Discrete Random Variables: Joint PMFs, Conditioning and Independence

Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University

Reference:

- D. P. Bertsekas, J. N. Tsitsiklis, *Introduction to Probability* , Sections 2.5-2.7

Motivation

- Given an experiment, e.g., a medical diagnosis
  - The results of blood test is modeled as numerical values of a random variable X
  - The results of magnetic resonance imaging (MRI,核磁共振攝影) is also modeled as numerical values of a random variable Y

We would like to consider probabilities involving simultaneously the numerical values of these two variables and to investigate their mutual couplings

$$\mathbf{P}\left(\{X = x\} \cap \{Y = y\}\right)?$$

Joint PMF of Random Variables

- Let X and Y be random variables associated with the same experiment, the **joint PMF** of X and Y is defined by

$$p_{X,Y}(x,y) = \mathbf{P}(\{X=x\}\cap\{Y=y\}) = \mathbf{P}(X=x, Y=y)$$

- if event A is the set of all pairs (x,y) that have a certain property, then the probability of A can be calculated by

$$\mathbf{P}((X,Y)\in A) = \sum_{(x,y)\in A} p_{X,Y}(x,y)$$

Marginal PMFs of Random Variables (1/2)

- The **PMFs** of random variables X and Y can be calculated from their **joint PMF**

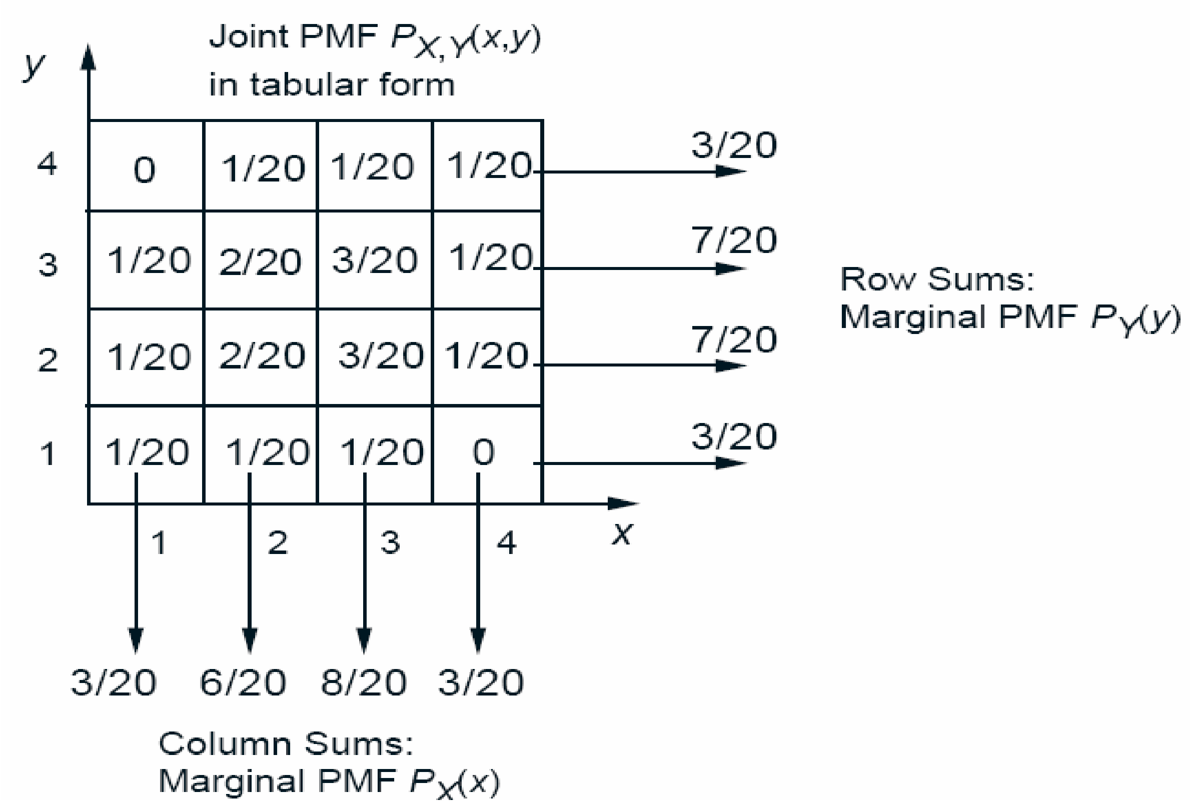
$$p_X(x) = \sum_y p_{X,Y}(x,y), \qquad p_Y(y) = \sum_x p_{X,Y}(x,y)$$

  - $p_X(x)$ and $p_Y(y)$ are often referred to as the **marginal PMFs**
  - The above two equations can be verified by

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

Marginal PMFs of Random Variables (2/2)

- **Tabular Method**: Given the joint PMF of random variables X and Y is specified in a two-dimensional table, the marginal PMF of X or Y at a given value is obtained by adding the table entries along a corresponding column or row, respectively



Functions of Multiple Random Variables (1/2)

- A function $Z = g(X,Y)$ of the random variables X and Y defines another random variable. Its PMF can be calculated from the joint PMF $p_{X,y}$

$$p_Z(z) = \sum_{\{(x,y) | g(x,y)=z\}} p_{X,Y}(x,y)$$

- The expectation for a function of several random variables

$$\mathbf{E}[Z] = \mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

Functions of Multiple Random Variables (2/2)

- If the function of several random variables is linear and of the form $Z = g(X, Y) = aX + bY + c$

$$\mathbf{E}[Z] = a\mathbf{E}[X] + b\mathbf{E}[Y] + c$$

  - How can we verify the above equation ?

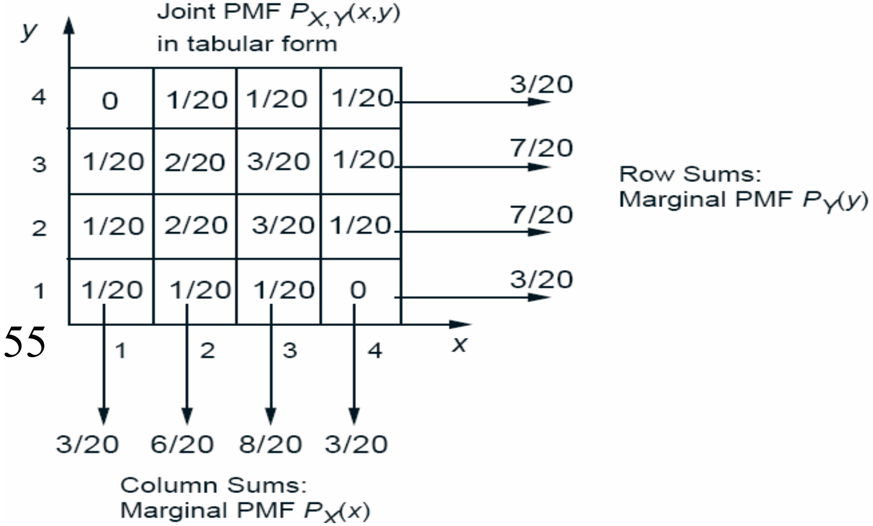
An Illustrative Example

- Given the random variables X and Y whose joint is given in the following figure, and a new random variable Z is defined by $Z = X + 2Y$, calculate $\mathbf{E}[Z]$
  - Method 1:

$$\mathbf{E}[X] = 1\cdot\frac{3}{20} + 2\cdot\frac{6}{20} + 3\cdot\frac{8}{20} + 4\cdot\frac{3}{20} = \frac{51}{20}$$

$$\mathbf{E}[Y] = 1\cdot\frac{3}{20} + 2\cdot\frac{7}{20} + 3\cdot\frac{7}{20} + 4\cdot\frac{3}{20} = \frac{50}{20}$$

$$\mathbf{E}[Z] = \mathbf{E}[X] + 2\mathbf{E}[Y] = \frac{51}{20} + 2\cdot\frac{50}{20} = \frac{151}{20} = 7.55$$



  - Method 2:

$$p_Z(z) = \sum_{\{(x,y) \mid x+2y=z\}} p_{X,Y}(x,y)$$

$$p_Z(3) = \frac{1}{20},\ p_Z(4) = \frac{1}{20},\ p_Z(5) = \frac{2}{20},\ p_Z(6) = \frac{2}{20}$$

$$p_Z(7) = \frac{4}{20},\ p_Z(8) = \frac{3}{20},\ p_Z(9) = \frac{3}{20},\ p_Z(10) = \frac{2}{20}$$

$$p_Z(11) = \frac{1}{20},\ p_Z(12) = \frac{1}{20}$$

$$\therefore \mathbf{E}[Z] = 3\cdot\frac{1}{20} + 4\cdot\frac{1}{20} + 5\cdot\frac{2}{20} + 6\cdot\frac{2}{20} + 7\cdot\frac{4}{20} + 8\cdot\frac{3}{20} + 9\cdot\frac{3}{20} + 10\cdot\frac{2}{20} + 11\cdot\frac{1}{20} + 12\cdot\frac{1}{20} = 7.55$$

More than Two Random Variables (1/2)

- The joint PMF of three random variables X , Y and Z is defined in analogy with the above as

$$p_{X,Y,Z}(x, y, z) = \mathbf{P}(X = x, Y = y, Z = z)$$

  – The corresponding marginal PMFs

$$p_{X,Y}(x, y) = \sum_{z} p_{X,Y,Z}(x, y, z)$$

and

$$p_{X}(x) = \sum_{y} \sum_{z} p_{X,Y,Z}(x, y, z)$$

More than Two Random Variables (2/2)

- The expectation for the function of random variables X, Y and Z

$$\mathbf{E}\left[g(X,Y,Z)\right] = \sum_x \sum_y \sum_z g(x,y,z) p_{X,Y,Z}(x,y,z)$$

  - If the function is linear and has the form $aX + bY + cZ + d$

$$\mathbf{E}\left[aX + bY + cZ + d\right] = aE[X] + bE[Y] + cE[Z] + d$$

    - A generalization to more than three random variables

$$\mathbf{E}\left[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n\right] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_n E[X_n]$$

An Illustrative Example

- **Example 2.10. Mean of the Binomial.** Your probability class has 300 students and each student has probability 1/3 of getting an A, independently of any other student.
  - What is the mean of X, the number of students that get an A?

Let

$$X_i = \begin{cases} 1, & \text{if the } i\text{th student gets an A} \\ 0, & \text{otherwise} \end{cases}$$

$\Rightarrow X_1, X_2, \ldots, X_{300}$ are bernoulli random variables with common mean $p = 1/3$

Their sum $X = X_1 + X_2 + \ldots + X_{300}$ can be interpreted as a binomial random variable with parameters n $(n = 300)$ and p $(p = 1/3)$. That is, X is the number of success in n $(n = 300)$ independent trials

$$\therefore \mathbf{E}[X] = \mathbf{E}[X_1 + X_2 + \ldots + X_{300}] = \sum_{i=1}^{300} \mathbf{E}[X_i] = 300 \cdot 1/3 = 100$$

Conditioning

- Recall that conditional probability provides us with a way to reason about the outcome of an experiment, based on partial information

- In the same spirit, we can define **conditional PMFs**, given the occurrence of a certain event or given the value of another random variable

Conditioning a Random Variable on an Event (1/2)

- The **conditional PMF** of a random variable X, conditioned on a particular event A with $\mathbf{P}(A) > 0$, is defined by (where X and A are associated with the same experiment)

$$P_{X|A}(x) = \mathbf{P}(X = x|A) = \frac{\mathbf{P}(\{X = x\} \cap A)}{\mathbf{P}(A)}$$

- Normalization Property
  - Note that the events $\mathbf{P}(\{X = x\} \cap A)$ are disjoint for different values of X, their union is A

$$\mathbf{P}(A) = \sum_x \mathbf{P}(\{X = x\} \cap A)$$

Total probability theorem

$$\therefore \sum_x P_{X|A}(x) = \sum_x \frac{\mathbf{P}(\{X = x\} \cap A)}{\mathbf{P}(A)} = \frac{\sum_x \mathbf{P}(\{X = x\} \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A)}{\mathbf{P}(A)} = 1$$

Conditioning a Random Variable on an Event (2/2)

- A graphical illustration

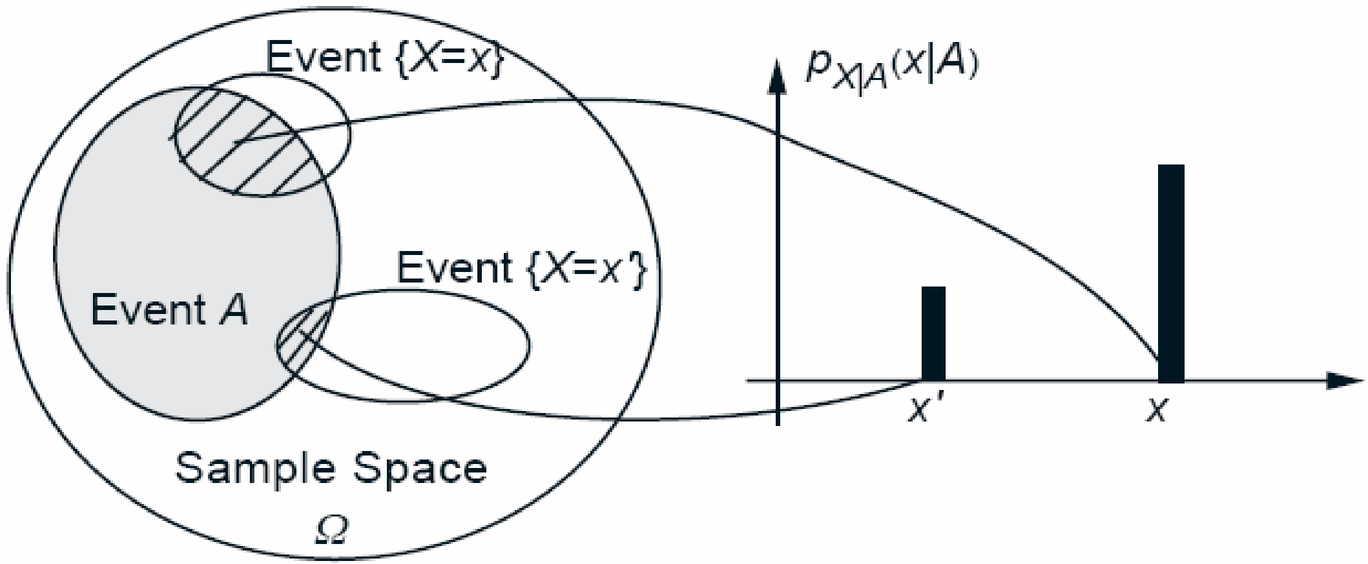


Figure 2.12: Visualization and calculation of the conditional PMF $p_{X|A}(x)$. For each x, we add the probabilities of the outcomes in the intersection $\{X = x\} \cap A$ and normalize by diving with $\mathbf{P}(A)$.

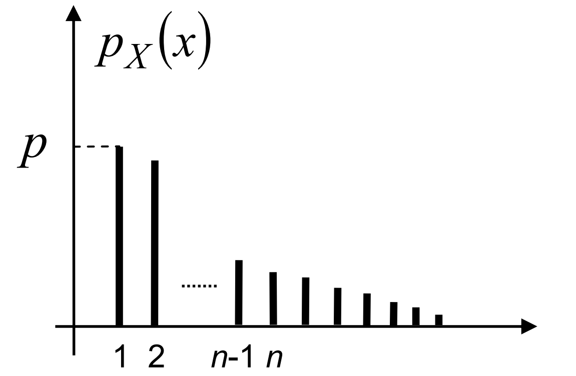
Illustrative Examples (1/2)

- **Example 2.12.** Let X be the roll of a fair six-sided die and A be the event that the roll is an even number

$$
\begin{aligned}
P_{X|A}(x) &= \mathbf{P}(X = x \mid \text{roll is even}) \\
&= \frac{\mathbf{P}(X = x \text{ and } X \text{ is even})}{\mathbf{P}(X \text{ is even})} \\
&= \begin{cases} 1/3, & \text{if } x = 2,4,6 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

Illustrative Examples (2/2)

- **Example 2.14.** A student will take a certain test repeatedly, up to a maximum of n times, each time with a probability p of passing, independently of the number of previous attempts.
  - What is the PMF of the number of attempts given that the student passes the test ?

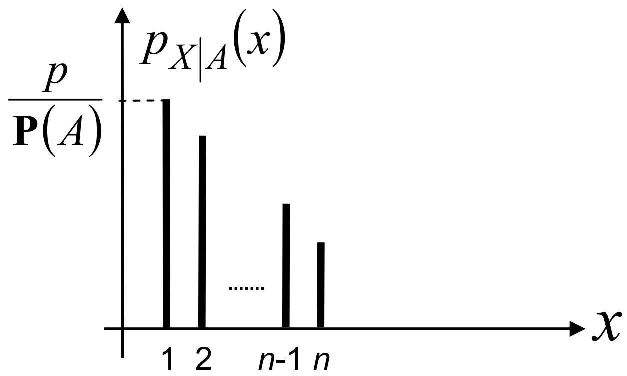


Let X be a geometric random variable with parameter p,
representi ng the number of attempts until the fist success comes up

$$p_X(x) = (1-p)^{x-1}p$$

Let A be the event that the student pass the test w ithin n attempts ($A = \{X \le n\}$)

$$\therefore\ p_{X|A}(x) = \begin{cases} \dfrac{(1-p)^{x-1}p}{\sum_{m=1}^{n}(1-p)^{m-1}p}, & \text{if } x = 1,2,\dots,n \\ 0, & \text{otherwise} \end{cases}$$



Conditioning a Random Variable on Another (1/2)

- Let X and Y be two random variables associated with the same experiment. The conditional PMF $p_{X|Y}$ of X given Y is defined as

$$p_{X|Y}(x|y) = \mathbf{P}(X = x | Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)}$$

$$= \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Y is fixed on some value y

- Normalization Property $\sum_x p_{X|Y}(x|y) = 1$

- The conditional PMF is often convenient for the calculation of the joint PMF

multiplication (chain) rule

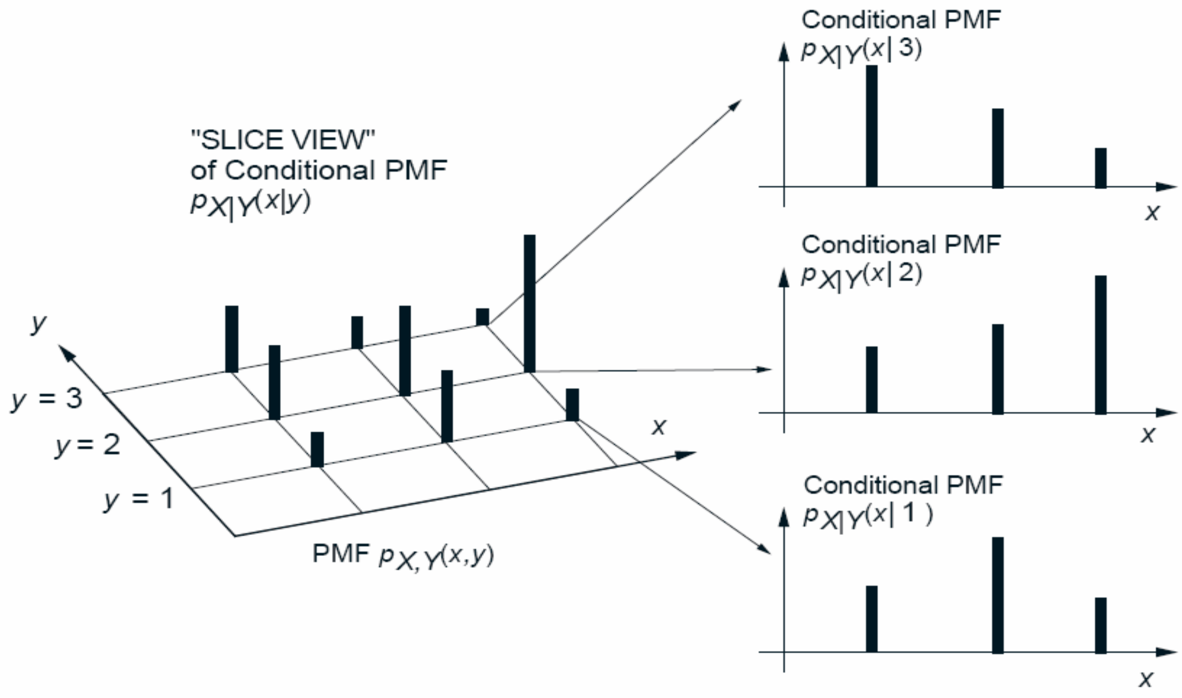
$$p_{X,Y}(x, y) = p_Y(y) p_{X|Y}(x|y) \; (= p_X(x) p_{Y|X}(y|x))$$

Conditioning a Random Variable on Another (2/2)

- The conditional PMF can also be used to calculate the marginal PMFs

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p_Y(y) p_{X|Y}(x|y)$$

- Visualization of the conditional PMF $p_{X|Y}$



$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{X,Y}(x,y)}{\sum_x p_{X,Y}(x,y)}$$

An Illustrative Example (1/2)

- **Example 2.14.** Professor May B. Right often has her facts wrong, and answers each of her students' questions incorrectly with probability 1/4, independently of other questions. In each lecture May is asked 0, 1, or 2 questions with equal probability 1/3.
  - What is the probability that she gives at least one wrong answer ?

Let X be the number of questions asked,

Y be the number of questions answered wrong

$$\binom{n}{k} p^k \cdot (1\text{-}p)^{n-k}$$

$$\mathbf{P}(Y \ge 1) = \mathbf{P}(Y = 1) + \mathbf{P}(Y = 2)$$
$$= \mathbf{P}(x = 1, y = 1) + \mathbf{P}(x = 2, y = 1) + \mathbf{P}(x = 2, y = 2)$$

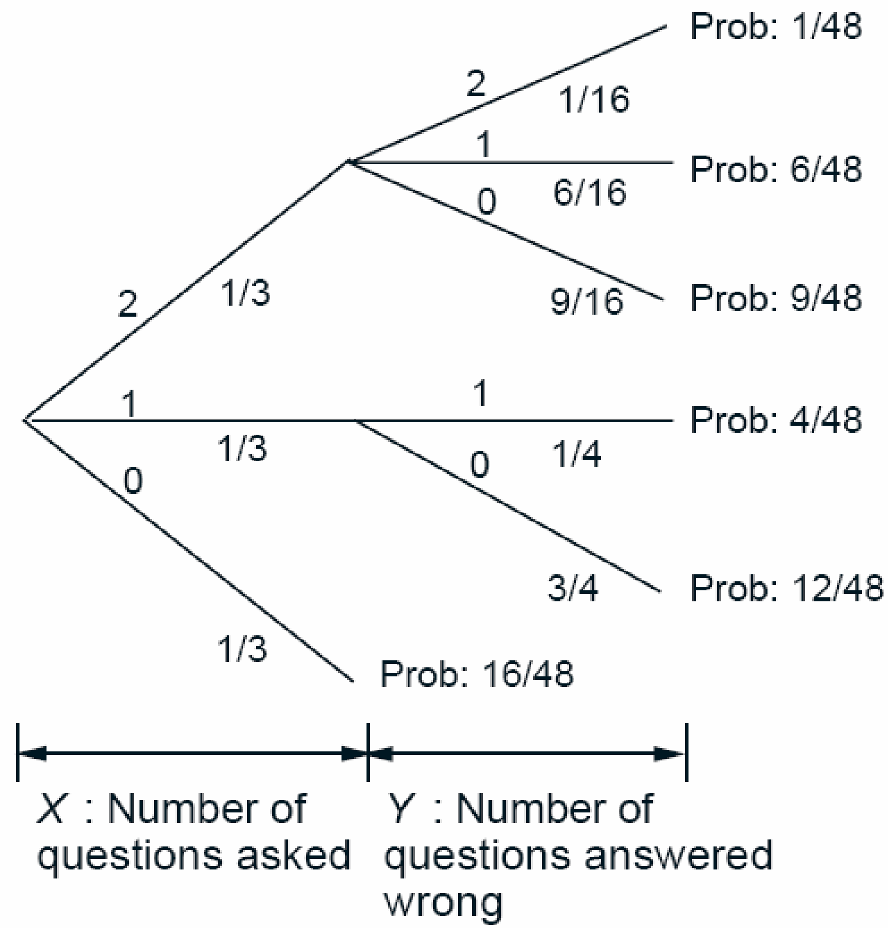
modeled as binomial distributions

$$\therefore \mathbf{P}(Y \ge 1) = \mathbf{P}(x = 1)\mathbf{P}(y = 1 | x = 1) + \mathbf{P}(x = 2)\mathbf{P}(y = 1 | x = 2) + \mathbf{P}(x = 2)\mathbf{P}(y = 2 | x = 2)$$

$$= \frac{1}{3}\cdot\frac{1}{4} + \frac{1}{3}\cdot\left[\binom{2}{1}\frac{1}{4}\cdot\frac{3}{4}\right] + \frac{1}{3}\cdot\left[\binom{2}{2}\frac{1}{4}\cdot\frac{1}{4}\right] = \frac{11}{48}$$

An Illustrative Example (2/2)

- Calculation of the joint PMF $p_{X,Y}(x,y)$ in Example 2.14.



y \ x	0	1	2
2	0	0	1/48
1	0	4/48	6/48
0	16/48	12/48	9/48

Joint PMF $P_{X,Y}(x,y)$ in tabular form

Conditional Expectation

- Recall that a conditional PMF can be thought of as an ordinary PMF over a new universe determined by the conditioning event

- In the same spirit, a conditional expectation is the same as an ordinary expectation, except that it refers to the new universe, and all probabilities and PMFs are replaced by their conditional counterparts

Summary of Facts About Conditional Expectations

- Let X and Y be two random variables associated with the same experiment
  - The conditional expectation of X given an event A with $\mathbf{P}(A) > 0$, is defined by

$$\mathbf{E}[X \mid A] = \sum_{x} x p_{X \mid A}(x)$$

    - For a function $g(X)$, it is given by

$$\mathbf{E}[g(X) \mid A] = \sum_{x} g(x) p_{X \mid A}(x)$$

Total Expectation Theorem (1/2)

- The conditional expectation of X given a value y of Y is defined by

$$\mathbf{E}[X \mid Y = y] = \sum_x x p_{X|Y}(x|y)$$

  - We have

$$\mathbf{E}[X] = \sum_y p_Y(y) \sum_x x p_{X|Y}(x|y)$$

- Let $A_1, \cdots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i . Then,

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{P}(A_i)\, \mathbf{E}[X \mid A_i]$$

Total Expectation Theorem (2/2)

- Let $A_1, \cdots, A_n$ be disjoint events that form a partition of an event B, and assume that $P(A_i \cap B) > 0$, for all i. Then,

$$\mathbf{E}[X \mid B] = \sum_{i=1}^{n} \mathbf{P}(A_i \mid B)\,\mathbf{E}[X \mid A_i \cap B]$$

- Verification of total expectation theorem

$$\begin{aligned}
\mathbf{E}[X] &= \sum_x x p_X(x) = \sum_x x \sum_y p_{X,Y}(x, y) \\
&= \sum_x x \sum_y p_Y(y) p_{X|Y}(x \mid y) \\
&= \sum_y p_Y(y) \sum_x x\, p_{X|Y}(x \mid y) \\
&= \sum_y p_Y(y)\,\mathbf{E}[X \mid Y = y]
\end{aligned}$$

An Illustrative Example (1/2)

- **Example 2.17.** Mean and Variance of the Geometric Random Variable
  - A geometric random variable X has PMF $p_X(x)=(1-p)^{x-1}p, \quad x=1,2,\ldots$

Let A_1 be the event that $\{X=1\}$

A_2 be the event that $\{X>1\}$

$$\mathbf{E}[X]=\mathbf{P}(A_1)\mathbf{E}[X|A_1]+\mathbf{P}(A_2)\mathbf{E}[X|A_2]$$

where

$$\mathbf{P}(A_1)=p,\ \mathbf{P}(A_2)=1-p \quad (??)$$

$$p_{X|A_1}(x)=\begin{cases}\dfrac{p}{p}=1, & x=1\\ 0, & \text{otherwise}\end{cases}$$

$$p_{X|A_2}(x)=\begin{cases}(1-p)^{x-2}p \quad (??), & x>1\\ 0, & \text{otherwise}\end{cases}$$

Note that (See Example 2.13) :

$$p_{X|A}(x)=\begin{cases}\dfrac{(1-p)^{x-1}p}{\sum_{m=1}^{n}(1-p)^{m-1}p}, & \text{if } x=1,2,\ldots,n\\ 0, & \text{otherwise}\end{cases}$$

$$\mathbf{E}[X|A_1]=1\cdot 1+\sum_{x=2}^{\infty}x\cdot 0=1$$

$$\begin{aligned}\mathbf{E}[X|A_2]&=1\cdot 0+\sum_{x=2}^{\infty}x\cdot\left[(1-p)^{x-2}p\right]\\&=\sum_{x=2}^{\infty}x\cdot\left[(1-p)^{x-2}p\right]\\&=\sum_{x'=1}^{\infty}(x'+1)(1-p)^{x'-1}p\\&=\left[\sum_{x'=1}^{\infty}x'(1-p)^{x'-1}p\right]+\underbrace{\left[\sum_{x'=1}^{\infty}(1-p)^{x'-1}p\right]}_{1}\\&=\mathbf{E}[X]+1\end{aligned}$$

$$\begin{aligned}\Rightarrow \mathbf{E}[X]&=\mathbf{P}(A_1)\mathbf{E}[X|A_1]+\mathbf{P}(A_2)\mathbf{E}[X|A_2]\\&=\mathbf{P}(A_1)\cdot 1+(1-p)(\mathbf{E}[X]+1)\end{aligned}$$

$$\therefore\ \mathbf{E}[X]=\frac{1}{p}$$

An Illustrative Example (2/2)

$$\mathbf{E}\left[X^2\right] = \mathbf{P}(A_1)\mathbf{E}\left[X^2 | A_1\right] + \mathbf{P}(A_2)\mathbf{E}\left[X^2 | A_2\right]$$

$$\mathbf{E}\left[X^2 | A_1\right] = 1^2 \cdot 1 + \sum_{x=2}^{\infty} x^2 \cdot 0 = 1$$

$$\mathbf{E}\left[X^2 | A_2\right] = 1^2 \cdot 0 + \sum_{x=2}^{\infty} x^2 \cdot (1-p)^{x-2} p \qquad \left(\because\ x^2 = (x-1)^2 + 2x - 1\right)$$

$$= \left[\sum_{x=2}^{\infty} (x-1)^2 \cdot (1-p)^{x-2} p\right] + 2\left[\sum_{x=2}^{\infty} x \cdot (1-p)^{x-2} p\right] - \left[\sum_{x=2}^{\infty} (1-p)^{x-2} p\right]$$

$$= \left[\sum_{x'=1}^{\infty} x'^2 \cdot (1-p)^{x'-1} p\right] + 2\left[\sum_{x=2}^{\infty} (x-1) \cdot (1-p)^{x-2} p\right] + 2\left[\sum_{x=2}^{\infty} (1-p)^{x-2} p\right] - \left[\sum_{x=2}^{\infty} (1-p)^{x-2} p\right]$$

$$= \mathbf{E}\left[X^2\right] + 2\left[\sum_{x'=1}^{\infty} x' \cdot (1-p)^{x'-1} p\right] + \left[\sum_{x'=1}^{\infty} (1-p)^{x'-1} p\right] \qquad \left(\text{set } x' = x-1\right)$$

$$= \mathbf{E}\left[X^2\right] + 2\mathbf{E}[X] + 1$$

$$\Rightarrow \mathbf{E}\left[X^2\right] = p \cdot 1 + (1-p)\left(\mathbf{E}\left[X^2\right] + 2\mathbf{E}[X] + 1\right)$$

$$\mathbf{E}\left[X^2\right] = \frac{1 + 2(1-p)\mathbf{E}[X]}{p} \qquad \left(\text{we have shown that } \quad \mathbf{E}[X] = \frac{1}{p}\right)$$

$$\mathbf{E}\left[X^2\right] = \frac{2}{p^2} - \frac{1}{p}$$

$$\therefore \operatorname{var}(X) = \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2 = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$$

Independence of a Random Variable from an Event

- A random variable X is **independent of an event** A if

$$\mathbf{P}(X = x \text{ and } A) = \mathbf{P}(X = x)\mathbf{P}(A), \quad \text{for all } x$$

- If a random variable X is **independent of an event** A and $\mathbf{P}(A) > 0$

$$\begin{aligned} p_{X|A}(x) &= \frac{\mathbf{P}(X = x \text{ and } A)}{\mathbf{P}(A)} \\ &= \frac{\mathbf{P}(X = x)\mathbf{P}(A)}{\mathbf{P}(A)} \\ &= \mathbf{P}(X = x) \\ &= p_X(x), \quad \text{for all } x \end{aligned}$$

An Illustrative Example

- **Example 2.19.** Consider two independent tosses of a fair coin.
  - Let random variable X be the number of heads
  - Let random variable Y be 0 if the first toss is head, and 1 if the first toss is tail
  - Let A be the event that the number of head is even
    - Possible outcomes (T,T), (T,H), (H,T), (H,H)

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

$$p_{X|A}(x) = \begin{cases} 1/2, & \text{if } x = 0 \\ 0, & \text{if } x = 1 \\ 1/2, & \text{if } x = 2 \end{cases}$$

$$p_{X|A}(x) \neq p_X(x) \Rightarrow X \text{ and } A \text{ are not independent!}$$

$$p_Y(y) = \begin{cases} 1/2, & \text{if } y = 0 \\ 1/2, & \text{if } y = 1 \end{cases}$$

$$p_{Y|A}(y) = \frac{\mathbf{P}(Y = y \text{ and A})}{\mathbf{P}(A)} = \begin{cases} 1/2, & \text{if } y = 0 \\ 1/2, & \text{if } y = 1 \end{cases}$$

$$\mathbf{P}(A) = 1/2$$

$$p_{Y|A}(y) = p_Y(y) \Rightarrow Y \text{ and } A \text{ are independent!}$$

Independence of a Random Variables (1/2)

- Two **random variables** X and Y are **independent** if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y), \quad \text{for all } x, y$$

$$\text{or} \quad \mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y), \quad \text{for all } x, y$$

- If a random variable X is **independent of an random variable** Y

$$p_{X|Y}(x|y) = p_X(x), \quad \text{for all } y \text{ with } p_Y(y) > 0 \text{ all } x$$

$$\begin{aligned} p_{X|Y}(x|y) &= \frac{p_{X,Y}(x,y)}{p_Y(y)} \\ &= \frac{p_X(x)p_Y(y)}{p_Y(y)} \\ &= p_X(x), \quad \text{for all } y \text{ with } p(y) > 0 \text{ and all } x \end{aligned}$$

Independence of a Random Variables (2/2)

- Random variables X and Y are said to be **conditionally independent**, given a positive probability event A, if

$$p_{X,Y|A}(x,y) = p_{X|A}(x)\,p_{Y|A}(y), \quad \text{for all } x, y$$

  – Or equivalently,

$$p_{X|Y,A}(x|y) = p_{X|A}(x), \quad \text{for all } y \text{ with } p_{Y|A}(y) > 0 \text{ and all } x$$

- Note here that, as in the case of events, conditional independence may not imply unconditional independence and vice versa

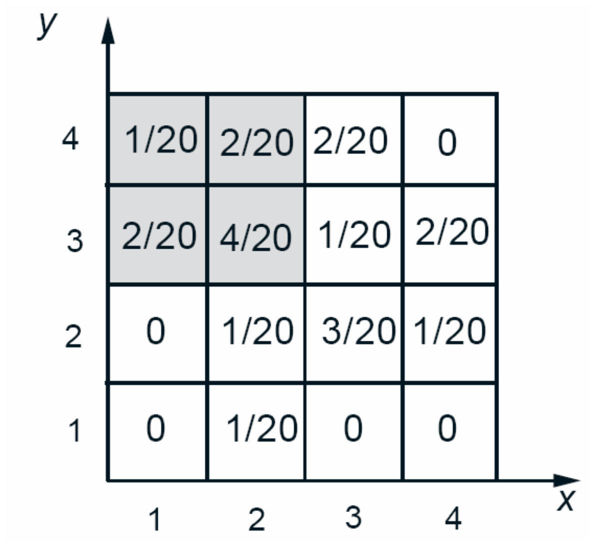
An Illustrative Example (1/2)

- **Figure 2.15:** Example illustrating that conditional independence may not imply unconditional independence
  - For the PMF shown, the random variables X and Y are not independent
    - To show X and Y are not independent, we only have to find a pair of values (x, y) of X and Y that

$$p_{X|Y}(x|y) \neq p_X(x)$$

  - For example, X and Y are not independent

$$p_{X|Y}(1|1) = 0 \neq p_X(1) = \frac{3}{20}$$



An Illustrative Example (2/2)

- To show X and Y are not dependent, we only have to find all pair of values (x, y) of X and Y that

$$p_{X|Y}(x|y) = p_X(x)$$

  – For example, X and Y are independent, conditioned on the event $A = \{X \leq 2, Y \geq 3\}$

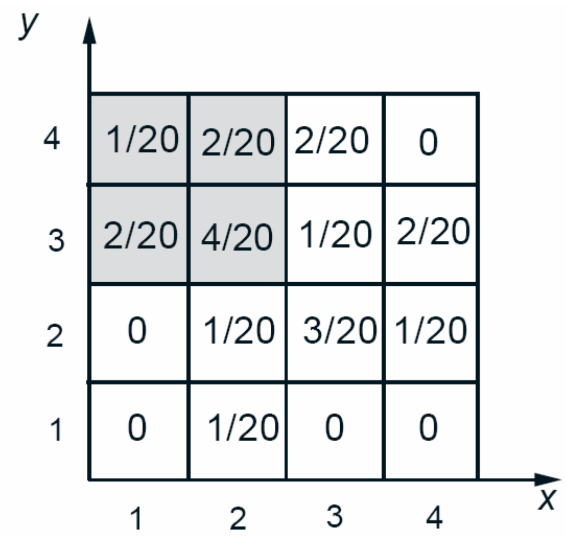
$$\mathbf{P}(A) = \frac{9}{20}, \quad p_{X|Y,A}(x|y) = \frac{\mathbf{P}(X = x \cap Y = y \cap A)}{\mathbf{P}(Y = y \cap A)}$$

$$p_{X|Y,A}(1|3) = \frac{2/20}{6/20} = \frac{1}{3}, \quad p_{X|A}(1) = \frac{3/20}{9/20} = 1/3$$

$$p_{X|Y,A}(1|4) = \frac{1/20}{3/20} = \frac{1}{3}$$

$$p_{X|Y,A}(2|3) = \frac{4/20}{6/20} = \frac{2}{3}, \quad p_{X|A}(2) = \frac{6/20}{9/20} = 2/3$$

$$p_{X|Y,A}(2|4) = \frac{2/20}{3/20} = \frac{2}{3}$$



Functions of Two Independent Random Variables

- Given X and Y be two independent random variables, let $g(X)$ and $h(Y)$ be two functions of X and Y, respectively. Show that $g(X)$ and $h(Y)$ are independent.

Let $U = g(X)$ and $V = h(Y)$, then

$$
\begin{aligned}
p_{U,V}(u,v) &= \sum_{\{(x,y) \mid g(x)=u,\, h(y)=v\}} p_{X,Y}(x,y) \\
&= \sum_{\{(x,y) \mid g(x)=u,\, h(y)=v\}} p_X(x)\, p_Y(y) \\
&= \sum_{\{x \mid g(x)=u\}} p_X(x) \sum_{\{y \mid h(y)=v\}} p_Y(y) \\
&= p_U(u)\, p_V(v)
\end{aligned}
$$

More Factors about Independent Random Variables (1/2)

- If X and Y are independent random variables, then

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$$

  - As shown by the following calculation

$$\begin{aligned}\mathbf{E}[XY] &= \sum_x \sum_y xy\underline{p_{X,Y}(x,y)} \\ &= \sum_x \sum_y xy\underline{p_X(x)p_Y(y)} \\ &= \sum_x xp_X(x)\left[\sum_y yp_Y(y)\right] \\ &= \mathbf{E}[X]\mathbf{E}[Y]\end{aligned}$$

by independence

- Similarly, if X and Y are independent random variables, then

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)]$$

More Factors about Independent Random Variables (2/2)

- If X and Y are independent random variables, then

$$\mathrm{var}(X+Y)=\mathrm{var}(X)+\mathrm{var}(Y)$$

  - As shown by the following calculation

$$\begin{aligned}
\mathrm{var}(X+Y)&=\mathbf{E}\left[\left((X+Y)-\mathbf{E}[X+Y]\right)^2\right]\\
&=\mathbf{E}\left[(X+Y)^2-2(X+Y)(\mathbf{E}[X]+\mathbf{E}[Y])+(\mathbf{E}[X]+\mathbf{E}[Y])^2\right]\\
&=\left[\sum_{x,y}(x+y)^2 p_{X,Y}(x,y)\right]-2(\mathbf{E}[X]+\mathbf{E}[Y])\mathbf{E}[X]-2(\mathbf{E}[X]+\mathbf{E}[Y])\mathbf{E}[Y]+\\
&\quad+(\mathbf{E}[X])^2+2\cdot\mathbf{E}[X]\mathbf{E}[Y]+(\mathbf{E}[Y])^2\\
&=\left[\sum_{x,y}x^2 p_{X,Y}(x,y)\right]+\left[\sum_{x,y}y^2 p_{X,Y}(x,y)\right]+\cancel{2\left[\sum_{x,y}xyp_{X,Y}(x,y)\right]}\\
&\quad-(\mathbf{E}[X])^2-(\mathbf{E}[Y])^2\cancel{-2\mathbf{E}[X]\mathbf{E}[Y]}\\
&=\left(\mathbf{E}\left[X^2\right]-(\mathbf{E}[X])^2\right)+\left(\mathbf{E}\left[Y^2\right]-(\mathbf{E}[Y])^2\right)=\mathrm{var}(X)+\mathrm{var}(Y)
\end{aligned}$$

More than Two Random Variables

- Independence of several random variables
  - Three random variable X , Y and Z are independent if

$$p_{X,Y,Z}(x, y, z) = p_X(x) p_Y(y) p_Z(z)$$

    - Any three random variables of the form $f(X)$, $g(X)$ and $h(X)$ are also independent

- Variance of the sum of independent random variables
  - If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n)$$

Illustrative Examples (1/3)

- **Example 2.20. Variance of the Binomial.** We consider n independent coin tosses, with each toss having probability p of coming up a head. For each i, we let X_i be the Bernoulli random variable which is equal to 1 if the *i*-th toss comes up a head, and is 0 otherwise.
  - Then, $X = X_1 + X_2 + \cdots + X_n$ is a binomial random variable.

$$\because\ \mathrm{var}(X_i) = p(1-p), \quad \text{for all } i$$

$$\therefore\ \mathrm{var}(X) = \sum_{i=1}^{n} \mathrm{var}(X_i) = np(1-p) \qquad (\text{Note that } X_i\text{'s are independent!})$$

Illustrative Examples (2/3)

- **Example 2.21. Mean and Variance of the Sample Mean.** We wish to estimate the approval rating of a president, to be called B. To this end, we ask n persons drawn at random from the voter population, and we let X_i be a random variable that encodes the response of the *i*-th person:

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th person approves B's performance} \\ 0, & \text{if the } i\text{-th person disapproves B's performance} \end{cases}$$

  - Assume that X_i independent, and are the same random variable (Bernoulli) with the common parameter (p for Bernoulli), which is unknown to us
    - X_i are independent, and identically distributed (i.i.d.)
  - If the sample mean S_n (is a random variable) is defined as

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Illustrative Examples (3/3)

– The expectation of S_n will be the true mean of X_i

$$\mathbf{E}[S_n] = \mathbf{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i]$$
$$= \mathbf{E}[X_i] \text{ (} = p \text{ for the Bernoulli we assumed here)}$$

– The variance of S_n will approximate 0 if n is large enough

$$\lim_{n \to \infty} \operatorname{var}(S_n) = \operatorname{var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$
$$= \lim_{n \to \infty} \frac{\sum_{i=1}^{n}\operatorname{var}(X_i)}{n^2} = \lim_{n \to \infty} \frac{np(1-p)}{n^2} = \lim_{n \to \infty} \frac{p(1-p)}{n} = 0$$

- Which means that S_n will be a good estimate of $\mathbf{E}[X_i]$ if n is large enough

Recitation

- SECTION 2.5 Joint PMFs of Multiple Random Variables
  - Problems 27, 28, 30
- SECTION 2.6 Conditioning
  - Problems 33, 34, 35, 37
- SECTION 2.6 Independence
  - Problems 42, 43, 45, 46