ENGG2430A Probability and Statistics for Engineers

Chapter 2: Discrete Random Variables

Instructor: Shengyu Zhang

Content

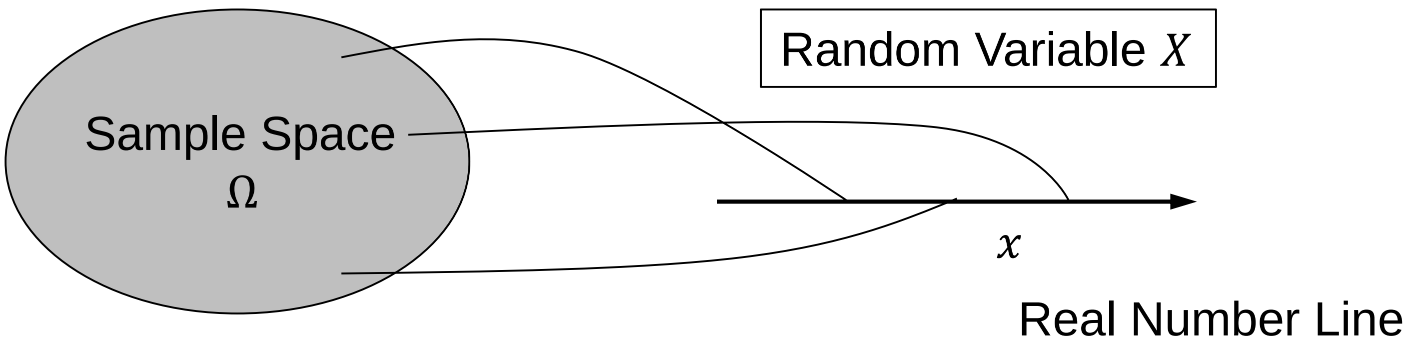
- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Basic Concepts

- In some experiments, the outcomes are numerical.
  - E.g. stock price.
- In some other experiments, the outcomes are not numerical, but they may be associated with some numerical values of interest.
- *Example*. Selection of students from a given population, we may wish to consider their grade point average.
  - The students are not numerical, but their GPA scores are.

Basic Concepts

- When dealing with these numerical values, it is useful to assign probabilities to them.
- This is done through the notion of a random variable.



Main Concepts Related to Random Variables

- Starting with a probabilistic model of an experiment:
- A random variable is a real-valued function of the outcome of the experiment.
- A function of a random variable defines another random variable.

Examples

- 5 tosses of a coin.
- This is a random variable:

The number of heads

- This is not:

Main Concepts Related to Random Variables

- We can associate with each random variable certain "averages" of interest, such as the mean and the variance.
- A random variable can be conditioned on an event or on another random variable.
- Notion of independence of a random variable from an event or from another random variable.
- We'll talk about all these in this lecture.

Discrete Random Variable

- A random variable is called discrete if its range is either finite or countably infinite.

- *Example*. Two rolls of a die.
  - The sum of the two rolls.
  - The number of sixes in the two rolls.
  - The second roll raised to the fifth power.

Continuous random variable

- *Example*. Pick a real number a and associate to it the numerical value a^2.
- The random variable a^2 is continuous, not discrete.
- We'll talk about continuous random variables later.
- The following random variable is discrete:

$$sign(a) = \begin{cases} 1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases}.$$

Discrete Random Variables: Concepts

- A discrete random variable is a real-valued function of the outcome of a discrete experiment.
- A discrete random variable has an associated probability mass function (PMF), which gives the probability of each numerical value that the random variable can take.
- A function of a discrete random variable defines another discrete random variable, whose PMF can be obtained from the PMF of the original random variable.

Content

- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Probability Mass Function

- For a discrete random variable X, the probability mass function (PMF) of X captures the probabilities of the values that it can take.

- If x is any possible value of X, the probability mass of x, denoted $p_X(x)$, is the probability of the event $\{X = x\}$ consisting of all outcomes that give rise to a value of X equal to x :

$$p_X(x) = P(\{X = x\})$$

Example

- Two independent tosses of a fair coin
- X: the number of heads obtained

- The PMF of X is

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

Probability Mass Function

- Upper case characters to denote random variables
  - $X, Y, Z, \ldots$
- Lower case characters to denote real numbers
  - $x, y, z, \ldots$
  - the numerical values of a random variable
- We'll write $P(X = x)$ in place of the notation $P(\{X = x\})$.
- Similarly, we'll write $P(X \in S)$ for the probability that X takes a value within a set S.

Probability Mass Function

- Follows from the additivity and normalization axioms

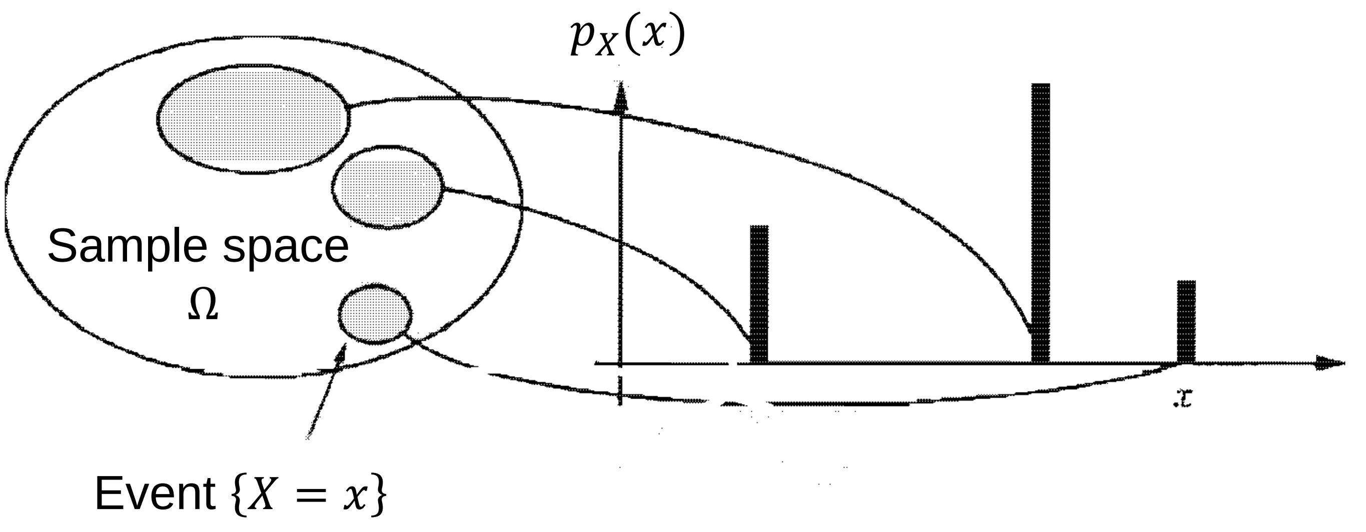
$$\sum_{\substack{x:\, all\ possible \\ values\ of\ X}} p_X(x) = 1$$

  - The events $\{X = x\}$ are disjoint, and they form a partition of the sample space
- For any set S of real numbers

$$P(X \in S) = \sum_{x \in S} p_X(x)$$

Probability Mass Function

- For each possible value x of X:
  - Collect all the possible outcomes that give rise to the event $\{X = x\}$.
  - Add their probabilities to obtain $p_X(x)$.



Important specific distributions

- Binomial random variable
- Geometric random variable
- Poisson random variable

Bernoulli Random Variable

- The Bernoulli random variable takes the two values 1 and 0

$$X \in \{0,1\}$$

- Its PMF is

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

Example of Bernoulli Random Variable

- The state of a telephone at a given time that can be either free or busy.
- A person who can be either healthy or sick with a certain disease.
- The preference of a person who can be either for or against a certain political candidate.

The Binomial Random Variable

- A biased coin is tossed n times.
- Each toss is independently of prior tosses
  - Head with probability p.
  - Tail with probability $1 - p$.
- The number X of heads up is a binomial random variable.

The Binomial Random Variable

- We refer to X as a binomial random variable with parameters n and p.

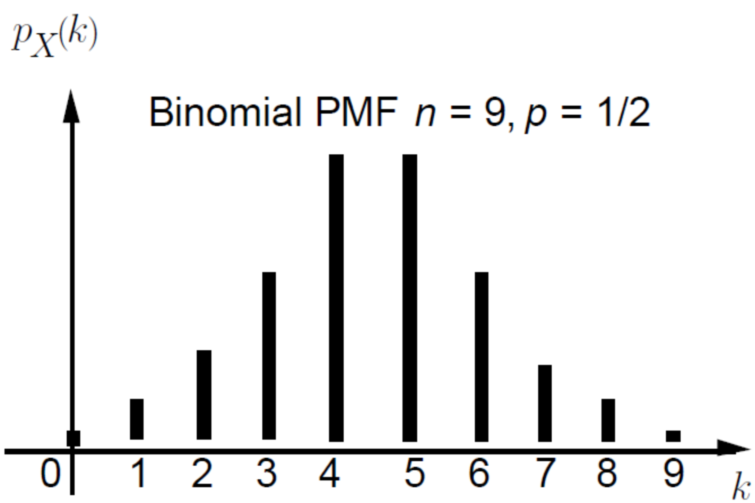
- For $k = 0,1, \dots, n$.

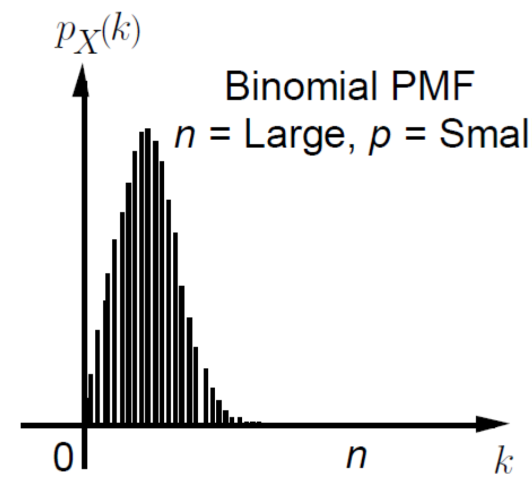
$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The Binomial Random Variable

- Normalization

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$$





The Geometric Random Variable

- Independently and repeatedly toss a biased coin with probability of a head p, where $0 < p < 1$.

- The geometric random variable is the number X of tosses needed for a head to come up for the first time.

The Geometric Random Variable

- The PMF of a geometric random variable

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

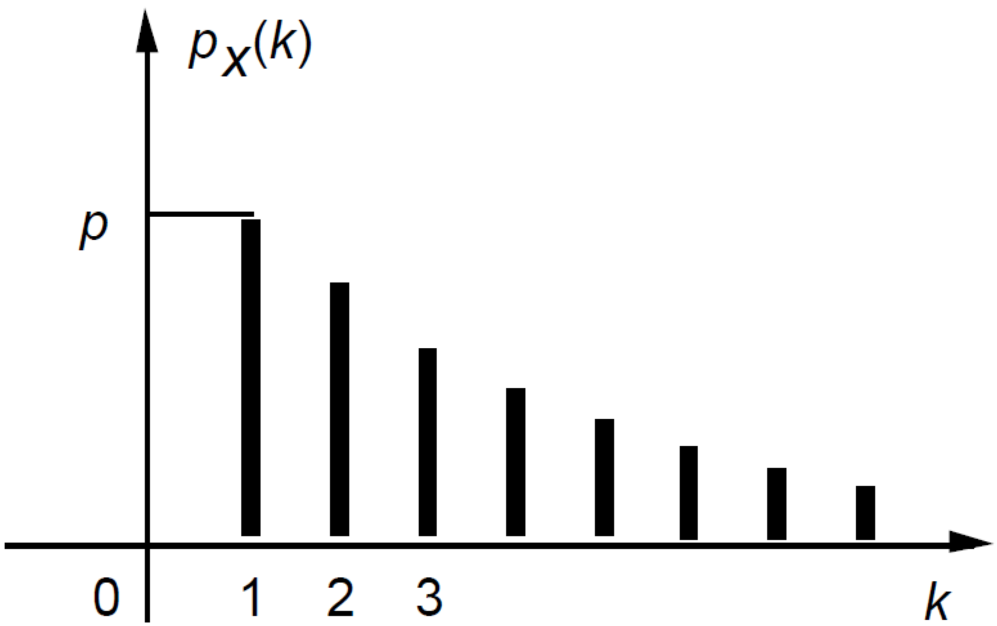
  - $k-1$ tails followed by a head.
- Normalization condition is satisfied:

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} (1-p)^{k-1}p = p\sum_{k=0}^{\infty} (1-p)^k$$

$$= p \cdot \frac{1}{1-(1-p)} = 1$$

The Geometric Random Variable

- The $p_X(k) = (1-p)^{k-1}p$ decreases as a geometric progression with parameter $1-p$.



The Poisson Random Variable

- A Poisson random variable takes nonnegative integer values.
- The PMF

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots,$$

- Normalization condition

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right)$$

$$= e^{-\lambda}e^{\lambda} = 1$$

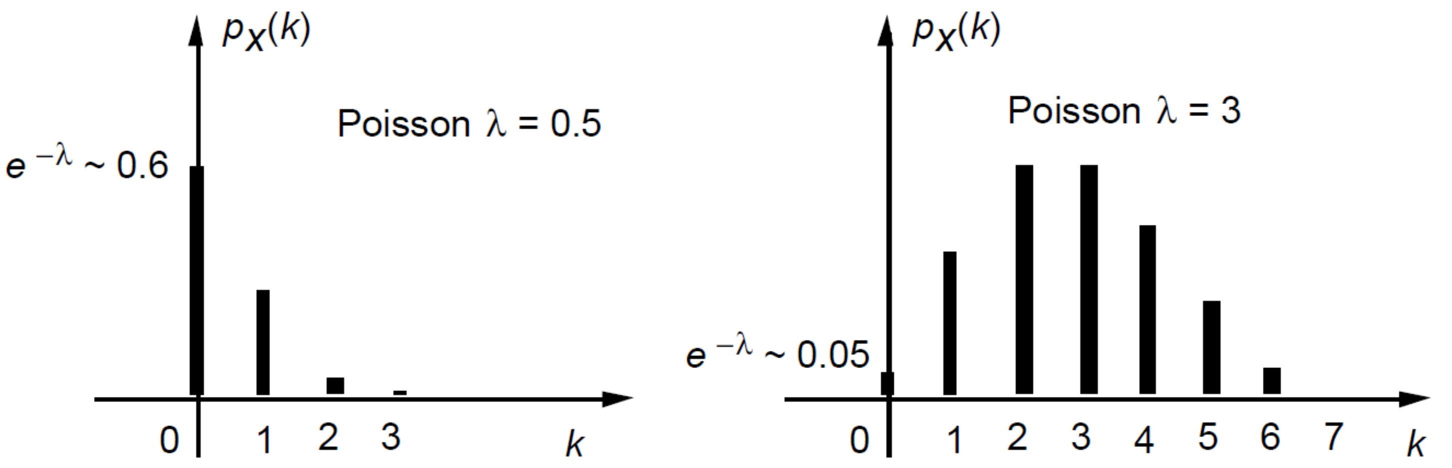
- Poisson random variable can be viewed as a binomial random variable with very small p and very large n.
- More precisely, the Poisson PMF with parameter λ is a good approximation for a binomial PMF with parameters n and p where $\lambda = np$, n is large and p is small.
  - See the wiki page for a proof.

Examples

- Because of the above connection, Poisson random variables are used in many scenarios.
- X is the number of typos in a book of n words.
  - The probability that any one word is misspelled is very small.
- X is the number of cars involved in accidents in a city on a given day.
  - The probability that any one car is involved in an accident is very small.

The Poisson Random Variable

- For Poisson random variable $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$
  - $\lambda \le 1$, monotonically decreasing
  - $\lambda > 1$, first increases and then decreases



Content

- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Functions of Random Variables

- Consider a probability model of today's weather
  - X = the temperature in degrees Celsius
  - Y = the temperature in degrees Fahrenheit
- Their relation is given by

$$Y = 1.8X + 32$$

- In this example, Y is a linear function of X, of the form

$$Y = g(X) = aX + b$$

Functions of Random Variables

- We may also consider nonlinear functions, such as

$$Y = \log(X)$$

- In general, if $Y = g(X)$ is a function of a random variable X, then Y is also a random variable.
- The PMF p_Y of $Y = g(X)$ can be calculated from PMF p_X of X

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x)$$

Example

- The PMF of X is

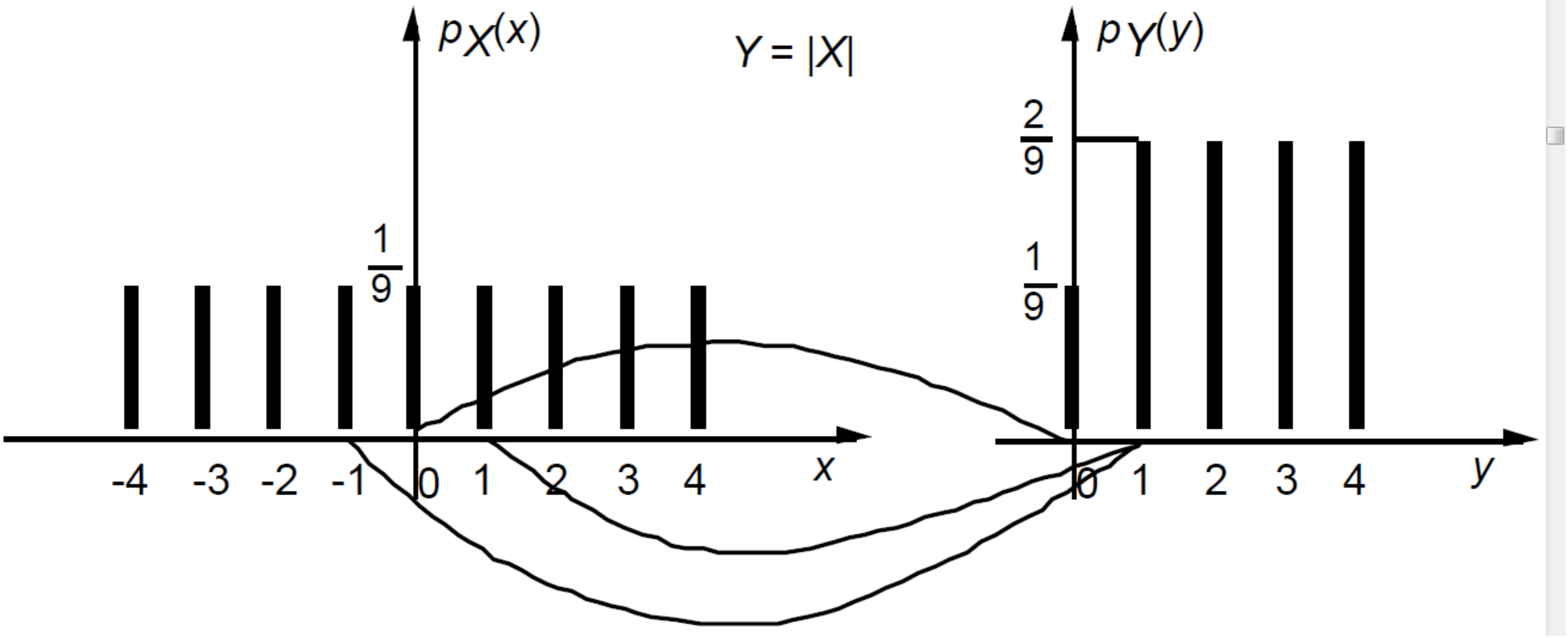
$$p_X(x) = \begin{cases} 1/9 & \text{if } x \text{ is an integer and } x \in [-4,4] \\ 0 & \text{otherwise} \end{cases}$$

- Let $Y = |X|$. Then the PMF of Y is

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

Example

- Visualization of the relation between X and Y



Example

- Let $Z = X^2$. Then the PMF of Z is

$$p_Z(z) = \begin{cases} 2/9 & \text{if } z = 1, 4, 9, 16 \\ 1/9 & \text{if } z = 0 \\ 0 & \text{otherwise} \end{cases}$$

Content

- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Expectation

- Sometimes it is desirable to summarize the values and probabilities by one number.
- The expectation of X is a weighted average of the possible values of X.
  - Weights: probabilities.
- Formally, the expected value of a random variable X, with PMF $p_X(x)$, is

$$\mathbf{E}[X] = \sum_x x p_X(x)$$

- Names: *expected value*, *expectation*, *mean*

Example

- Two independent coin tosses
  - $P(H) = \frac{3}{4}$
  - $X =$ the number of heads

- Binomial random variable with parameters $n = 2$ and $p = 3/4$.

Example

- The PMF is

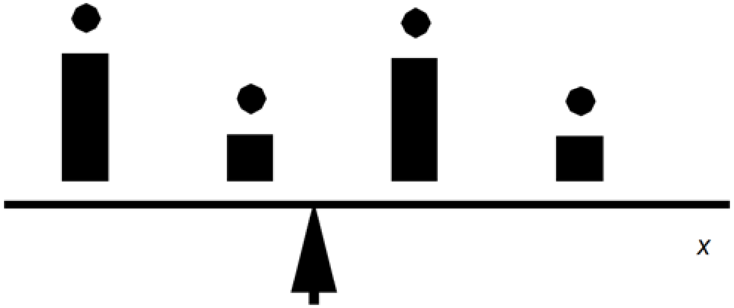
$$p_X(k) = \begin{cases} (1/4)^2 & \text{if } k = 0 \\ 2 \cdot (1/4) \cdot (3/4) & \text{if } k = 1 \\ (3/4)^2 & \text{if } k = 2 \end{cases}$$

- The mean is

$$\mathbf{E}[X] = 0 \cdot \left(\frac{1}{4}\right)^2 + 1 \cdot \left(2 \cdot \frac{1}{4} \cdot \frac{3}{4}\right) + 2 \cdot \left(\frac{3}{4}\right)^2 = \frac{3}{2}$$

Expectation

- Consider the mean as the center of gravity of the PMF



Center of gravity

$c = \text{mean} = \mathbf{E}[X]$

$$\sum_x (x - c) p_X(x) = 0$$

$$\Rightarrow c = \sum_x x p_X(x) \quad .$$

Variance

- Besides the mean, there are several other important quantities.
- The kth moment is $\mathbf{E}\left[X^k\right]$
  - So the first moment is just the mean.
- Variance of X, denoted by $\mathrm{var}(X)$, is
$$\mathrm{var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2]$$
  - The second moment of $X - \mathbf{E}[X]$.
- The variance is always non-negative:
$$var(X) \geq 0$$

Standard deviation

- Variance is closely related to another measure.

- Standard deviation of X, denoted by σ_X, is

$$\sigma_X = \sqrt{\mathrm{var}(X)}$$

Example

- Suppose that the PMF of X is

$$p_X(x) = \begin{cases} 1/9 & \text{if } x \text{ is an integer and } x \in [-4,4] \\ 0 & \text{otherwise} \end{cases}$$

- The expectation

$$\mathbf{E}[X] = \sum_x x p_X(x) = \frac{1}{9} \sum_{x=-4}^{4} x = 0$$

- Can also be seen from symmetry.

Example

- Let $Z = (X - \mathbf{E}[X])^2 = X^2$. The PMF of Z

$$p_Z(z) = \begin{cases} 2/9 & \text{if } z = 1, 4, 9, 16 \\ 1/9 & \text{if } z = 0 \\ 0 & \text{otherwise} \end{cases}$$

- The variance of X is then

$$\mathrm{var}(X) = \mathbf{E}[Z] = \sum_z z p_Z(z)$$

$$= 0 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 4 \cdot \frac{2}{9} + 9 \cdot \frac{2}{9} + 16 \cdot \frac{2}{9} = \frac{60}{9}$$

Expectation for $g(X)$

- There is a simpler way of computing $var\big(g(X)\big)$.
- Let X be a random variable with PMF $p_X(x)$, and let $g(X)$ be a real-valued function of X.
- The expected value of the random variable $Y = g(X)$ is

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$$

Expectation for $g(X)$

- Using the formula $p_Y(y) = \sum_{\{x|g(x)=y\}} p_X(x)$:

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

Variance example

- The PMF of X

$$p_X(x) = \begin{cases} 1/9 & \text{if } x \text{ is an integer and } x \in [-4,4] \\ 0 & \text{otherwise} \end{cases}$$

- The variance

$$\begin{aligned} \mathrm{var}(X) &= \mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= \sum_x (x - \mathbf{E}[X])^2 p_X(x) \\ &= \frac{1}{9} \sum_{x=-4}^{4} x^2 \\ &= (16 + 9 + 4 + 1 + 0 + 1 + 9 + 16)/9 \\ &= \frac{60}{9} \end{aligned}$$

Mean of $aX + b$

- Let Y be a linear function of X

$$Y = aX + b$$

- The mean of Y

$$\mathbf{E}[Y] = \sum_x (ax + b) p_X(x)$$

$$= a \sum_x x p_X(x) + b \sum_x p_X(x) = a\mathbf{E}[X] + b$$

- The expectation scales *linearly*.

Variance of $aX + b$

- Let Y be a linear function of X

$$Y = aX + b$$

- The variance of Y

$$\begin{aligned}\mathrm{var}(Y) &= \sum_x (ax + b - \mathbf{E}[aX + b])^2 p_X(x) \\ &= \sum_x (ax + b - a\mathbf{E}[X] - b)^2 p_X(x) \\ &= a^2 \sum_x (x - \mathbf{E}[X])^2 p_X(x) \\ &= a^2 \mathrm{var}(X)\end{aligned}$$

- The variance scales *quadratically*.

Variance as moments

- Fact. $var(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$.

- $$\begin{aligned} var(X) &= \mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + (\mathbf{E}[X])^2] \\ &= \mathbf{E}[X^2] - 2\mathbf{E}\big[X\mathbf{E}[X]\big] + (\mathbf{E}[X])^2 \\ &= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + (\mathbf{E}[X])^2 \\ &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \end{aligned}$$

Example: Average time

- Distance between class and home is 2 miles
- $P(\text{weather is good}) = 0.6$
- Speed:
  - $V = 5$ miles/hour if weather is good.
  - $V = 30$ miles/hour if weather is bad.

- *Question*: What is the mean of the time T to get to class?

Example: Average time

- The PMF of T

$$p_T(t) = \begin{cases} 0.6 & \text{if } t = \dfrac{2}{5}\ hours \\ 0.4 & \text{if } t = \dfrac{2}{30}\ hours \end{cases}$$

- The mean of T

$$\mathbf{E}[T] = 0.6 \cdot \frac{2}{5} + 0.4 \cdot \frac{2}{30} = \frac{4}{15}$$

Example: Average time

- Wrong calculation by speed V
- The mean of speed V

$$\mathbf{E}[V] = 0.6 \cdot 5 + 0.4 \cdot 30 = 15$$

- The mean of time T

$$\frac{2}{\mathbf{E}[V]} = \frac{2}{15}$$

- To summarize, in this example we have

$$T = \frac{2}{V} \quad \text{and} \quad \mathbf{E}[T] = \mathbf{E}\left[\frac{2}{V}\right] \neq \frac{2}{\mathbf{E}[V]}$$

Example: Bernoulli

- Consider the Bernoulli random variable X with PMF

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

- Its mean, second moment, and variance:

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

$$\mathbf{E}[X^2] = 1^2 \cdot p + 0 \cdot (1 - p) = p$$

$$\mathrm{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = p - p^2 = p(1 - p)$$

Example: Uniform

- What is the mean and variance of the roll of a fair six-sided die?

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

- The mean $\mathbf{E}[X] = 3.5$ and the variance

$$\begin{aligned} \mathrm{var}(X) &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \\ &= \frac{1}{6}\left(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2\right) - 3.5^2 \\ &= 35/12 \end{aligned}$$

Example: Uniform integers

- General, a discrete uniformly distributed random variable
  - Range: contiguous integer values $a, a+1, \ldots, b$
  - Probability: equal probability
- The PMF is

$$p_X(k) = \begin{cases} \dfrac{1}{b-a+1} & \text{if } k = a, a+1, \ldots, b \\ 0 & \text{otherwise} \end{cases}$$

Example: Uniform integers

- The mean

$$\mathbf{E}[X] = \frac{a+b}{2}$$

- For variance, first consider $a = 1$ and $b = n$
- The second moment

$$\mathbf{E}[X^2] = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{6}(n+1)(2n+1)$$

Example: Uniform integers

- The variance for special case

$$\begin{aligned}\mathrm{var}(X) &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \\ &= \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 \\ &= \frac{n^2-1}{12}\end{aligned}$$

Example: Uniform integers

- For the case of general integers a and b
  - X: discrete uniform over $[a, b]$
  - Y: discrete uniform over $[1, b - a + 1]$
- Relation between X and Y

$$Y = X - a + 1$$

- Thus

$$\mathrm{var}(X) = \mathrm{var}(Y) = \frac{(b - a + 1)^2 - 1}{12}$$

Example: Poisson

- Recall Poisson PMF

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad\qquad k = 0,1,2,\ldots,$$

- Mean:

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}$$

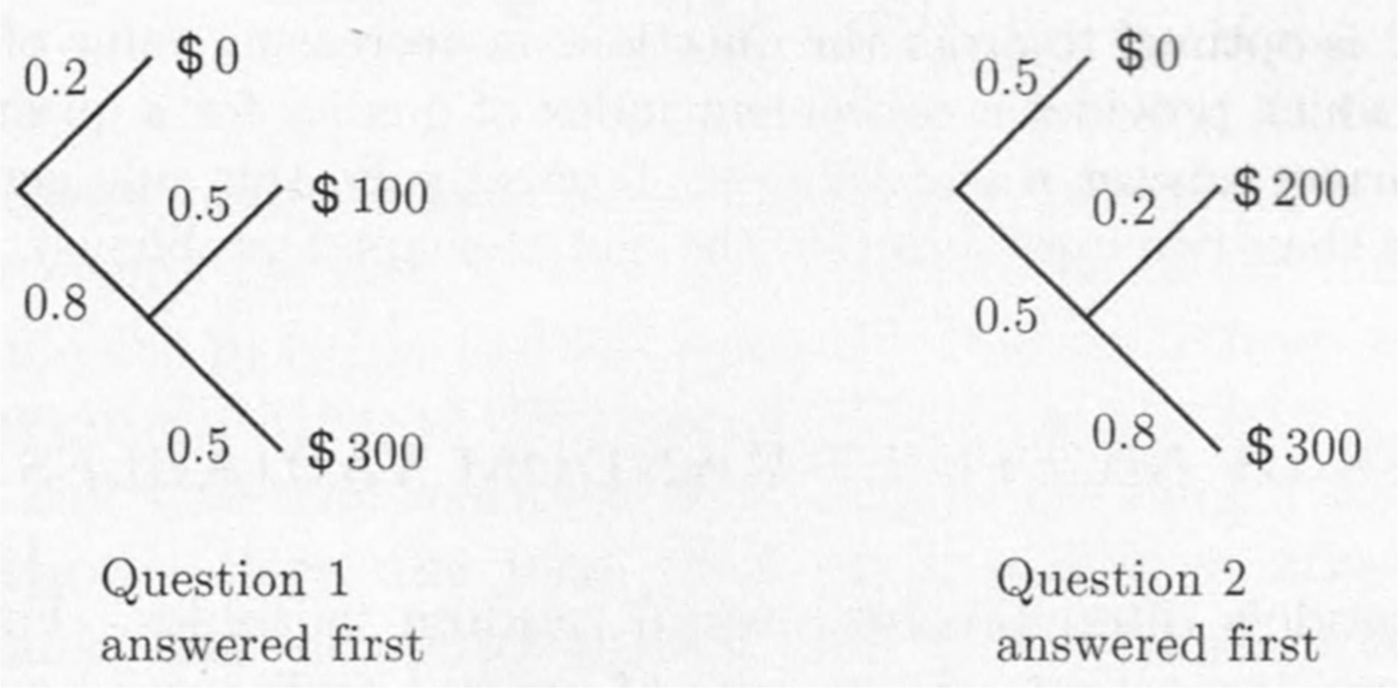
$$= \lambda$$

- Variance: $var[X] = \lambda$.
  - Verification left as exercise.

The Quiz Problem

- A person is given two questions and must decide which question to answer first.
  - $P(\text{question 1 correct}) = 0.8$ Prize=$100
  - $P(\text{question 2 correct}) = 0.5$ Prize=$200
  - If incorrectly answer the first question, then no second question.
- How to choose the first question so that maximize the expected prize?

Tree illustration



The Quiz Problem

- Answer question 1 first: Then the PMF of X is

$$p_X(0) = 0.2$$
$$p_X(100) = 0.8 \cdot 0.5$$
$$p_X(300) = 0.8 \cdot 0.5$$

- We have

$$\mathbf{E}[X] = 0.8 \cdot 0.5 \cdot 100 + 0.8 \cdot 0.5 \cdot 300 = 160$$

The Quiz Problem

- Answer question 2 first: Then the PMF of X is

$$p_X(0) = 0.5$$
$$p_X(200) = 0.5 \cdot 0.2$$
$$p_X(300) = 0.5 \cdot 0.8$$

- We have

$$\mathbf{E}[X] = 0.5 \cdot 0.2 \cdot 200 + 0.5 \cdot 0.8 \cdot 300 = 140$$

- It is better to answer question 1 first.

The Quiz Problem

- Let us now generalize the analysis.
  - p_1: $P(\text{correctly answering question 1})$
  - p_2: $P(\text{correctly answering question 2})$
  - v_1: prize for question 1
  - v_2: prize for question 2

The Quiz Problem

- Answer question 1 first

$$\mathbf{E}[X] = p_1(1 - p_2)v_1 + p_1p_2(v_1 + v_2)$$
$$= p_1v_1 + p_1p_2v_2$$

- Answer question 2 first

$$\mathbf{E}[X] = p_2(1 - p_1)v_2 + p_2p_1(v_2 + v_1)$$
$$= p_2v_2 + p_2p_1v_1$$

The Quiz Problem

- It is optimal to answer question 1 first if and only if

$$p_1 v_1 + p_1 p_2 v_2 \geq p_2 v_2 + p_2 p_1 v_1$$

- Or equivalently

$$\frac{p_1 v_1}{1 - p_1} \geq \frac{p_2 v_2}{1 - p_2}$$

- Rule: Order the questions in decreasing value of the expression $pv/(1-p)$

Content

- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Multiple Random Variables

- Probabilistic models often involve several random variables of interest.

- Example: In a medical diagnosis context, the results of several tests may be significant.
- Example: In a networking context, the workloads of several gateways may be of interest.

Joint PMFs of Multiple Random Variables

- Consider two discrete random variables X and Y associated with the same experiment.
- The joint PMF of X and Y is denoted by $p_{X,Y}$. It specifies the probability of the values that X and Y can take.
- If (x, y) is a pair of values that (X, Y) can take, then the probability mass of (x, y) is the probability of the event $\{X = x, Y = y\}$:

$$P_{X,Y}(x, y) = P(X = x, Y = y).$$

- The joint PMF determines the probability of any event that can be specified in terms of the random variables X and Y.

- For example, if A is the set of all pairs (x, y) that have a certain property, then
$$P\big((X, Y) \in A\big) = \sum_{(x,y)\in A} p_{X,Y}(x, y)$$

Joint PMFs of Multiple Random Variables

- The PMFs of X and Y

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

- The formula can be verified by

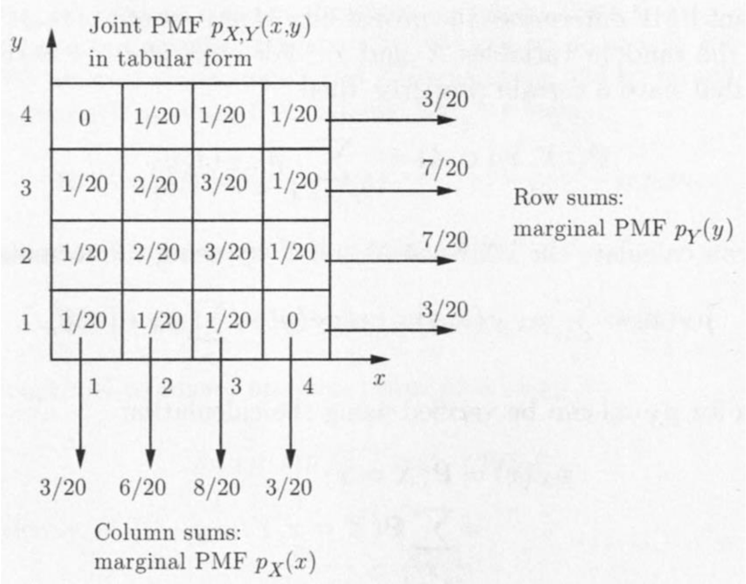
$$\begin{aligned} p_X(x) &= P(X = x) \\ &= \sum_y P(X = x, Y = y) \\ &= \sum_y p_{X,Y}(x,y) \end{aligned}$$

- p_X, p_Y are the marginal PMFs.

Joint PMFs of Multiple Random Variables

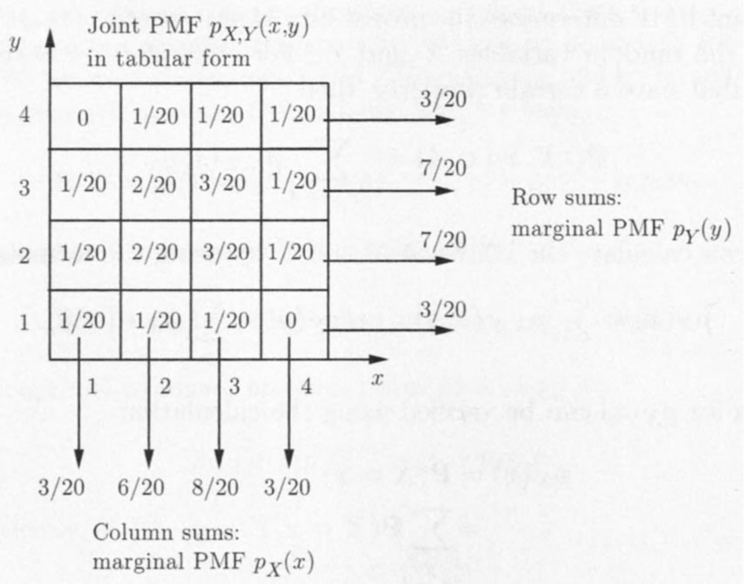
- Computing the marginal MPFs p_X and p_Y of $p_{X,Y}$ from table.

- The joint PMF $p_{X,Y}$ is arranged in a two-dimensional table.



Joint PMFs of Multiple Random Variables

- 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

- One can generate new random variables by applying functions on several random variables.
- Consider $Z = g(X, Y)$.
- Its PMF can be calculated from the joint PMF $p_{X,Y}$ according to

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

Functions of Multiple Random Variables

- The expected value rule for multiple variables

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

- For special case, g is linear and of the form $aX + bY + c$, we have

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

- "linearity of expectation" --- regardless of dependence of X and Y.

More than Two Random Variables

- We can also consider three or more random variables.
- The joint PMF of three random variables X, Y, and Z

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

- The marginal PMFs are

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

and

$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x, y, z)$$

More than Two Random Variables

- The expected value rule for functions

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

- If g is linear and of the form

$$g(X,Y,Z) = aX + bY + cZ + d$$

then

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

More than Two Random Variables

- Generalization to more than three random variables.

- For any random variables $X_1, X_2, \ldots, X_n$ and any scalars $a_1, a_2, \ldots, a_n$, we have
$$\mathbf{E}[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n]$$
$$= a_1 \mathbf{E}[X_1] + a_2 \mathbf{E}[X_2] + \cdots + a_n \mathbf{E}[X_n]$$

Example: Mean of the Binomial

- 300 students in probability class
- Each student has probability 1/3 of getting an A, independently of any other student.
- X: the number of students that get an A.

- Question: What is the mean of X?

Example: Mean of the Binomial

- Let X_i be the random variable for ith student

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

- Each X_i is a Bernoulli random variable
  - $\mathbf{E}[X_i] = p = 1/3$
  - $\mathbf{Var}[X_i] = p(1-p) = (1/3)(2/3) = 2/9$

Example: Mean of the Binomial

- The random variable X can be expressed as their sum

$$X = X_1 + X_2 + \cdots + X_n$$

- Using the linearity of X as a function of the X_i

$$\mathbf{E}[X] = \sum_{i=1}^{300} \mathbf{E}[X_i] = \sum_{i=1}^{300} \frac{1}{3} = 300 \cdot \frac{1}{3} = 100$$

Example: Mean of the Binomial

- If we repeat this calculation for a general number of students n and probability of A equal to p, we obtain

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

Example: The Hat Problem

- Suppose that n people throw their hats in a box.
- Each picks up one hat at random.
- X: the number of people that get back their own hat

- Question: What is the expected value of X?

Example: The Hat Problem

- For the ith person, we introduce a random variable X_i

$$X_i = \begin{cases} 1 & \text{if the } i\text{th his own} \\ 0 & \text{otherwise} \end{cases}$$

- Since $P(X_i = 1) = \frac{1}{n}$ and $P(X_i = 0) = 1 - \frac{1}{n}$

$$E[X_i] = 1 \cdot \frac{1}{n} + 0 \cdot \left(1 - \frac{1}{n}\right) = \frac{1}{n}$$

Example: The Hat Problem

- We know

$$X = X_1 + X_2 + \cdots + X_n$$

- Thus

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n] = n \cdot \frac{1}{n} = 1$$

Summary of Facts About Joint PMFs

- The joint PMF of X and Y is defined by

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

- The marginal PMFs of X and Y can be obtained from the joint PMF, using the formulas

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

Summary of Facts About Joint PMFs

- A function $g(X,Y)$ of X and Y defines another random variable

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

If g is linear, of the form $aX + bY + c$,

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

- These naturally extend to more than two random variables.

Content

- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Conditioning

- In a probabilistic model, a certain event A has occurred
- Conditional probability captures this knowledge.
- Conditional probabilities are like ordinary probabilities (satisfy the three axioms) except
  - refer to a new universe: event A is known to have occurred

Conditioning a Random Variable on an Event

- The conditional PMF of a random variable X, conditioned on a particular event A with $P(A) > 0$, is defined by

$$\begin{aligned} p_{X|A}(x) &= P(X = x|A) \\ &= \frac{P(\{X = x\} \cap A)}{P(A)} \end{aligned}$$

Conditioning a Random Variable on an Event

- Consider the events $\{X = x\} \cap A$:
  - They are disjoint for different values of x.
  - Their union is A.
- Thus $P(A) = \sum_x P(\{X = x\} \cap A)$
- Combining this and $p_{X|A}(x) = P(\{X = x\} \cap A)/P(A)$ (last slide), we can see that

$$\sum_x p_{X|A}(x) = 1$$

- So $p_{X|A}$ is a legitimate PMF.

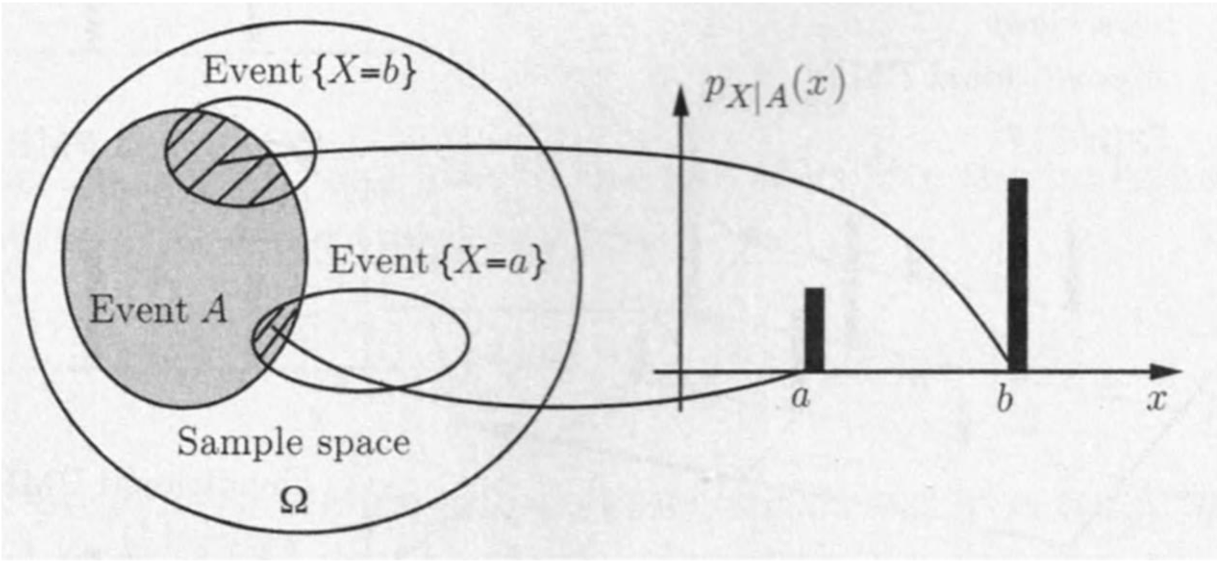
Conditioning a Random Variable on an Event

- The conditional PMF is calculated similar to its unconditional counterpart.

- To obtain $p_{X|A}(x)$
  - Add the probabilities of the outcomes $X = x$
  - Conditioning event A
  - Normalize by dividing with $P(A)$

Conditioning a Random Variable on an Event

- Visualization and calculation of the conditional PMF $p_{X|A}(x)$



Example: dice

- X: the roll of a fair 6-sided dice
- A: the roll is an even number

$$p_{X|A}(x) = P(X = x|A)$$
$$= \frac{P(X = x \; and \; A)}{P(A)}$$
$$= \begin{cases} \frac{1}{3} & \text{if } x = 2,4,6 \\ 0 & \text{otherwise} \end{cases}$$

Conditioning one random variable on another

- We have talked about conditioning a random variable X on an event A.
- Now let's consider conditioning a random variable X on another random variable Y.
- Let X and Y be two random variables associated with the same experiment.
- The experimental value $Y = y$ ($p_Y(y) > 0$) provides partial knowledge about the value of X.

Conditioning one random variable on another

- The knowledge is captured by the conditional PMF $p_{X|Y}$ of X given Y, which is defined as $p_{X|A}$ for $A = \{Y = y\}$:

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

- Using the definition of conditional probabilities

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

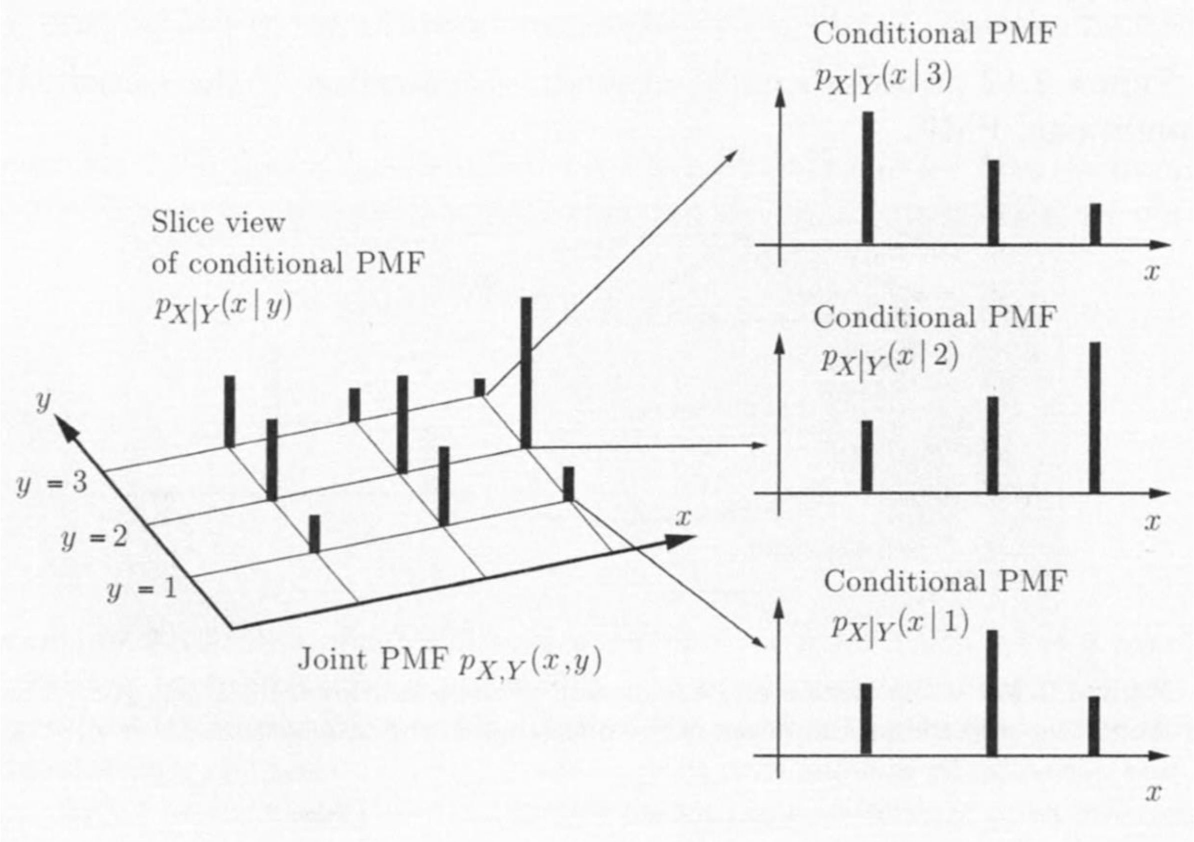
Conditioning one random variable on another

- Fix some y, with $p_Y(y) > 0$ and consider $p_{X|Y}(x|y)$ as a function of x.

- This function is a valid PMF for X:
  - Assigns nonnegative values to each possible x
  - These values add to 1
  - Has the same shape as $p_{X,Y}(x, y)$
  - $\sum_x p_{X|Y}(x|y) = 1$

Conditioning one random variable on another

- Visualization of the conditional PMF $p_{X|Y}(x|y)$



Conditioning one random variable on another

- It is convenient to calculate the joint PMF by a sequential approach and the formula

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

- Or its counterpart

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

- This method is entirely similar to the use of the multiplication rule from previous lectures.

Example: Question answering

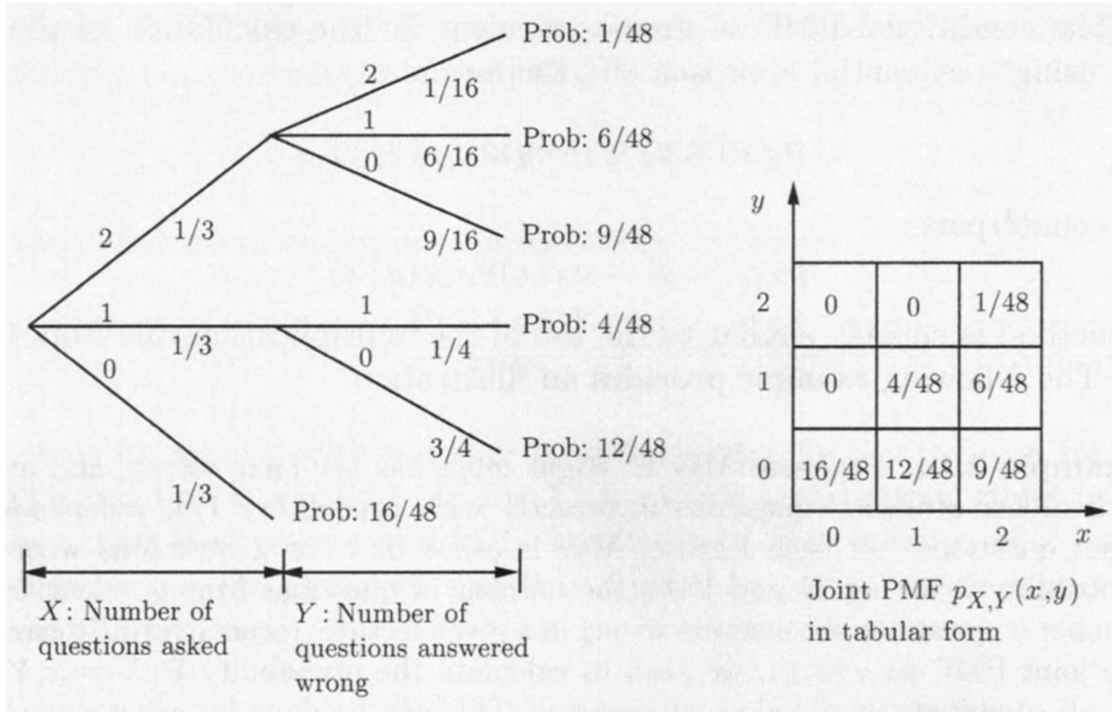
- A professor independently answers each of her students' questions incorrectly with probability ¼.
- In each lecture the professor is asked $0,1,$ or 2 questions with equal probability $1/3$.
  - X: the number of questions professor is asked
  - Y: the number of questions she answers wrong in a given lecture

Example: Question answering

- Construct the joint PMF $p_{X,Y}(x, y)$: calcualte all the probabilities $P(X = x, Y = y)$.

- Using a sequential description of the experiment and the multiplication rule
$$p_{X,Y}(x, y) = p_Y(y)p_{X|Y}(x|y)$$

Example: Question answering



- For example,

$$p_{X,Y}(1,1) = p_X(x)p_{Y|X}(y,x) = \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$$

Example: Question answering

- We can compute other useful information from two-dimensional table.
- For example,

$$P(\text{at least one wrong answer})$$
$$= p_{X,Y}(1,1) + p_{X,Y}(2,1) + p_{X,Y}(2,2)$$
$$= \frac{4}{48} + \frac{6}{48} + \frac{1}{48} = \frac{11}{48}$$

Conditioning one random variable on another

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

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

- This formula provides a divide-and-conquer method for calculating marginal PMFs.

Summary of Facts About Conditional PMFs

- Conditional PMFs are similar to ordinary PMFs, but refer to a universe where the conditioning event is known to have occurred.
- The conditional PMF of X given an event A with $P(A) > 0$, is defined by

$$p_{X|A}(x) = P(X = x|A)$$

and satisfies

$$\sum_x p_{X|A}(x) = 1$$

Summary of Facts About Conditional PMFs

- The conditional PMF of X given Y can be used to calculate the marginal PMFs with the formula

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

This is analogous to the divide-and-conquer approach for calculating probabilities using the total probability theorem.

Conditional Expectations

- The conditional expectation of X given an event A with $P(A) > 0$, is defined by

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

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

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

Conditional Expectations

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

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

- The total expectation theorem

$$\boldsymbol{E}[X] = \sum_y p_Y(y)\,\mathbf{E}[X|Y=y]$$

Conditional Expectations

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

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

- Indeed,

$$\begin{aligned}
\mathbf{E}[X] &= \sum_x x p_X(x) \\
&= \sum_x x \sum_{i=1}^{n} P(A_i) p_{x|A_i}(x|A_i) \\
&= \sum_{i=1}^{n} P(A_i) \sum_x x p_{x|A_i}(x|A_i) \\
&= \sum_{i=1}^{n} P(A_i)\, \mathbf{E}[X|A_i]
\end{aligned}$$

Conditional Expectation

- Messages transmitted by a computer in Boston through a data network are destined
  - for New York with probability 0.5
  - for Chicago with probability 0.3
  - for San Francisco with probability 0.2
- The transit time X of a message is random
  - $\mathbf{E}[X] = 0.05$ for New York
  - $\mathbf{E}[X] = 0.1$ for Chicago
  - $\mathbf{E}[X] = 0.3$ for San Francisco

Conditional Expectation

- By total expectation theorem

$$\mathbf{E}[X] = 0.5 \cdot 0.05 + 0.3 \cdot 0.1 + 0.2 \cdot 0.3$$
$$= 0.115$$

Mean and Variance of the Geometric Random Variable

- You write a software program over and over,
  - probability p that it works correctly
  - independently from previous attempts
- X: the number of tries until the program works correctly

- Question: What is the mean and variance of X?

Mean and Variance of the Geometric Random Variable

- X is a geometric random variable with PMF

$$p_X(k) = (1-p)^{k-1}p \qquad k = 1,2,\ldots$$

- The mean and variance of X

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

$$\mathrm{var}(X) = \sum_{k=1}^{\infty}(k-\mathbf{E}[X])^2(1-p)^{k-1}p$$

Mean and Variance of the Geometric Random Variable

- Evaluating these infinite sums is somewhat tedious.
- As an alternative, we will apply the total expectation theorem.
- Let

$$A_1 = \{X = 1\} = \{\text{first try is a success}\}$$

and

Mean and Variance of the Geometric Random Variable

- If the first try is successful, we have $X = 1$

$$\mathbf{E}[X|X = 1] = 1$$

- If the first try fails ($X > 1$), we have wasted one try, and we are back where we started.
  - The expected number of remaining tries is $\mathbf{E}[X]$
- We have

$$\mathbf{E}[X|X > 1] = 1 + \mathbf{E}[X]$$

Mean and Variance of the Geometric Random Variable

- Thus

$$\mathbf{E}[X] = P(X = 1)\mathbf{E}[X|X = 1] + P(X > 1)\mathbf{E}[X|X > 1] = p + (1 - p)(1 + \mathbf{E}[X])$$

- Solving this equation gives

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

Mean and Variance of the Geometric Random Variable

- Similar reasoning

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

and

$$\mathbf{E}\left[X^2 \middle| X > 1\right] = \mathbf{E}\left[(1 + X)^2\right]$$
$$= 1 + 2\mathbf{E}[X] + \mathbf{E}[X^2]$$

- So

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

Mean and Variance of the Geometric Random Variable

- We obtain

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

- and conclude that

$$\begin{aligned}\mathbf{Var}(X) &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2 \\ &= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}\end{aligned}$$

Content

- Basic Concepts
- Probability Mass Function
- Functions of Random Variables
- Expectation, Mean, and Variance
- Joint PMFs of Multiple Random Variables
- Conditioning
- Independence

Independence of a r.v. from an event

- Idea is similar to the independence of two events.

- Knowing the occurrence of the conditioning event tells us nothing about the value of the random variable.

Independence of a r.v. from an event

- Formally, the random variable X is independent of the event A if
$$P(X = x \text{ and } A) = P(X = x)P(A) = p_X(x)P(A)$$

- Same as requiring that the events $\{X = x\}$ and A are independent, for any choice x.

Independence of a r.v. from an event

- Consider $P(A) > 0$
- By the definition of the conditional PMF
$$p_{X|A}(x) = P(X = x \text{ and } A)/P(A)$$
- Independence is the same as the condition
$$p_{X|A}(x) = p_X(x) \quad \text{for all } x$$

Independence of a r.v. from an event

- Consider two independent tosses of a fair coin.
  - X: the number of heads
  - A: the number of heads is even

- The PMF of X

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

Independence of a r.v. from an event

- We know $P(A) = \frac{1}{2}$
- The conditional PMF

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

- The PMFs p_X and $p_{X|A}$ are different
  $\Rightarrow$ X and A are not independent

Independence of random variables

- The notion of independence of two random variables is similar.
- Two random variables X and Y are independent if
$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y$$
- Same as requiring that the two events $\{X = x\}$ and $\{Y = y\}$ be independent for every x and y.

Independence of random variables

- By the formula

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

- Independence is equivalent to the condition

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

for all y with $p_Y(y) > 0$ and all x.

- Independence means that the experimental value of Y tells us nothing about the value of X.

Independence of random variables

- X and Y are conditionally independent, if given a positive probability event A
$P(X = x, Y = y|A) = P(X = x|A)P(Y = y|A)$
- Using this chapter's notation

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

- Or equivalently,

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

for all x, y such that $p_{Y|A}(y) > 0$.

Independence of random variables

- If X and Y are independent random variables, then

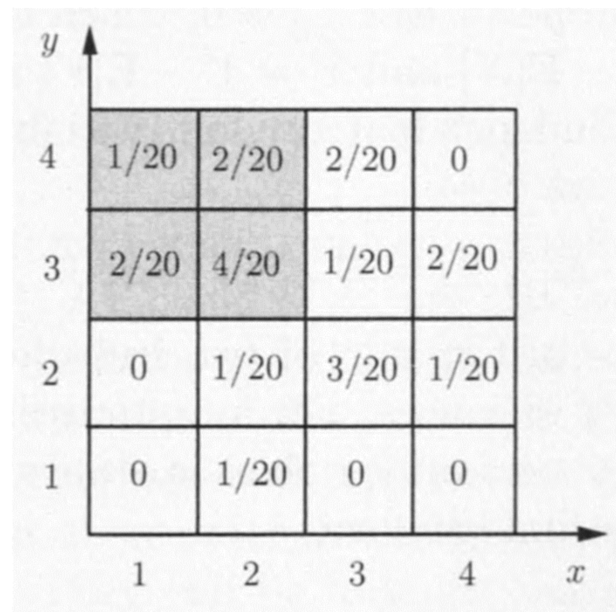
$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$$

- Shown by the following calculation

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

Independence of random variables

- Conditional independence may not imply unconditional independence.
- X and Y are not independent
  - $p_{X|Y}(1|1) = P(X = 1|Y = 1)$
    $= 0 \neq P(X = 1) = p_X(1)$
- Condition on
  $A = \{X \leq 2, Y \geq 3\}$
  - They are independent



Independence of random variables

- A very similar calculation shows that if X and Y are independent, then so are $g(X)$ and $h(Y)$ for any functions g and h.
- $\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)]$
- Next, we consider variance of sum of independent random variables.

Independence of random variables

- Consider $Z = X + Y$, where X and Y are independent.
- $\mathbf{Var}(Z) = \mathbf{E}[(X + Y - \mathbf{E}[X + Y])^2]$

$$= \mathbf{E}[(X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2]$$

$$= \mathbf{E}\left[\big((X - \mathbf{E}[X]) + (Y - \mathbf{E}[Y])\big)^2\right]$$

$$= \mathbf{E}[(X - \mathbf{E}[X])^2] + \mathbf{E}[(Y - \mathbf{E}[Y])^2]$$

$$+2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

Independence of random variables

- Now we compute $\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$.
- Since X and Y are independent, so are $X - \mathbf{E}[X]$ and $Y - \mathbf{E}[Y]$.
  - As they are two functions of X and Y, respectively.
- Thus $\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$

$$= \mathbf{E}[(X - \mathbf{E}[X])] \cdot \mathbf{E}[(Y - \mathbf{E}[Y])]$$

$$= 0 \cdot 0 = 0$$

- So $\mathbf{Var}(Z) = \mathbf{E}[(X - \mathbf{E}[X])^2] + \mathbf{E}[(Y - \mathbf{E}[Y])^2]$

$$= \mathbf{Var}[X] + \mathbf{Var}[Y]$$

Summary of independent r.v.'s

- X is independent of the event A if

$$p_{X|A}(x) = p_X(x)$$

that is, if for all x, the events $\{X = x\}$ and A are independent.

- X and Y are independent if for all possible pairs (x, y), the events $\{X = x\}$ and $\{Y = y\}$ are independent

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

Summary of Facts About Independent Random Variables

- If X and Y are independent random variables, then
  1. $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$
  2. $\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)]$, for any functions g and h.
  3. $\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$

Independence of Several Random Variables

- All previous results have natural extensions to more than two random variables.
- Example: Random variables X, Y, and Z are independent if

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

- Example: If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\begin{aligned}&\mathbf{Var}(X_1 + X_2 + \cdots + X_n)\\&= \mathbf{Var}(X_1) + \mathbf{Var}(X_2) + \cdots + \mathbf{Var}(X_n)\end{aligned}$$

Variance of the Binomial

- Consider n independent coin tosses
  - $P(H) = p$
  - X_i: Bernoulli random variable for ith toss

- Its PMF

$$p_{X_i}(x) = \begin{cases} 1 & i\text{th toss comes up a head} \\ 0 & \text{otherwise} \end{cases}$$

Variance of the Binomial

- Let $X = X_1 + X_2 + \cdots + X_n$ be a binomial random variable.

- By the independence of the coin tosses

$$\mathbf{Var}(X) = \sum_{i=1}^{n} \mathbf{Var}(X_i) = np(1-p)$$

Mean and Variance of the Sample Mean

- Estimate the approval rating of a president C.
  - Ask n persons randomly from the voters
  - X_i response of the ith person

- $$X_i = \begin{cases} 1 & i\text{th person approves } C \\ 0 & i\text{th person disapproves } C \end{cases}$$

Mean and Variance of the Sample Mean

- Model $X_1, X_2, \dots, X_n$ as independent Bernoulli random variables
  - mean p
  - variance $p(1-p)$

- The sample mean

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

Mean and Variance of the Sample Mean

- S_n is the approval rating of C within our n-person sample.
- Using the linearity of S_n as a function of the X_i

$$\mathbf{E}[S_n] = \sum_{i=1}^{n} \frac{1}{n} \mathbf{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} p = p$$

and

$$\mathbf{Var}(S_n) = \sum_{i=1}^{n} \frac{1}{n^2} \mathbf{Var}(X_i) = \frac{p(1-p)}{n}$$