CS145: Probability & Computing

Lecture 7: Multiple Discrete Variables: Independence, Expectation

Instructors: Alessio Mazzetto and Eli Upfal

Brown University Computer Science

IN DEO SPERAMUS

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

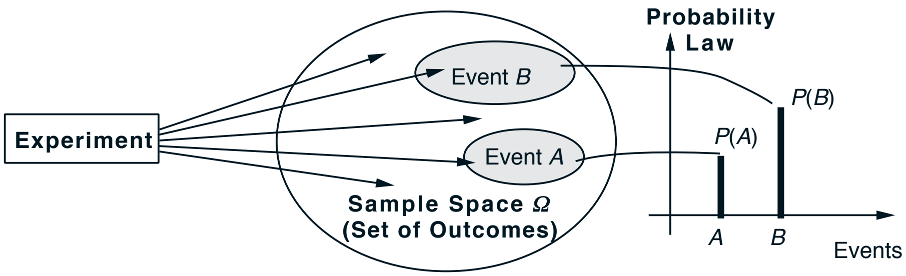
CS145: Lecture 6 Outline

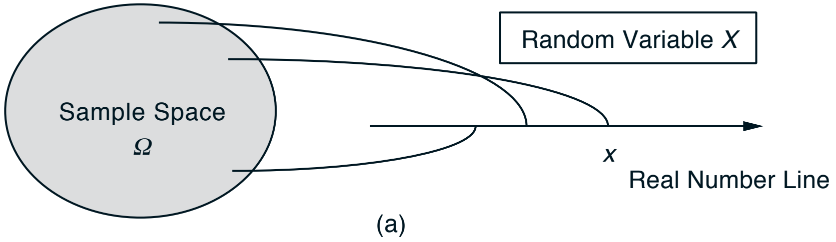
- Independent random variables
- Expectations of multiple discrete variables

Discrete Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$





- The *range* of a random variable is the set of values with positive probability

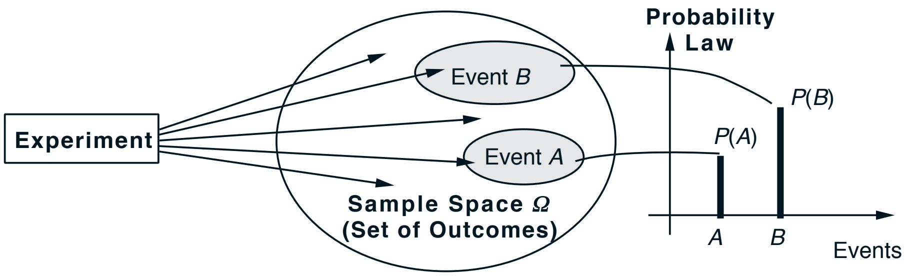
$$\mathcal{X} = \{x \in \mathbb{R} \mid X(\omega) = x \text{ for some } \omega \in \Omega, P(\omega) > 0\}$$

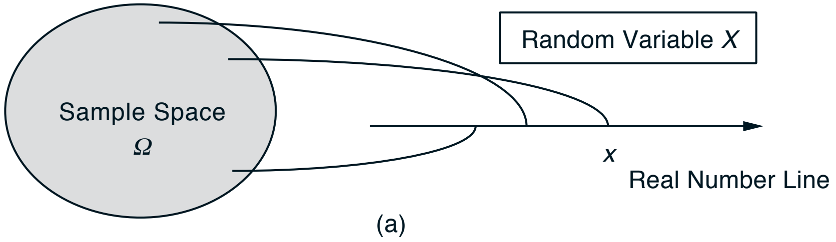
For a *discrete random variable*, the range is finite or countably infinite (we can map it to the integers). *Coming later: continuous random variables.*

Discrete Random Variables

- A *random variable* assigns values to outcomes of uncertain experiments

$$X : \Omega \to \mathbb{R} \qquad x = X(\omega) \in \mathbb{R} \text{ for } \omega \in \Omega$$





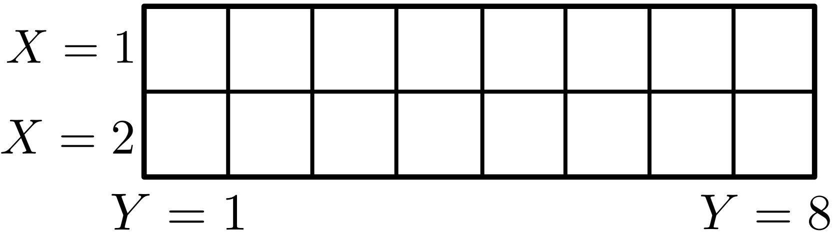
- The *probability mass function (PMF)* or *probability distribution* of variable:

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

$$p_X(x) \geq 0, \qquad \sum_{x \in \mathcal{X}} p_X(x) = 1.$$

If range is finite, this is a vector of non-negative numbers that sums to one.

Joint Probability Distributions



In this example, N=2 and M=8, and the joint PMF is a 2x8 matrix.

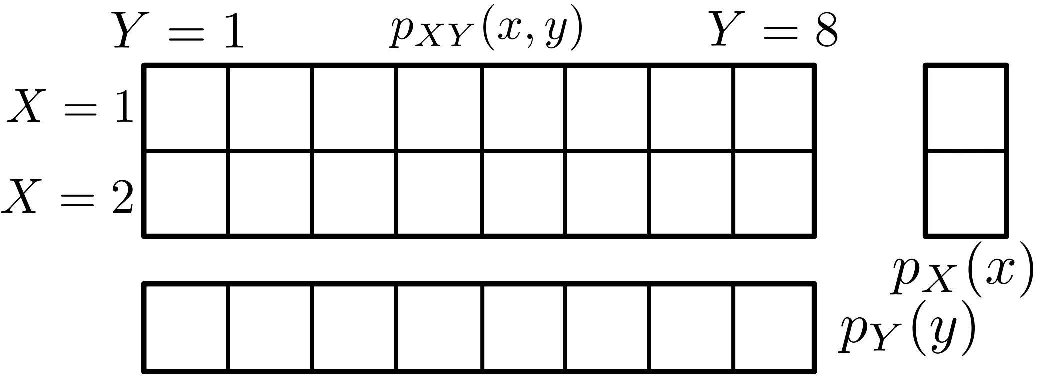
- Consider two random variables *X, Y.*
  Suppose range of X is size N, range of Y is size M.
- The *joint probability mass function* or *joint distribution* of two variables:

$$p_{XY}(x, y) = P(X = x \text{ and } Y = y)$$

$$p_{XY}(x, y) \geq 0, \qquad \sum_x \sum_y p_{XY}(x, y) = 1.$$

- The joint distribution is uniquely specified by *NM-1* numbers

Marginal Probability Distributions



The marginal distributions are defined by (N-1)+(M-1) numbers.
Many joint distributions may have the same marginals.

- The *joint probability mass function* or *joint distribution* of two variables:

$$p_{XY}(x, y) = P(X = x \text{ and } Y = y)$$

- The range of each variable defines a partition of the sample space, so the *marginal distributions* can be computed from the joint distribution:

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

Marginal Probability Distributions

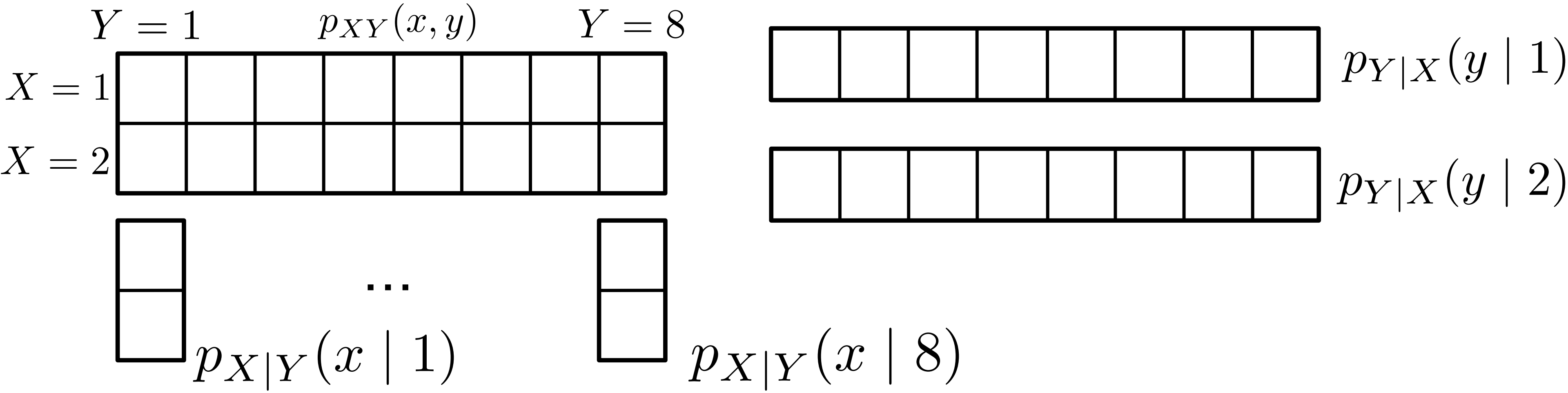
Joint PMF $P_{X,Y}(x,y)$ in tabular form

y					
4	0	1/20	1/20	1/20	3/20
3	1/20	2/20	3/20	1/20	7/20
2	1/20	2/20	3/20	1/20	7/20
1	1/20	1/20	1/20	0	3/20
	1	2	3	4	x
	3/20	6/20	8/20	3/20	

Row Sums: Marginal PMF $P_Y(y)$

Column Sums: Marginal PMF $P_X(x)$

Conditional Probability Distributions



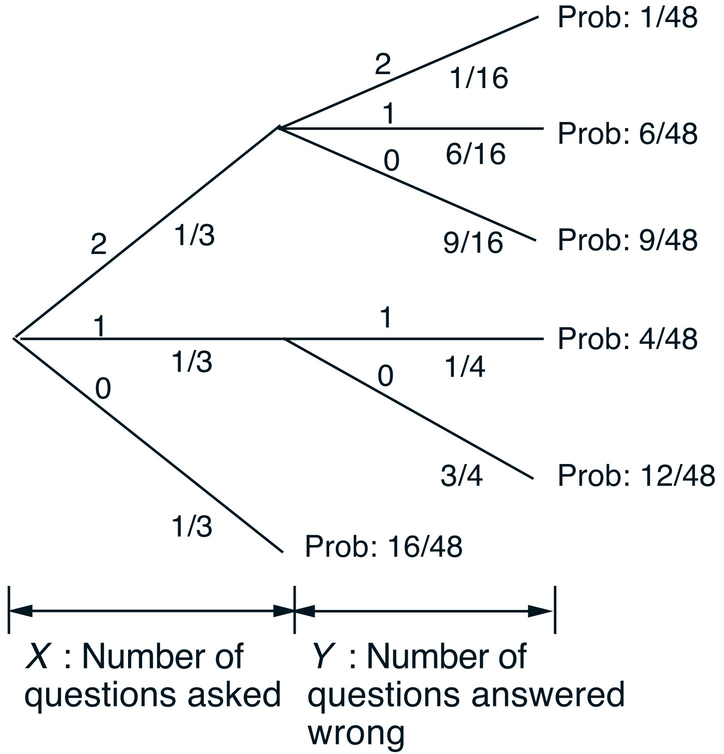
- By the definition of conditional probability:

$$P(X=x \mid Y=y) = \frac{P(X=x \text{ and } Y=y)}{P(Y=y)}$$

- The *conditional probability mass function* is then:

$$p_{X|Y}(x \mid y) = P(X=x \mid Y=y) = \frac{p_{XY}(x,y)}{p_Y(y)} = \frac{p_{XY}(x,y)}{\sum_{x'} p_{XY}(x',y)}$$

Example: The Absent-Minded Prof



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

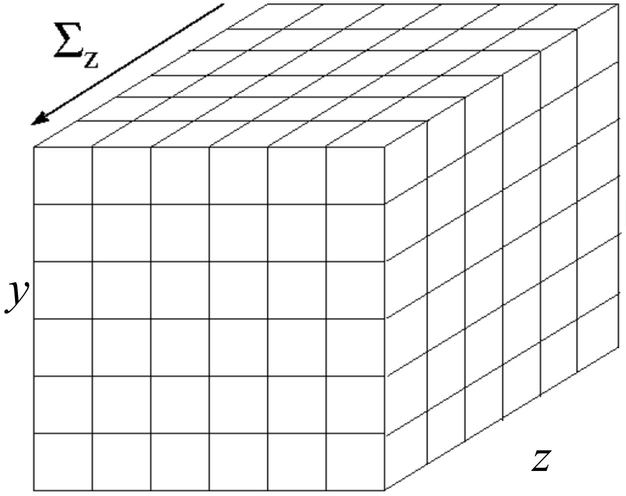
Joint PMF $P_{X,Y}(x,y)$ in tabular form

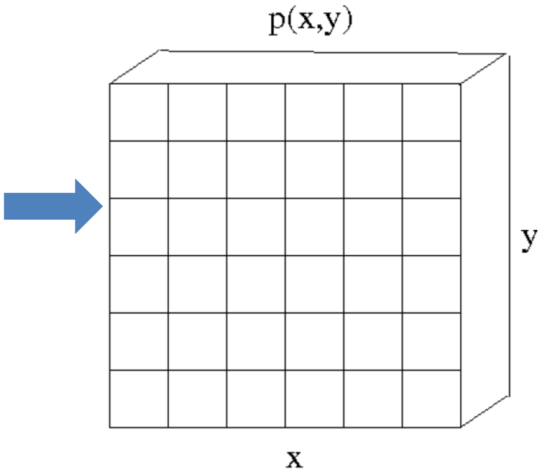
$$\mathbf{P}(\text{at least one wrong answer}) = p_{X,Y}(1,1) + p_{X,Y}(2,1) + p_{X,Y}(2,2) = \frac{4}{48} + \frac{6}{48} + \frac{1}{48}.$$

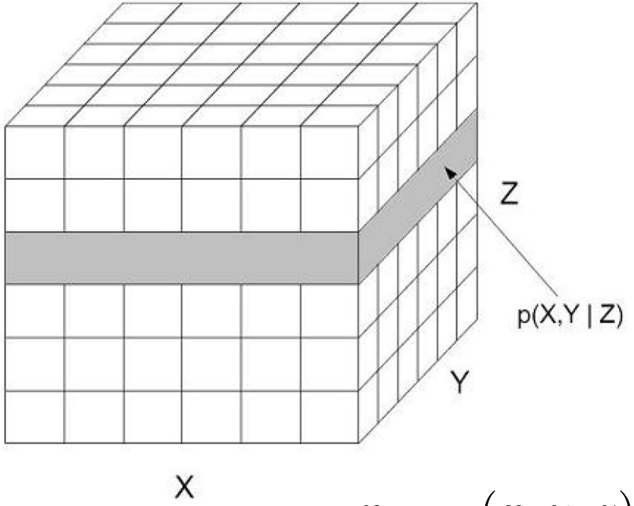
- At office hours, a Professor gets 0, 1, or 2 questions with equal probability
- Each question is answered correctly with probability ¾ (independently)

Several Random Variables

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







$$p_{XY}(x, y) = \sum_{z \in \mathcal{Z}} p_{XYZ}(x, y, z) \qquad p_X(x) = \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \qquad p_{XY|Z}(x, y \mid z) = \frac{p_{XYZ}(x, y, z)}{p_Z(z)}$$

Marginal and conditional define new probability spaces.

May compute marginal and conditioned on any other set of variables.

CS145: Lecture 6 Outline

- Independent random variables
- Expectations of multiple discrete variables

Reminder: Independence of Events

Independence of Two Events: $$P(A \cap B) = P(A)P(B)$$

This implies that $$P(A \mid B) = P(A), P(B \mid A) = P(B).$$

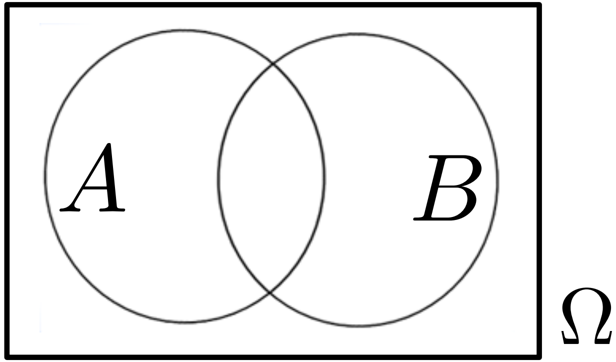
- Observing *B* provides no information about whether *A* occurred
- Observing *A* provides no information about whether *B* occurred

Definition of Conditional Probabilities:

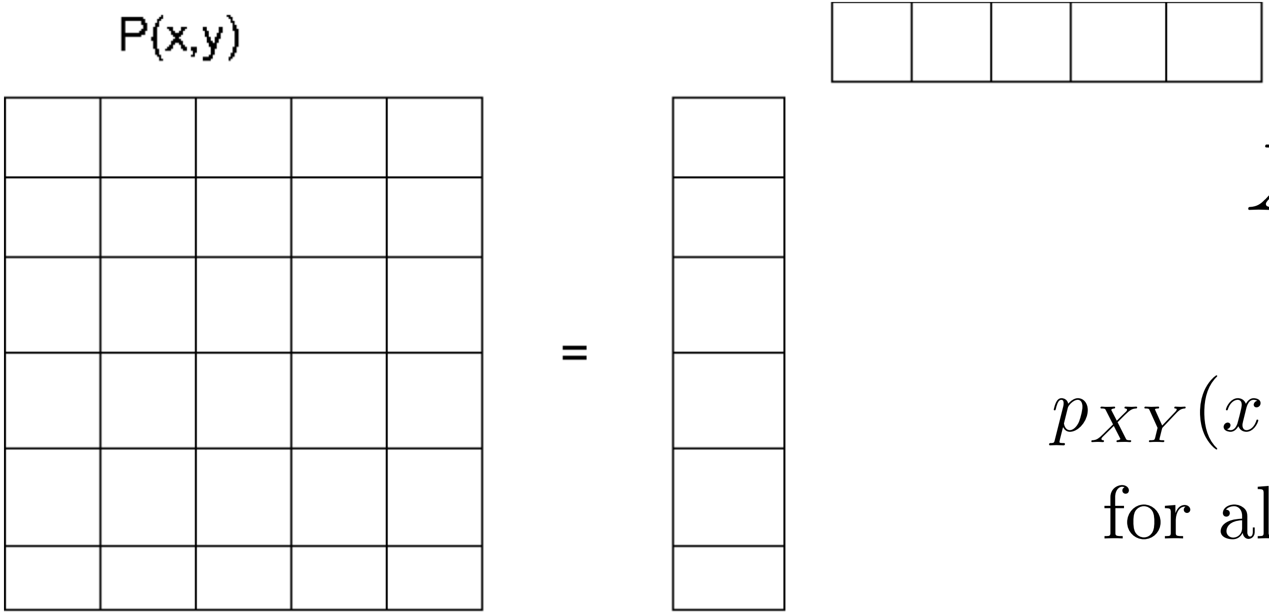
- **Definition:** Assuming $\mathbf{P}(B) \neq 0$,

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

$\mathbf{P}(A \mid B)$ undefined if $\mathbf{P}(B) = 0$



Independent Random Variables



$$X \perp Y$$

$$\Updownarrow$$

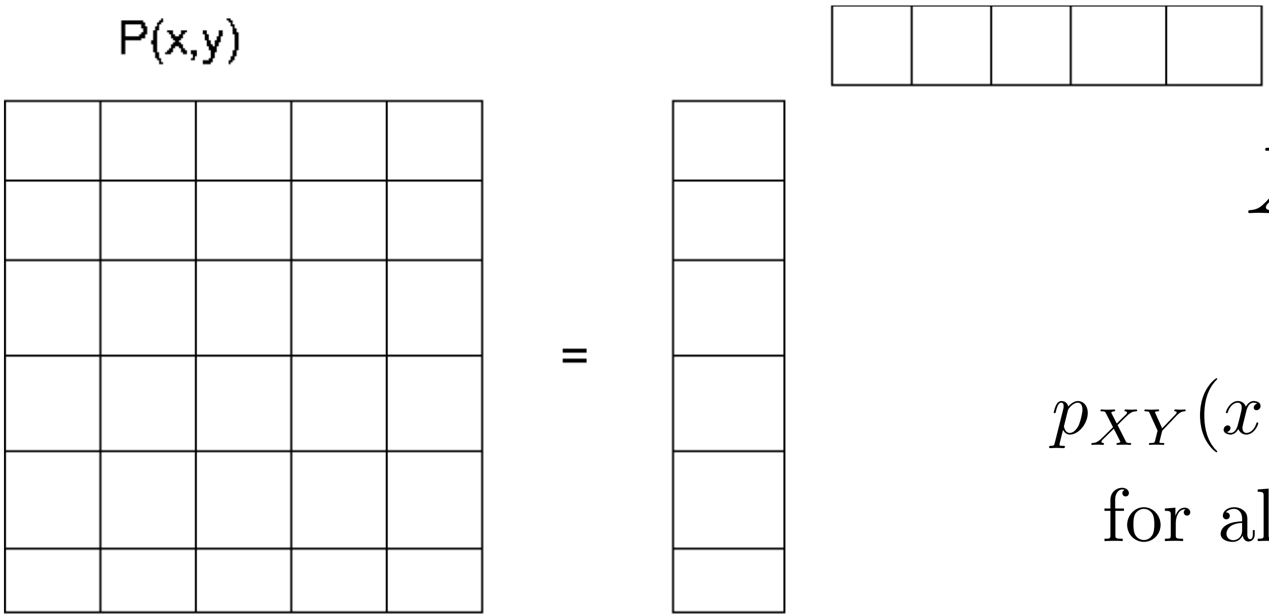
$$p_{XY}(x, y) = p_X(x) p_Y(y)$$
$$\text{for all } x \in \mathcal{X}, y \in \mathcal{Y}$$

- Equivalent conditions on conditional probabilities:

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

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

Independent Random Variables



$$X \perp Y$$

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

- For a given set of marginal distributions, there exists a *unique* joint distribution under which those variables are independent
- Three random variables are independent if and only if

$$p_{XYZ}(x,y,z) = p_X(x)p_Y(y)p_Z(z)$$

Example: Independence

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

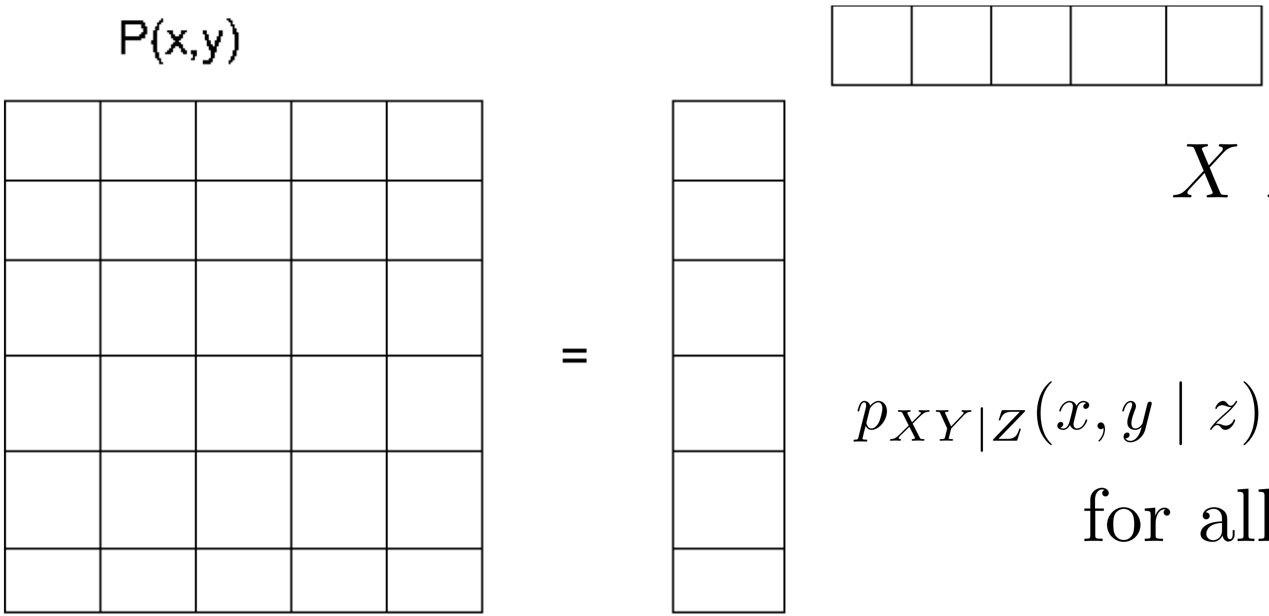
$$p_{XY}(x, y) = P(X = x \text{ and } Y = y)$$

Verify that *X* and *Y*
are *not* independent:

$$p_X(x) =$$

$$p_Y(y) =$$

Conditional Independence



$$X \perp Y \mid Z = z$$

$$p_{XY|Z}(x, y \mid z) = p_{X|Z}(x \mid z) p_Y(y \mid z)$$
$$\text{for all } x \in \mathcal{X}, y \in \mathcal{Y}$$

- Apply the same definition of independence for *X* and *Y*, but *condition* all probability distributions on some other variable *Z*
- Independence does *not* always imply conditional independence, and conditional independence does *not* always imply independence

Example: (Conditional) Independence

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

$$p_{XY}(x, y) = P(X = x \text{ and } Y = y)$$

Verify that *X* and *Y* are *not* independent:

$$p_X(x) =$$

$$p_Y(y) =$$

y \ x	1	2	3	4
4	1/20	2/20	2/20	0
3	2/20	4/20	1/20	2/20
2	0	1/20	3/20	1/20
1	0	1/20	0	0

But *X* and *Y* are *conditionally independent* given

$$Z = 1_{\{X \le 2, Y \ge 3\}}$$

$$p_{X|Z}(x \mid 1) =$$

$$p_{Y|Z}(y \mid 1) =$$

Example:

Consider the following game:
We roll a dice until we obtain an even number.

Define: X = the number of rolls in a game,
Y= the value of the last roll (Y can be either 2, 4 or 6).

1) What is the distribution of (X,Y)?

2) What is the distribution of X?

3) What is the distribution of Y?

4) Are X and Y independent?

CS145: Lecture 6 Outline

- Independent random variables
- Expectations of multiple discrete variables

Expectation

- The *expectation* or *expected value* of a discrete random variable is:

$$E[X] = \sum_{x \in \mathcal{X}} x p_X(x)$$

- The expectation is a single number, not a random variable.
  It encodes the "center of mass" of the probability distribution:

$$x_{\min} \leq E[x] \leq x_{\max}$$

$$x_{\min} = \min\{x \mid x \in \mathcal{X}\}$$

$$x_{\max} = \max\{x \mid x \in \mathcal{X}\}$$

- The expectation is an average or interpolation. It is possible that

$$p_X(E[x]) = 0 \text{ for some random variables } X.$$

Expected Values of Functions

- Consider a *non-random (deterministic) function* of a random variable:

$$Y = g(X)$$

$$p_X(x) = P(X = x) \quad \Rightarrow \quad p_Y(y) = \sum_{\{x | g(x) = y\}} p_X(x)$$

- What is the expected value of random variable *Y*? $E[Y] = E[g(X)]$

- Correct approach #1: $E[Y] = \sum_y y p_Y(y)$

- Correct approach #2: $E[Y] = E[g(X)] = \sum_x g(x) p_X(x)$

- Incorrect approach: $g(E[X]) \neq E[g(X)]$ *(except in special cases)*

Examples

X = 1,2,3 with probability 1/3,1/3,1/3

Y = X^2

$$E[Y] = E[X^2] = \frac{1}{2} + \frac{4}{3} + \frac{9}{3} = 4\frac{2}{3}$$

$$(E[X])^2 = 2^2 = 4$$

$X = -1, 0, +1$ with probabity $1/3, 1/3, 1/3$.
$E[X^2] = \frac{2}{3}$, while $(E[X])^2 = 0$

Expectation of Multiple Variables

- The *expectation* or *expected value* of a function of two discrete variables:

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

- A similar formula applies to functions of 3 or more variables

- Expectations of sums of functions are sums of expectations:

$$E[g(X)+h(Y)] = E[g(X)] + E[h(Y)] = \left[\sum_{x\in\mathcal{X}} g(x) p_X(x)\right] + \left[\sum_{y\in\mathcal{Y}} h(y) p_Y(y)\right]$$

- This is always true, *whether or not X and Y are independent*
- Specializing to *linear functions*, this implies that:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

Examples:

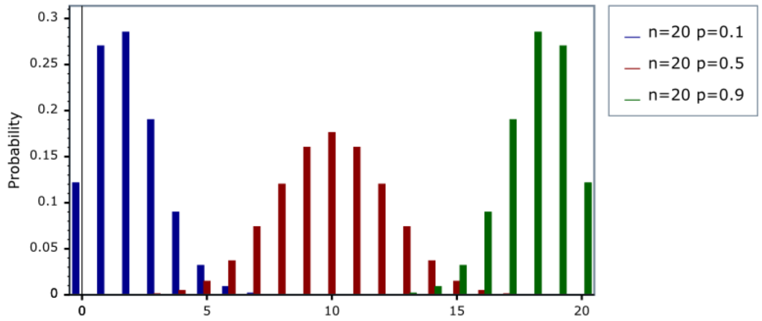
We role 2 die and get the sum of the first role plus the square of the second role

$$
\begin{aligned}
E[X+Y^2] &= \sum_{x=1}^{6}\sum_{y=1}^{6}(x+y^2)Pr(X=x,Y=y) \\
&= \sum_{x=1}^{6}\sum_{y=1}^{6}xPr(X=x,Y=y)+\sum_{x=1}^{6}\sum_{y=1}^{6}y^2Pr(X=x,Y=y) \\
&= \sum_{x=1}^{6}xPr(X=x)+\sum_{x=1}^{6}y^2Pr(Y=y) \\
&= E[X]+E[Y^2]
\end{aligned}
$$

Mean of Binomial Probability Distribution

- Suppose you flip *n* coins with bias *p*, count number of heads
- A *binomial* random variable *X* has parameters *n, p*:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \mathcal{X} = \{0, 1, 2, \ldots, n\}$$



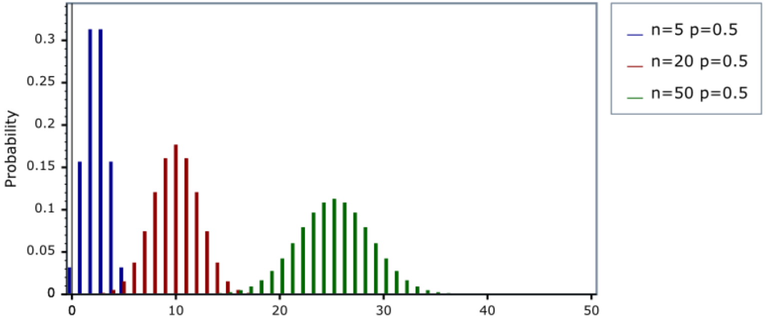
- For binomial, *expected values are expected counts of events:*

$$E[X] = pn$$

- Simple proof uses indicator variables X_i for whether each of n tosses is heads:

$$\mathbf{E}[X_i] = p \cdot 1 + (1-p) \cdot 0 = p = \Pr(X_i = 1).$$

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np.$$



Binomial Mean: The Hard Way

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} \\
&= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \\
&= \sum_{j=1}^{n} \frac{n!}{(j-1)!(n-j)!} p^j (1-p)^{n-j} \\
&= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1} (1-p)^{(n-1)-(j-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np.
\end{aligned}
$$