$
- $E[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$
- $Var[X] = E[X^2] - (E[X])^2 = \frac{91}{6} - (\frac{21}{6})^2 = \frac{105}{36} = \frac{35}{12} \approx 2.9$

Conditional Variance Example

- Let X be the outcome of a roll of a die and $A = \{1, 6\}$.
- Find its conditional mean and variance. Are they larger or smaller than before?

. . . .

$$\begin{aligned} E[X|A] &= \sum_{x \in A} x p_{X|B}(x) \\ &= 1 \cdot \frac{1}{2} + 6 \cdot \frac{1}{2} \\ &= 3.5 = E[X] \end{aligned}$$

$$\begin{aligned} Var[X|A] &= \sum_{x \in A} (x - \underbrace{3.5}_{E[X|A]})^2 p_{X|B}(x) \\ &= (1 - 3.5)^2 \cdot \frac{1}{2} + (6 - 3.5)^2 \cdot \frac{1}{2} = 8.75 > 2.9 = Var[X] \end{aligned}$$

This means that conditional variance may be larger than variance!

Computing Expectations by Conditioning: Example and New Trick

- Determine the mean and variance of $X \sim \text{geometric}(p)$.
- Computing
$$E[X] = \sum_{x=1}^{\infty} x p_X(x) = \sum_{x=1}^{\infty} x p(1-p)^{x-1} = \frac{1}{p}$$
was messy!
- Consider the partition $A_1 = \{X = 1\}$ and $A_2 = \{X > 1\}$, then
$$E[X] = E[X|X=1]P[X=1] + E[X|X>1]P[X>1]$$
$$= 1 \cdot p + (1 + E[X])(1-p) \Rightarrow E[X] = \frac{1}{p}$$
- How would you do the same to compute the variance?
- Answer: Compute $E[X^2]$ using same strategy.
$\Rightarrow Var[X] = \frac{1-p}{p^2}$

Multiple Discrete RVs

- **Motivation:** Study dependence relationships and mutual coupling between multiple RVs associated with the same experiment
  - e.g., in medical diagnosis, the joint results from multiple tests may be significant.

- **Recall:** RVs are not just functions! To analyze multiple RVs, they need to share the *same* underlying probability model!

- An experiment produces both X and Y, e.g.,
  $X =$ minutes you wait for the number 40B bus to campus
  $Y =$ no. of other buses that pass by

Multiple Discrete RVs

- **Idea:** X and Y are two RVs modeling some phenomenon, both random together.

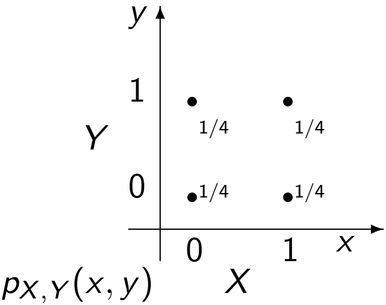
$$\xrightarrow{X} \boxed{\;?\;} \xrightarrow{Y}$$

- **Definition:(Joint PMF)** The joint PMF of X and Y is given by
  $$p_{X,Y}(x,y) = P[X = x, Y = y]$$

- **Definition:(Support)** of (X, Y) is the set of all possible values of the pair (X, Y)
  $$S_{X,Y} = \{(x,y) | p_{X,Y}(x,y) > 0\}$$

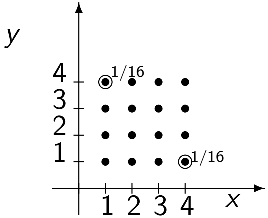
Support Example

- PMF for $(X, Y) \sim$ uniform$\{(0,0),(0,1),(1,0),(1,1)\}$.
- Draw the support.



Multiple Discrete RVs: Example

- Two tosses of a tetrahedral die (X,Y)



- Let $M = min[X, Y]$
  $N = max[X, Y]$
- Consider the pair (M, N).
- What are $S_{M,N}$ and $p_{M,N}(m, n)$?

. . . .

$(M, N) = (1, 4)$

$$p_{M,N}(1, 4) = P[M = 1, N = 4]$$

$$= P[(min(X, Y) = 1, max(X, Y) = 4] = \frac{1}{16} + \frac{1}{16}$$

Properties of Joint PMF

1. All the probabilities add up to 1.

$$\sum_{(x,y)} \sum_{\in S_{X,Y}} p_{X,Y}(x,y) = 1$$

2. $p_{X,Y}(x,y) \geq 0$ for all pairs (x,y)

3. Given a subset B of the plane

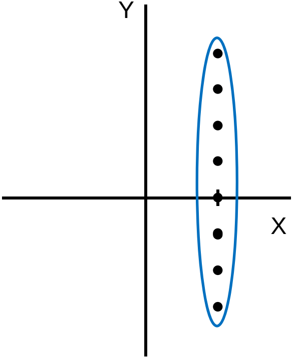
$$P[(X,Y) \in B] = P[B] = \sum_{(x,y)} \sum_{\in B} p_{X,Y}(x,y)$$

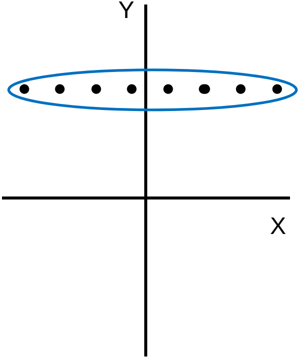
Marginal PMF

- **Definition:(Marginal PMFs)** of a joint distribution for X, Y are the PMFs of X and Y.

$p_{X,Y}(x,y)$ joint PMF

marginal PMF $p_X(x)$ $p_Y(y)$





Computing Marginal PMFs from Joint PMFs

- Suppose you are given $p_{X,Y}(x,y)$.
- What is $p_X(x)$?

. . . .

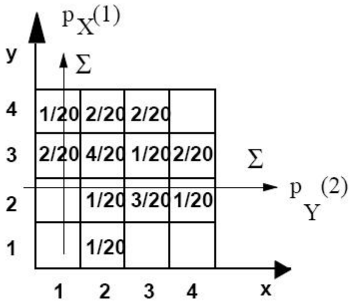
$$\begin{aligned} p_X(x) &= P[X = x] \\ &= \sum_y p_{X,Y}(x,y) \end{aligned}$$

- Similarly,

$$\begin{aligned} p_Y(y) &= P[Y = y] \\ &= \sum_x p_{X,Y}(x,y) \end{aligned}$$

Computing Marginal PMFs from Joint PMFs: Interpretation

- Consider the joint PMF exhibited in the table below.



- The marginals are obtained by summing rows and columns.

- Find $P[X + Y \leq 4]$.

Computing Marginal PMFs from Joint PMFs: Interpretation (cont.)

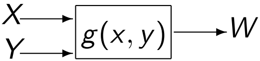
$$p_X(x) = \begin{cases} 3/20 & , \text{x} = 1,\ 4 \\ 8/20 & , \text{x} = 2 \\ 6/20 & , \text{x} = 3 \\ 0 & , \text{otherwise} \end{cases}$$

$$p_Y(y) = \begin{cases} 1/20 & , \text{y} = 1 \\ 5/20 & , \text{y} = 2,\ 4 \\ 9/20 & , \text{y} = 3 \\ 0 & , \text{otherwise} \end{cases}$$

$$P[X+Y \leq 4] = 4/20$$

Functions of Two Random Variables

- $W = g(X, Y)$



- Given $p_{X,Y}(x, y)$, what is the PMF of W?

$$p_W(w) = P[W = w] = \sum_{(x,y):\, g(x,y)=w} \sum p_{X,Y}(x, y)$$

Functions of Two Random Variables: Example 1

- Let $W = X \cdot Y$
- Suppose $p_{X,Y}(x,y)$ is as shown in the table

	$p_{X,Y}(x,y)$	Y = 0	Y = 1
X	0	1/4	1/4
	1	1/4	1/4

- What is the PMF of W?

. . . .

- $p_W(0) = 3/4$
- $p_W(1) = 1/4$

Functions of Two Random Variables: Example 2

- Let $W = X \cdot Y$
- Suppose $p_{X,Y}(x,y)$ is as shown in the table

	$p_{X,Y}(x,y)$	Y: 0	Y: 1
X	0	1/4	0
	1	1/4	1/2

- What is $p_W(w)$?

. . . .

- $S_W = \{0, 1\}$

$$p_W(w) = \begin{cases} 1/2 & , \ w = 0 \\ 1/2 & , \ w = 1 \\ 0 & , \ \text{otherwise} \end{cases}$$

Expectation of Functions

- In many situations we need to know only the expected value of a derived random variable rather than the entire probability model.
- In these situations, we can obtain the expected value directly from the joint PMF of the random variable pair.
  - We do not have to compute the PMF of the derived random variable.

Expectation of Functions

- **Theorem:** If $W = g(X, Y)$, then

$$E[W] = E[g(X, Y)] = \sum_{x} \sum_{y} g(x, y) p_{X,Y}(x, y)$$

- **Theorem:** If $g(X, Y) = g_1(X, Y) + g_1(X, Y) + ... + g_n(X, Y)$, then

$$E[g(X, Y)] = E[g_1(X, Y)] + ... + E[g_n(X, Y)]$$

"Expectation of the sum is equal to the sum of expectations!"

Expectation of Functions: Proof of Theorem

$$
\begin{aligned}
E[g(X,Y)] &= \sum_x \sum_y g(x,y) p_{X,Y}(x,y) \\
&= \sum_x \sum_y (g_1(x,y) + \ldots + g_n(x,y)) p_{X,Y}(x,y) \\
&= \sum_x \sum_y [g_1(x,y) p_{X,Y}(x,y) + \ldots + g_n(x,y) p_{X,Y}(x,y)] \\
&= \underbrace{\sum_x \sum_y g_1(x,y) p_{X,Y}(x,y)}_{E[g_1(X,Y)]} + \ldots \\
&\qquad + \underbrace{\sum_x \sum_y g_n(x,y) p_{X,Y}(x,y)}_{E[g_n(X,Y)]} \\
&= E[g_1(X,Y)] + \ldots + E[g_n(X,Y)]
\end{aligned}
$$

Expectation of Functions: Sum of Two RVs

- Let (X, Y) have joint PMF $p_{X,Y}(x, y)$.
- Then,

$$E[X + Y] = E[X] + E[Y]$$

$$\begin{aligned} Var[X + Y] &= E[(X + Y - \mu_X - \mu_Y)^2] = E[(X - \mu_X + Y - \mu_Y)^2] \\ &= E[(X - \mu_X)^2 + 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2] \\ &= Var[X] + Var[Y] + 2E[(X - \mu_X)(Y - \mu_Y)] \end{aligned}$$

Covariance

- **Definition:(Covariance)** Covariance of two RVs (X, Y) is

$$\begin{aligned} Cov[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY - X\mu_Y - \mu_X Y + \mu_X \mu_Y] \\ &= E[XY] - E[X\mu_Y] - E[\mu_X Y] + E[\mu_X \mu_Y] \end{aligned}$$

$$E[X\mu_Y] = \sum_x (x\mu_Y) p_X(x) = \mu_Y E[X] = \mu_X \mu_Y$$

$$\begin{aligned} Cov[X, Y] &= E[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\ &= E[XY] - \mu_X \mu_Y \end{aligned}$$

- **Note:** Suppose $Cov[X, Y] > 0$. This suggests that on average,
  - either $X > \mu_X$ and $Y > \mu_Y$
  - or $X < \mu_X$ and $Y < \mu_Y$
- **Interpretation:** If $Cov[X, Y] > 0$, then $X - \mu_X$ and $Y - \mu_Y$ tend to stray on the same side of their means. If $Cov[X, Y] < 0$, they tend to stray in opposite directions.

Correlation

- **Recall:** $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$
- **Definition:(Correlation)** Correlation of X and Y is $E[X \cdot Y]$.
- **Note:** If $E[X] = E[Y] = 0$, then $Cov[X,Y] = E[X \cdot Y]$.
- **Definition:(Orthogonal)** X and Y are said to be orthogonal if $E[X \cdot Y] = 0$.

Correlation

- **Definition:(Uncorrelated RVs)** (X, Y) are such that $Cov[X, Y] = 0$ or alternatively $E[XY] = E[X]E[Y]$.
- **Note:** $Cov[X, X] = Var[X]$
- **Recall:** $Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$
- **Note:** If X and Y are uncorrelated
$$Var[X + Y] = Var[X] + Var[Y]$$

Correlation Coefficient

- **Definition:(Correlation Coefficient)** $\rho_{X,Y}$

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

- **Theorem:** The correlation coefficient is normalized, i.e.,

$$-1 \leq \rho_{X,Y} \leq 1$$

(For proof, refer to text.)

- **Theorem:** If $Y = aX + b$, then

$$\rho_{X,Y} = \begin{cases} -1 & , a < 0 \\ 1 & , a > 0 \\ 0 & , a = 0 \end{cases}$$

Correlation Coefficient: Proof of Theorem

$$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = E[(X-\mu_X)(aX+b-a\mu_X-b]$$

$$[\quad \mu_Y = E[Y] = E[aX+b] = aE[X]+b = a\mu_X + b \quad]$$

$$Cov[X,Y] = aE[(X-\mu_X)^2] = aVar[X]$$

$$Var[Y] = a^2 Var(X)$$

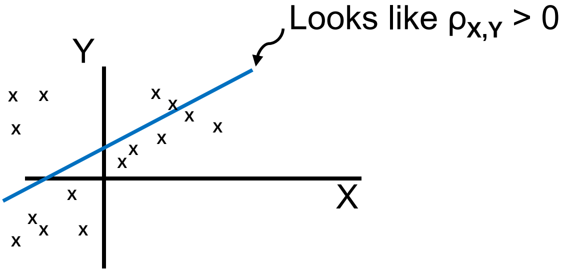
not a!

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} = \frac{aVar[X,Y]}{\sqrt{a^2 Var[X]^2}}$$

$$= \frac{a}{\sqrt{a^2}} = \frac{a}{|a|} = \begin{cases} 1 & \text{, if } a > 0 \\ -1 & \text{, if } a < 0 \end{cases}$$

Correlation Coefficient: Example

- Collect data $(x_1, y_1), \ldots, (x_n, y_n)$.



- And if $\rho_{X,Y} = 1$, the relationship between X and Y might be appropriately modeled by a straight line, with positive slope!