# COMPSCI 240: Reasoning Under Uncertainty

Nic Herndon and Andrew Lan

University of Massachusetts at Amherst

Spring 2019, Section 01

# Lecture 31: Review for Final Exam

# Topics

- Basic counting problems
- Probability
- Discrete random variables
- Midterm Exam #1
- Continuous random variables
- Central limit theorem
- Probabilistic reasoning
- Game theory
- Midterm Exam #2
- Markov chains
- Bayesian network
- Final Exam

# Overview

- Basic counting problems
  - Set theory: size of, subset, disjoint sets, partitions, power set, universal set, operations (complement, union, intersection)
  - Counting: permutations, $k$-permutations, combinations, partitions
- Probability
  - Probability axioms
  - Conditional probability (sequential model)
  - Multiplication rule
  - Total probability theorem
  - Bayes' rule
  - Independence
  - Conditional independence
- Discrete random variables
  - Probability mass function (PMF)
  - Common discrete RVs: uniform, Bernoulli, binomial, geometric, Poisson
  - Expectation and Variance + their properties (e.g., functions of RVs)
  - Multiple RVs (joint, marginal, conditional PMF; functions of two RVs, expectation and variance)

# Problems from MIT OCW:

## Probabilistic Systems Analysis and Applied Probability

## Quiz 1 Spring 2009

# Problem 1

Which of the following statements is NOT true?

(a) If $A \subset B$, then $P(A) \leq P(B)$.

(b) If $P(B) > 0$, then $P(A|B) \geq P(A)$.

(c) $P(A \cap B) \geq P(A) + P(B) - 1$.

(d) $P(A \cap B^c) = P(A \cup B) - P(B)$.

# Problem 1

Which of the following statements is NOT true?

(a) If $A \subset B$, then $P(A) \leq P(B)$.

(b) If $P(B) > 0$, then $P(A|B) \geq P(A)$.

(c) $P(A \cap B) \geq P(A) + P(B) - 1$.

(d) $P(A \cap B^c) = P(A \cup B) - P(B)$.

**Solution: (b)**

A counter example: if we have two events $A$, $B$ such that $P(B) > 0$ and $P(A) > 0$, but $A \cap B = \emptyset$, then $P(A|B) = 0$, but $P(A) > P(A|B)$. It's easy to come up with examples like this: for example, take any sample space with event $A$ such that $P(A) > 0$, and $P(A^c > 0)$, it follows that $P(A|A^c) = 0$, but $P(A) > 0$.

# Problem 2

We throw $n$ identical balls into $m$ urns at random, where each urn is equally likely and each throw is independent of any other throw. What is the probability that the $i$-th urn is empty?

(a) $\left(1 - \frac{1}{m}\right)^n$

(b) $\left(1 - \frac{1}{n}\right)^m$

(c) $\binom{m}{n}\left(1 - \frac{1}{n}\right)^m$

(d) $\binom{n}{m}\left(\frac{1}{m}\right)^n$

# Problem 2

We throw $n$ identical balls into $m$ urns at random, where each urn is equally likely and each throw is independent of any other throw. What is the probability that the $i$-th urn is empty?

(a) $\left(1 - \frac{1}{m}\right)^n$
(b) $\left(1 - \frac{1}{n}\right)^m$
(c) $\binom{m}{n}\left(1 - \frac{1}{n}\right)^m$
(d) $\binom{n}{m}\left(\frac{1}{m}\right)^n$

**Solution: (a)**
The probability of the $j$th ball going into the $i$th urn is $1/m$. Hence, the probability of the $j$th ball not going into the $i$th urn is $(1 - 1/m)$. Since all throws are independent from one another, we can multiply these probabilities: the probability of all $n$ balls not going into the $i$th urn, i.e. it is empty, is $\left(1 - \frac{1}{m}\right)^n$.

# Problem 3

We toss two fair coins simultaneously and independently. If the outcomes of the two coins are the same, we win; otherwise, we lose. Let $A$ be the event that the first coin comes up heads, $B$ be the event that the second coin comes up heads, and $C$ be the event that we win. Which of the following statements is false?

(a) Events $A$ and $B$ are independent.

(b) Events $A$ and $C$ are independent.

(c) Events $A$ and $B$ are *not* conditionally independent given $C$.

(d) The probability of winning is $1/2$.

# Problem 3

We toss two fair coins simultaneously and independently. If the outcomes of the two coins are the same, we win; otherwise, we lose. Let $A$ be the event that the first coin comes up heads, $B$ be the event that the second coin comes up heads, and $C$ be the event that we win. Which of the following statements is false?

(a) Events $A$ and $B$ are independent.

(b) Events $A$ and $C$ are independent.

(c) Events $A$ and $B$ are *not* conditionally independent given $C$.

(d) The probability of winning is $1/2$.

**Solution: (b)**
The sample space in this case is $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. The probability law is a uniform distribution over this space. We have $A = \{(H, H), (H, T)\}$, $B = \{(H, H), (T, H)\}$, and $C = \{(H, H), (T, T)\}$. By the discrete uniform law, $P(A) = P(B) = P(C) = 1/2$. We also have $P(A \cap C) = 1/4$, hence $P(A \cap C) = P(A)P(C)$, and the two events are independent. Intuitively, knowing that you won adds no information about whether your coin turned up heads or not: stating this formally, we have $P(A|C) = P(A)$.

# Problem 4

For a biased coin, the probability of "heads" is $1/3$. Let $h$ be the number of heads in five independent coin tosses. What is the probability $P(\text{first toss is a head}|h = 1 \text{ or } h = 5)$?

(a) $\dfrac{\frac{1}{3}(\frac{2}{3})^4}{5\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}$

(b) $\dfrac{\frac{1}{3}(\frac{2}{3})^4}{\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}$

(c) $\dfrac{\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}{5\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}$

(d) $\dfrac{1}{5}$

# Problem 4

For a biased coin, the probability of "heads" is $1/3$. Let $h$ be the number of heads in five independent coin tosses. What is the probability
$P(\text{first toss is a head}|h = 1 \text{ or } h = 5)$?

(a) $\dfrac{\frac{1}{3}(\frac{2}{3})^4}{5\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}$

(b) $\dfrac{\frac{1}{3}(\frac{2}{3})^4}{\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}$

(c) $\dfrac{\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}{5\frac{1}{3}(\frac{2}{3})^4 + (\frac{1}{3})^5}$

(d) $\dfrac{1}{5}$

**Solution: (c)**

Let $A$ be the event that the first toss is a head.

$$
\begin{aligned}
P(A|\{H=1\} \text{ or } \{H=5\}) &= \frac{P(A \cap (\{H=1\} \cup \{H=5\}))}{P(\{H=1\} \cup \{H=5\})} \\
&= \frac{P((A \cap \{H=1\}) \cup (A \cap \{H=5\}))}{P(\{H=1\} \cup \{H=5\})} \\
&= \frac{P(\{H=1\}) + P(A \cap \{H=5\})}{P(\{H=1\}) + P(\{H=5\})} = \frac{(1/3)^5 + (1/3)(2/3)^4}{\binom{5}{1}(1/3)(2/3)^4 + \binom{5}{5}(1/3)^5}
\end{aligned}
$$

# Problem 5

A well-shuffled deck of 52 cards is dealt evenly to two players (26 cards each). What is the probability that player 1 gets all the aces?

(a) $\dfrac{\binom{48}{22}}{\binom{52}{26}}$

(b) $\dfrac{4\binom{48}{22}}{\binom{52}{26}}$

(c) $\dfrac{48!}{22!}\,\dfrac{52!}{26!}$

(d) $\dfrac{4!\binom{48}{22}}{\binom{52}{26}}$

# Problem 5

A well-shuffled deck of 52 cards is dealt evenly to two players (26 cards each). What is the probability that player 1 gets all the aces?

(a) $\dfrac{\binom{48}{22}}{\binom{52}{26}}$

(b) $\dfrac{4\binom{48}{22}}{\binom{52}{26}}$

(c) $\dfrac{48!}{22!}\dfrac{52!}{26!}$

(d) $\dfrac{4!\binom{48}{22}}{\binom{52}{26}}$

**Solution: (a)**

Let $A$ be the event that player 1 gets all aces. By the discrete uniform law,

$$P(A) = |A|/|\Omega|$$

$|\Omega| = \binom{52}{16}$ is the number of hands (26 cards from 52) player 1 can have. Additionally, once we have given player 1 all aces, then they must be given an additional 22 cards from the remaining 48 cards in the deck. Hence,

$$P(A) = \binom{48}{22} \Big/ \binom{52}{26}$$

# Problem 6

Suppose $X$, $Y$ and $Z$ are three independent discrete random variables. Then, $X$ and $Y+Z$ are

(a) always independent

(b) sometimes independent

(c) never independent

# Problem 6

Suppose $X$, $Y$ and $Z$ are three independent discrete random variables. Then, $X$ and $Y + Z$ are

(a) always independent

(b) sometimes independent

(c) never independent

**Solution: (a)**
Since $X$ is independent of $Y$ and $Z$, $X$ is independent of $g(Y, Z)$ for any function $g(Y, Z)$, including $g(Y, Z) = Y + Z$.

# Problem 7

To obtain a driving license, Mina needs to pass her driving test. Every time Mina takes a driving test, with probability $1/2$, she will clear the test independent of her past. Mina failed her first test. Given this, let $Y$ be the additional number of tests Mina takes before obtaining a license. Then,

(a) $E[Y] = 1$

(b) $E[Y] = 2$

(c) $E[Y] = 0$

# Problem 7

To obtain a driving license, Mina needs to pass her driving test. Every time Mina takes a driving test, with probability $1/2$, she will clear the test independent of her past. Mina failed her first test. Given this, let $Y$ be the additional number of tests Mina takes before obtaining a license. Then,

(a) $E[Y] = 1$

(b) $E[Y] = 2$

(c) $E[Y] = 0$

**Solution: (b)**

$Y$ is defined as the number of additional tests Mina takes, so this is independent of the fact that she failed her first test. $Y$ is a geometric RV with $p = 1/2$. Hence, $E[Y] = 1/p = 2$.

# Problem 8

Let $X_i$, $1 \leq i \leq 4$ be independent Bernoulli random variable each with mean $p = 0.1$. Let $X = \sum_{i=1}^{4} X_i$. That is, $X$ is a Binomial random variable with parameters $n = 4$ and $p = 0.1$. Then,

(a) $E[X_1 | X = 2] = 0.1$

(b) $E[X_1 | X = 2] = 0.5$

(c) $E[X_1 | X = 2] = 0.25$

# Problem 8

Let $X_i$, $1 \leq i \leq 4$ be independent Bernoulli random variable each with mean $p = 0.1$. Let $X = \sum_{i=1}^{4} X_i$. That is, $X$ is a Binomial random variable with parameters $n = 4$ and $p = 0.1$. Then,

(a) $E[X_1|X = 2] = 0.1$

(b) $E[X_1|X = 2] = 0.5$

(c) $E[X_1|X = 2] = 0.25$

**Solution: (b)**
We have $P(X_1 = 1|X = 2) = 0.5$, because

$$P(X_1 = 1|X = 2) = \frac{P(X_1 \cap X = 2)}{P(X = 2)} = \frac{p \cdot \binom{3}{1} p(1-p)^2}{\binom{4}{2} p^2 (1-p)^2} = \frac{\binom{3}{1}}{\binom{4}{2}} = 0.5$$

Note that $\binom{4}{2} p^2 (1-p)^2$ is the probability of seeing 2 heads out of 4 tosses, and $\binom{3}{1} p(1-p)^2$ is the probability of seeing 1 head in the last 3 tosses.

Hence,

$$E[X_1|X = 2] = 1 \cdot P(X_1 = 1|X = 2) + 0 \cdot P(X_1 = 0|X = 2) = 0.5$$

# Lectures' summaries

# How do we reason under uncertainty?

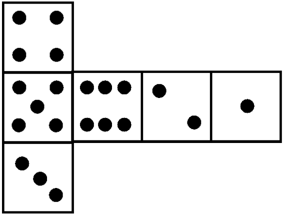
- Using Probability Theory
- Main idea: Assign each event a measure between 0 to 1: to signify its likelihood
- Then proceed very carefully - or our intuitions and observations will not match

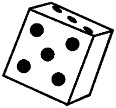
# Set theory

# Back to basics: Set theory

- A set is a collection of objects, which are the elements of the set
- If $S$ is a set and $x$ is an element of $S$, we write $x \in S$.
- If $x$ is not an element of $S$, we write $x \notin S$. A set can have no elements, in which case it is called the empty set, denoted by $\emptyset$.
- Apple $\in$ { Orange, Apple, Pear }
  Strawberry $\notin$ { Orange, Apple, Pear }

# Back to basics: Set theory





- Two ways of writing a set down:

$$S = \{1, 2, 3, 4, 5, 6\}$$

or

$$S = \{x | x \text{ is a possible outcome of a throw of a die}\}$$

"The collection of all elements that satisfy a certain condition is a set"

# Set theory

- Size of a set $S$ is denoted by $|S|$
- $|\{$ Orange, Apple, Pear $\}| = 3$
- $S$ is a subset of $T$, $S \subset T$, means every element of $S$ is also an element of $T$:
  $\forall x \in S, x \in T$
  - $\{$Apple, Pear $\} \subset \{$ Orange, Apple, Pear $\}$
  - $\{$ Orange, Apple, Pear $\} \subset \{$ Orange, Apple, Pear $\}$
  - $\{$Apple, Banana $\} \not\subset \{$ Orange, Apple, Pear $\}$
- If $S \subset T$ and $T \subset S$ then,

$$S = T$$

# Universal Set

- $\Omega$: contains all objects that could conceivably be of interest in a particular context.
- In the context of coin tossing, $\Omega = \{H, T\}$.
- In the context of dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$.

# Set operations

- Complement: $S^c = \{x \in \Omega | x \notin S\}$
  Example: $\Omega = \{1, 2, 3, 4, 5, 6\}; \quad S = \{2, 5\}$
  $S^c = \{1, 3, 4, 6\}$
  Note that, $\Omega^c = \emptyset$
- The union of two sets $S$ and $T$ is the set of all elements that belong to $S$ or $T$ (or both), and is denoted by $S \cup T$.

$$S \cup T = \{x | x \in S \text{ or } x \in T\}$$

- The intersection of two sets $S$ and $T$ is the set of all elements that belong to both $S$ and $T$ , and is denoted by $S \cap T$.

$$S \cap T = \{x | x \in S \text{ and } x \in T\}$$

# Power Set

By default: $\emptyset \subset S \subset \Omega$.
**Power Set**: Set of all subsets

$$S = \{1, 2, 3\}$$

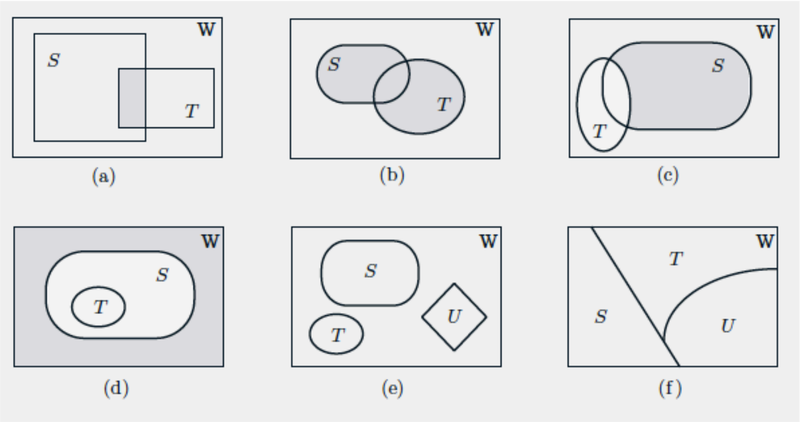
$$2^S = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

# Disjoint Set

$S$ and $T$ are **disjoint** if $S \cap T = \emptyset$

$S_1, S_2, \ldots, S_n$ forms a **partition** of $S$ if $S_i$ and $S_j$ are disjoint for any $i \neq j$ and $S_1 \cup S_2 \cup \cdots \cup S_n = S$.

# Venn Diagram




Courtesy: Textbook

# Venn Diagram - Partitions

- In (e), $S$, $T$, and $U$ **do not** form partitions of $W$. However, $S$, $T$, $U$, and $(S \cup T \cup U)^C$ form partitions of $W$.
- In (f), $S$, $T$, and $U$ form partitions of $W$.

Similarly,

- In (a),(b) and (c), $S$ and $T$ do not form partitions of $W$. However, $(S \cup T)$ and $(S \cup T)^C$ form partitions of $W$.
- In (d), $T$ and $T^C \cap S$ form partitions of $S$. Furthermore, $S$ and $S^C$ form partitions of $W$.

# Set Algebra

Using the above definitions, we can show that:

- Intersection Commutativity $S \cap T = T \cap S$
- Union Commutativity $S \cup T = T \cup S$
- Intersection Associativity $S \cap (T \cap U) = (S \cap T) \cap U$
- Union Associativity $S \cup (T \cup U) = (S \cup T) \cup U$
- Intersection Distributivity $S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$
- Union Distributivity $S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$

# Summary: Sets

- A set is a collection of objects, which are the elements of the set
- $x \in S$, $x \notin S$, empty set $\emptyset$, number of elements in a set $|S|$
- Subset: $S \subset T$
- Universal set $\Omega$, set complement: $S^c = \{x \in \Omega | x \notin S\}$
- Set union $S \cup T = \{x | x \in S \text{ or } x \in T\}$, intersection $S \cap T = \{x | x \in S \text{ and } x \in T\}$
- **Power Set**: Set of all subsets
- Disjoint sets $S \cap T = \emptyset$
- Partition of a set: $S_i$ and $S_j$ are disjoint for any $i \neq j$ and $S_1 \cup S_2 \cup \cdots \cup S_n = S$

# Model of Probability

# Model of Probability

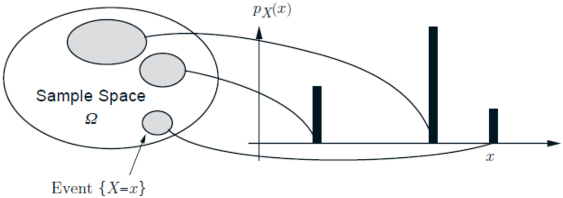
A probabilistic model is a mathematical description of an uncertain situation. Two fundamental elements of a probabilistic model are

- **Sample Space** $\Omega$: all possible outcomes of an experiment
- **Probability Law**:

$$A \subset \Omega; \quad P(A),$$

  where $A$ is an **event** (a set of possible outcomes) and $P(A)$ is a non-negative number presenting the **likelihood** of observing the event $A$.

Probabilistic model involves an **experiment**, which produces an **event** from the sample space.



# Probability Laws

- Probability represents likelihood of any outcomes or of any set of possible outcomes.
- The probability law assigns to every event $A$, a number $P(A)$, call the **probability** of $A$.

# Axioms of Probability

- **Nonnegativity**:

$$P(A) \geq 0$$

- **Additivity**: For any two disjoint sets $A$ and $B$,

$$P(A \cup B) = P(A) + P(B)$$

Holds for infinitely many disjoints events $A_1, A_2, A_3, \ldots$

$$P(\cup_i A_i) = \sum_i P(A_i).$$

- **Normalization**:

$$P(\Omega) = 1$$

# Discrete Probability Models

If $\Omega$ consists of a finite number of possible outcomes, we are dealing with discrete probability models.
For example,

- Coin Toss
- Dice Rolling

For discrete probabilistic models, the probability law is specified by the probabilities of the events that consists of a single element (that are disjoint by nature).

$$A = \{s_1, s_2, \ldots, s_n\} \subset \Omega$$

$$P(\Omega) = P(\{s_1, s_2, ..., s_n\}) = P(s_1) + P(s_2) + ... + P(s_n)$$

# Uniform Discrete Model

If $\Omega$ is finite and all possible outcomes are equally likely, it is a **uniform discrete model**. Then, the probability of each element of $\Omega$ has the probability of

$$\frac{1}{|\Omega|}$$

More generally, $\forall A \subset \Omega$

$$P(A) = \frac{|A|}{|\Omega|}$$

# Uniform Discrete Model - Example

Throwing a fair die is an example of a uniform discrete model.
$\Omega = \{1, 2, 3, 4, 5, 6\}$ Uniform model:

$$P(\{i\}) = \frac{1}{|\Omega|} = \frac{1}{6}$$

for $i = 1, 2, 3, 4, 5, 6$.
$A$: even number shows up

$$A = \{2, 4, 6\}$$

$$|A| = 3$$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

# Conditional Probability

# Conditional Probabilities

Conditional probability provides us with a way to reason about the outcome of an experiment based on **partial information** or **observations**.

Consider rolling a fair die. What is the probability that the outcome is 6 **given that** we know that the outcome is an even number.

- Suppose that you rolled a die while blindfolding yourself. Your friend next to you told you that the number is even. Does that change your probability space?

We can express this conditional probability using $P(A|B)$: conditional probability of $A$ given $B$, where $P(B) > 0$.
In the above example,

- $A = \{$ The outcome is 6 $\}$
- $B = \{$ The outcome is an even number $\}$

# Conditional Probabilities

- A new probability space to be defined.
- The universe (sample space) has been changed to $B$
- The probability has now to be normalized by $P(B)$

Definition of conditional probability,

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

- If $A$ and $B$ are disjoint, i.e., $A \cap B = \emptyset$, then $P(A|B) = 0$. Why?
  - In the case of disjoint $A$ and $B$, $A \cap B = \emptyset$.
  - Which means, $P(A \cap B) = 0$. So $P(A|B) = 0$.

# New probability space $P(\cdot|B)$

Verify that the axioms of probability are satisfied!

- **Nonnegativity**: $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$ since $P(A \cap B) \geq 0$
- **Additivity**: For any two disjoint sets $A$ and $C$, show that $P(A \cup C|B) = P(A|B) + P(C|B)$.

$$
\begin{aligned}
P(A \cup C|B) &= \frac{P((A \cup C) \cap B)}{P(B)} \\
&= \frac{P((A \cap B) \cup (C \cap B)))}{P(B)} = \frac{P(A \cap B) + P(C \cap B)}{P(B)} \\
&= P(A|B) + P(C|B).
\end{aligned}
$$

- **Normalization**: New sample space is $B$.

$$P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

.

# Example

Let us have two unfair coin tosses where the joint probability has

$$P(\{HH\}) = 1/2$$

$$P(\{HT\}) = 1/4$$

$$P(\{TH\}) = 1/8$$

$$P(\{TT\}) = 1/8$$

What is the probability that we have exactly one $H$ **given that** the second toss shows $H$?

Define $A$ and $B$ first.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{TH\})}{P(\{HH, TH\})} = \frac{1/8}{1/2 + 1/8} = \frac{1}{5}.$$

# Another Exercise

Example: Throw of two dice. Each of the 36 outcomes are equally likely

- $A =$ max of two dice is less than 5
- $B =$ min of the two dice is greater than 1

What is $P(A|B)$?

# Another Exercise

Example: Throw of two dice. Each of the 36 outcomes are equally likely

- $A =$ max of two dice is less than 5
- $B =$ min of the two dice is greater than 1

What is $P(A|B)$?

- $P(A) = \frac{16}{36}$
- $P(B) = \frac{25}{36}$
- $P(A \cap B) = \frac{9}{36}$
- $P(A|B) = \frac{9}{25}$

# Sequential Model for Conditional Probabilities

Many experiments have a sequential characteristic: the future outcomes depending on the past.

For example, consider an example involving three coin tosses.

- The first toss is unbiased (fair): $P(H) = 0.5$ and $P(T) = 0.5$.
- Based on the outcome of the first toss, the second toss is biased towards that outcome by 60%.
  - For example, if the outcome of the first toss is $H$, then the second toss has $P(H) = 0.6$ and $P(T) = 0.4$.
- Based on the outcome of the second toss, the third toss is biased towards that outcome by 70%.

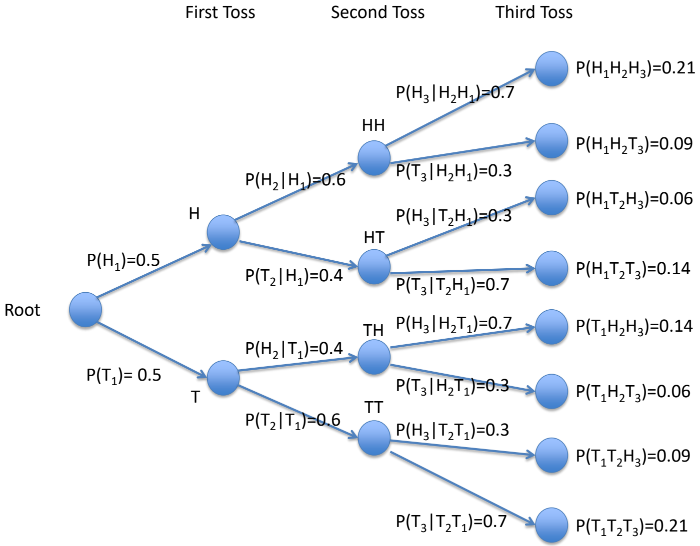
Let us draw a **tree-based sequential description**.

# Sequential Model for Conditional Probabilities

How to setup a tree-based sequential description and use it?

1. Leaves represent events of interest, which occur in a sequential manner
2. Branches represent the conditional probability
3. The probability of the end-leaf can be computed by multiplying conditional probabilities from the root.

# Sequential Model for Conditional Probabilities



# Multiplication Rule

- We learned conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

which can be re-written as

$$P(A \cap B) = P(B)P(A|B) = P(A)(B|A)$$

- Now, what about

$$P(A \cap B \cap C) =?$$

# Multiplication Rule

- We learned conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

which can be re-written as

$$P(A \cap B) = P(B)P(A|B) = P(A)(B|A)$$

- Now, what about

$$\begin{aligned} P(A \cap B \cap C) &= P((A \cap B) \cap C) \\ &= P(D \cap C), \text{ where } D = (A \cap B) \\ &= P(D)P(C|D) \\ &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

These are other equivalent results for $P(A \cap B \cap C)$.

# Multiplication Rule

In general,

$$
\begin{aligned}
P(\cap_{i=1}^{n} A_i) &\equiv P(A_1 \cap A_2 \cap \ldots A_n) \\
&= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n| \cap_{i=1}^{n-1} A_i)
\end{aligned}
$$

# Total Probability Theorem
# and Bayes' Rule

# Total Probability Theorem

- Let $A_1, A_2, \ldots, A_n$ form a partition of $\Omega$ and $P(A_i) > 0$
- Then, for any event $B$, we have

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n). \end{aligned}$$

# Bayes' Rule

Let $A_1, A_2, \ldots, A_n$ partition $\Omega$ and $P(A_i) > 0$. For any $B$ such that $P(B) > 0$,

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i)P(B|A_i)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)} \end{aligned}$$

# Independence

# Independence

- Consider flipping a fair coin twice in a row.
- If we know the coin is fair, does knowing the result of the first flip give us any information about the result of the second flip?
- What's the probability the coin comes up heads on the second flip?
- What's the probability the coin comes up heads on the second flip given that it came up heads on the first flip?

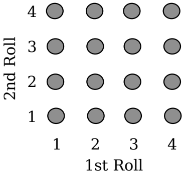
# Probabilistic Independence

- Intuitively, when knowing that one event occurred doesn't change the probability that another event occurred or will occur, we say that the two events are *probabilistically independent*.
- We say that two events $A$ and $B$ are independent if and only if (iff)

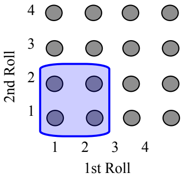
$$P(A \cap B) = P(A)P(B) \ .$$

  and this implies that $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

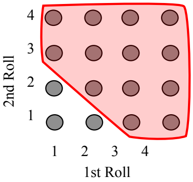
- **Question:** Suppose you roll two fair four sided dice. Are the events $A =$"maximum is less than 3" and $B=$"sum is greater than 3" independent?



Sample Space



$|A| = 4$



$|B| = 13$

- **Answer 2:** Formally,

$$P(A \cap B) = \frac{1}{16}\ ,\ \ P(A) = \frac{1}{4}\ \text{ and }\ P(B) = \frac{13}{16}\ .$$

Since $\frac{1}{16} \neq \frac{1}{4} \cdot \frac{13}{16}$, the events are not independent.

# An Event and Its Complement

- **Question:** Are $A$ and $A^c$ independent if $0 < P(A) < 1$?
- **Answer 1:** Intuitively, no. If you know $A$ happens, then you know $A^C$ does not happen.
- **Answer 2:** Formally, $P(A \cap A^C) = P(\emptyset) = 0$. If $0 < P(A) < 1$, then

$$P(A)P(A^C) \neq 0 \ .$$

# Independence of Three Events

- Three events $A$, $B$, and $C$ are independent iff:

$$P(A \cap B) = P(A)\,P(B)$$
$$P(A \cap C) = P(A)\,P(C)$$
$$P(B \cap C) = P(B)\,P(C)$$
$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

- First three conditions imply that any two events are independent (known as *pairwise independence*)
- *Pairwise independence* does not imply the *independence of all events*.
- Suppose we have a finite collection of events $A_1, A_2, ..., A_n$. These events are said to be independent iff

$$P\left(\cap_{i \in \mathcal{S}} A_i\right) = \prod_{i \in \mathcal{S}} P(A_i), \text{ for every subset } \mathcal{S} \text{ of } \{1, 2, ..., n\}$$

# Conditional Independence

- $A$ and $B$ are *conditionally independent* given $C$ iff

$$P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$$

- This is equivalent to $P(A \mid B \cap C) = P(A \mid C)$, assuming that $P(B \mid C) > 0$.
  - If $C$ is given, additional information of knowing $B$ has occurred does not change the conditional probability of $A$.
- This is equivalent to $P(B \mid A \cap C) = P(B \mid C)$, assuming that $P(A \mid C) > 0$.
  - If $C$ is given, additional information of knowing $A$ has occurred does not change the conditional probability of $B$.

# Counting

# Counting and Discrete Probability Laws

- If $\Omega$ is finite and all outcomes are equally likely, then

$$P(A) = \frac{|A|}{|\Omega|}$$

.

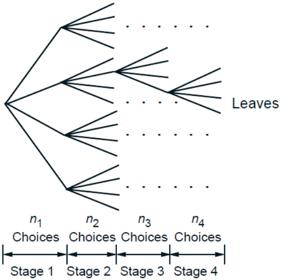
- The calculation of probabilities often involve **counting** the number of outcomes in various events.
- Sometimes it's challenging to compute $|A|$ and $|\Omega|$ and they are too large work out by hand...

We covered different counting methods:

- *Permutations*
- *k-Permutations*
- *Combinations*
- *Partitions*

# The Counting Principle

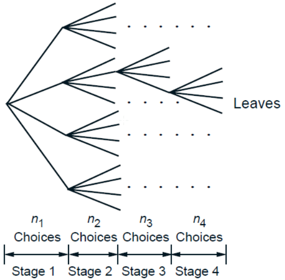
- Consider a sequential process with $s$ stages. At each stage $i$, there are $n_i$ possible results. How many outcomes does the process have?




- How many possible outcomes are possible from a sequence of $s$ stages?

# The Counting Principle

- Consider a sequential process with $s$ stages. At each stage $i$, there are $n_i$ possible results. How many outcomes does the process have?




- How many possible outcomes are possible from a sequence of $s$ stages?

$$n_1 \times n_2 \times \cdots \times n_s = \prod_{i=1}^{s} n_i .$$

# Counting Permutations

- Let $S$ be a set of $n$ objects.
- Consider an $n$-stage experiment where at each stage we choose one object without replacement.
  - We pick objects until there's no more objects to pick.
- This process produces an **ordering** or **permutation** of the $n$ objects.
  - For example, if $n = 3$ and $S = \{a, b, c\}$, one ordering can be $bac$.
- This is an $n$ stage process. We have $s_1 = n$, $s_2 = n - 1$,..., $s_n = 1$.
- By the counting principle, the number of permutations is

$$n(n-1)(n-2)\cdots 1 = n!$$

- For permutations, order matters, i.e., $abc \neq bac$.

# Counting $k$-Permutations

- Let $S$ be a set of $n$ objects.
- Consider a $k$-stage experiment where $k \leq n$. At each stage we choose one object without replacement.
  - We pick only $k$ objects.
- This process produces an ordering of the $k$ objects, which is also called a $k$-permutation.
  - For example, if $n = 3, k = 2$, and $S = \{a, b, c\}$, one possible 2-permutation is $ba$ and another is $ab$.
- This is a $k$-stage process where $s_1 = n$, $s_2 = n - 1$,..., $s_k = n - k + 1$.
- By the counting principle, the number of permutations is

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

- Order also matters for $k$-permutations.

# Counting Combinations

- Let $S$ be a set of $n$ objects. How many subsets of size $k$ are there?
- The number of $k$-permutations is $n!/(n-k)!$ but this over counts the number of subsets, e.g., $ab$ and $ba$ are different 2-permutations of $\{a, b, c\}$, but the same subset $\{a, b\}$.
  - Order does **NOT** matters for combinations.
- $k!$ different $k$-permutations belong to the same subset of $k$ objects, so the number of "$k$-combinations" is

$$\frac{\frac{n!}{(n-k)!}}{k!} = \frac{n!}{(n-k)!k!},$$

  which is denoted $\binom{n}{k}$, pronounced as "$n$ choose $k$".
- Note that $\binom{n}{0} = 1$

# Counting Partitions

- A combination divides items into one group of $k$ and one group of $n - k$. Thus, a combination can be viewed as a partition of the set in two.
- Consider an experiment where we divide $n$ objects into $\ell$ groups with sizes $n_1, n_2, ..., n_\ell$ such that $n = \sum_{i=1}^{\ell} n_i$.
- How many partitions are there?
- There are $\binom{n}{n_1}$ ways to choose the objects for the first partition. This leaves $n - n_1$ objects. There are $\binom{n-n_1}{n_2}$ ways to choose objects for the second partition. There are $\binom{n-n_1-n_2-...-n_{\ell-1}}{n_\ell}$ ways to choose the objects for the last group.

# Counting Partitions

- Using the counting principle, the number of partitions is thus:

$$\binom{n}{n_1} \cdot \binom{n - n_1}{n_2} \cdots \binom{n - n_1 - n_2 - \ldots - n_{\ell-1}}{n_\ell}$$

$$= \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdots \frac{(n-n_1-n_2-\ldots-n_{\ell-1})!}{n_\ell!(n-n_1-n_2-\ldots-n_\ell)!}$$

- Note that $(n - n_1 - n_2 - \ldots - n_\ell)! = 0! = 1$.
- Canceling terms yields the final result:

$$\frac{n!}{n_1! \cdots n_\ell!}$$

# Summary of Counting Problems

| Structure | Description | Order Matters | Formula |
|---|---|---|---|
| Permutation | Number of ways to order $n$ objects | Yes | $n!$ |
| $k$-Permutation | Number of ways to form a sequence of size $k$ using $k$ different objects from a set of $n$ objects | Yes | $\frac{n!}{(n-k)!}$ |
| Combination | Number of ways to form a set of size $k$ using $k$ different objects from a set of $n$ objects | No | $\frac{n!}{k!(n-k)!}$ |
| Partition | Number of ways to partition $n$ objects into $\ell$ groups of size $n_1, \ldots, n_\ell$ | No | $\frac{n!}{n_1!\ldots n_\ell!}$ |

# The Binomial Law

- If we toss $n$ coins, what's the probability of seeing $k$ heads, denoted as $P_n(k)$? (without exhaustively enumerating all sequences?)
- Any single sequence of length $n$ with $k$ heads has probability

$$p^k(1-p)^{n-k} \ .$$

- But how many different sequences of length $n$ contain $k$ heads?

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

  where $\binom{n}{0} = 1$.
- Thus,

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

# The Binomial Law

- The following equation is often called the **binomial probabilities**.

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $\binom{n}{k}$ is referred to as **binomial coefficient**.

# Discrete random variables
# and
# Probability mass functions

# Random Variables Give An Easy Way to Specify Events

- If we have a function $X : \Omega \rightarrow \mathbb{R}$, we can use it to construct a different event for each value of $x \in \mathbb{R}$:

$$\{X = x\} = \{o | o \in \Omega \text{ and } X(o) = x\}$$

- In the dice example, the event $\{X = x\}$ is the set of outcomes $o \in \Omega$ that are mapped to the the same value $x$ by the function $X$.
  For example,

$$\{X = 2\} = \{(1,2),(2,1),(2,2)\}$$

$$\{X = 3\} = \{(1,3),(2,3),(3,3)(3,1),(3,2)\}$$

$$\{X = 1\} = \{(1,1)\}$$

# Discrete Random Variables and Probability

- A random variable is called **discrete** if its input (sample space) is either finite or countably infinite.
- We can compute the probability of an event $\{X = x\}$ by decomposing it into atomic events and using the probability rule:

$$P(X = x) = p_X(x) = P(\{o | o \in \Omega \text{ and } X(o) = x\})$$

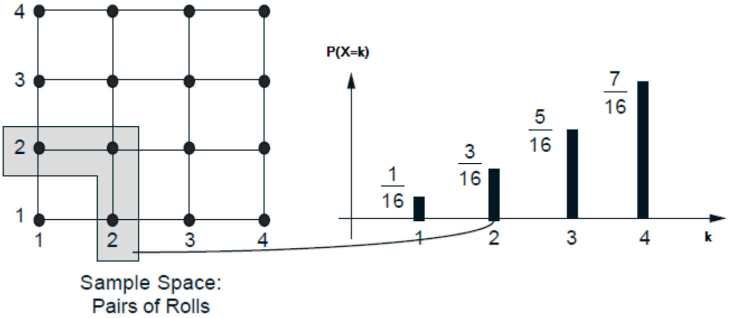
- *Probability law:* A function $p_X(x)$ that maps event to a number between 0 and 1 that satisfies the probability axioms:
  1. *Nonnegativity:* $p_X(x) \geq 0, \forall x$.
  2. *Normalization:* $\sum_x p_X(x) = 1$.

# Example: Maximum of Dice Rolls

- For example, in the event of $\{X = 2\}$ for the dice rolling example where $X(r_1, r_2) = \max(r_1, r_2)$

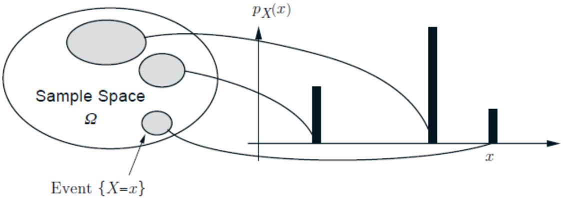
$$
\begin{aligned}
P(X = 2) &= P(\{(1,2),(2,1),(2,2)\}) \\
&= P((1,2)) + P((2,1)) + P((2,2)) = 3/16
\end{aligned}
$$

- We can work out the probability for all possible values of $x$:



# In general...

- The probability associated with the event $\{X = x\}$ for each element $x \in \mathbb{R}$ of a discrete random variable $X$ is referred to as the **probability mass function** or **PMF** of the random variable.
- The PMF is denoted by $P(X = x)$ or $p_X(x)$.




  - The x-axis represents all possible outcomes of the event
  - The y-axis represents the associated probabilities

# Common Discrete Random Variables

# Discrete Uniform Random Variables

- A *discrete uniform random variable* $X$ with range $[a, b]$ takes on any integer value between (and including) $a$ and $b$ with the same probability
- For example, the random variable that maps a fair six-sided dice roll to the number that comes up is a uniform random variable with $a = 1$, $b = 6$ and $P(X = k) = 1/6$ for $k = 1, ..., 6$.
- The PMF of a discrete uniform random variable $X$ is

$$P(X = k) = \frac{1}{b - a + 1} \text{ for } k = a, \ldots, b$$

- Used to model probabilistic situations where each of the values $a, ..., b$ are equally likely.

# Bernoulli Random Variables

- Suppose we have an experiment with two outcomes $H$ and $T$. $H$ happens with probability $p$ and $T$ with probability $1 - p$, $0 < p < 1$.
- We define a random variable $X$ such that $X(H) = 1$ and $X(T) = 0$.
- This is called a Bernoulli random variable $X$ that takes the two values 0 or 1.
- Its PMF looks like

$$P(X = k) = \begin{cases} 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

- You can also define $X(H) = 0$ and $X(T) = 1$, with $P(X = 1) = p' = 1 - p$

# Bernoulli Random Variables: Examples

- Whether a coin lands heads or tails.
- Whether a server is online or offline.
- Whether an email is spam or not.
- Whether a pixel in a black and white image is black or white.
- Whether a patient has a disease or not.

# Binomial Random Variable

- A binomial random variable is the combination of independent and identically distributed Bernoulli random variables
- Suppose we flip $n$ coins independently, where each coin has probability $p$ of being heads
- The set of outcomes is:

$$\Omega = \{(TTT \dots TT), (TTT \dots TH), \dots, (HHH \dots HH)\}$$

- Define a random variable $X$ where for each $o \in \Omega$,

$$X(o) = \text{ “the number of heads in outcome } o\text{”}$$

- We're already shown that $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

# Binomial Random Variables: Examples

- The number of heads in $N$ coin tosses.
- The number of servers that fail in a cluster of $N$ servers.
- The number of games a football team wins in a season of $N$ games (assuming i.i.d.).
- The number of True/False questions you get correct if you guess each of $N$ questions.

# Geometric Random Variables

- Suppose we flip a biased coin repeatedly until it lands heads. Let $X$ be the number of tosses needed for a head to come up for the first time.
- The PMF of a *geometric random* variable $X$ is

$$P(X = k) = (1 - p)^{k-1} \cdot p \quad \text{for } k = 1, 2, 3, \ldots$$

- Used to model the number of repeated independent trials up to (and including) the *first "successful" trial*.
- Example: the number of patients we test before the first one we find who has a given disease.

# Geometric Random Variables: Example

- Products made by a machine have a 3% defective rate.
- What is the probability that the first defect occurs in the fifth item inspected?

$$P(X = k) = (1 - p)^{k-1} \cdot p = (1 - 0.03)^{5-1} \cdot 0.03 = 0.0265 \ldots$$

# Poisson Random Variables

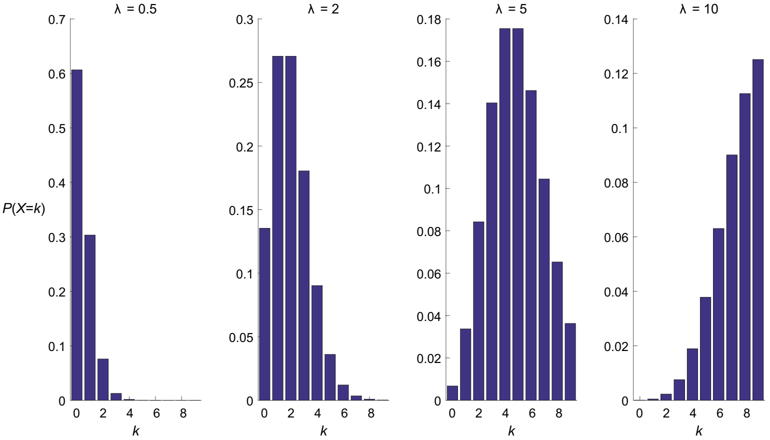
- A *Poisson random variable* $X$ is a random variable that has the following PMF

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

- The Poisson distribution is one of the most widely used probability distributions.
- Built based on Taylor series: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.
- Think about Poisson RV as a framework that provides approximation of a real-life random variable as a function of $\lambda$.

# Poisson Random Variables

- Think about Poisson RV as a framework that provides approximation of different PMFs as a function of $\lambda$.



# Poisson Random Variables

- It is generally used in scenarios where we are counting the occurrences of certain events within an interval of time or space.
  - The number of typos in a book with $n$ words.
  - The number of cars that crash in a city on a given day.
  - The number of phone calls arriving at a call center per minute etc.
- $\lambda$ represents the expected number of events (we will learn more about this).
  - The average number of typos in a book.
  - The average number of car crash per day.
  - The average number of phone calls per minute.

# Poisson Random Variables: Example

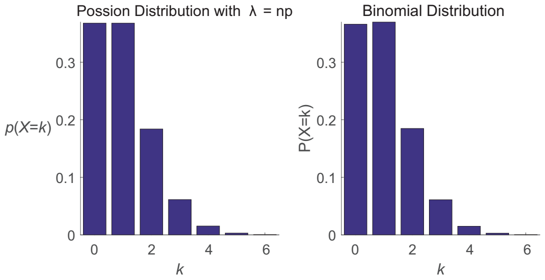
- Suppose that the number of phone calls arriving at a call center per minute can be modeled by a discrete Poisson PMF.
- In average, the call center receives 10 calls.
- What is the probability that the center will receive 5 calls?

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$P_X(k) = e^{-10} \frac{10^5}{5!} = 0.0378...$$

# Poisson Random Variables

- A Poisson PMF with $\lambda = np$ is a *good approximation* for a binomial PMF with very small $p$ and very large $n$ if $k \ll n$
  - A bionomial RV $X$ is the number of heads $(k)$ in the $n$-toss sequence, where the coin comes up a head with probability $p$.
- Example: $n = 100$ and $p = 0.01$ for the binomial r.v. where as $\lambda = np$ for the Poisson r.v.




- Poisson PMF provides much simpler models and calculations: $\binom{n}{k} p^k (1-p)^{n-k}$ **vs.** $e^{-\lambda} \frac{\lambda^k}{k!}$

# Summary: Discrete Random Variables

- **Uniform:** For $k = a, \ldots, b$:

$$P(X = k) = \frac{1}{b - a + 1}$$

- **Bernoulli:** For $k = 0$ or $1$:

$$P(X = k) = \begin{cases} 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

- **Binomial:** For $k = 0, \ldots, N$

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

- **Geometric:** For $k = 1, 2, 3, \ldots$, $P(X = k) = (1 - p)^{k-1} \cdot p$
- **Poisson:** $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$

# Expectation and Variance

# Expected Value

- For a random variable $X$, the expected value is defined to be:

$$E[X] = \sum_{x \in \mathbb{R}} x\, P(X = x)$$

  i.e., the probability-weighted average of the possible values of $X$.
- $E[X]$ is also called the **expectation** or **mean** of $X$.
- Why do we care to know about the expected value?
- Given a certain PMF, what is the "average" outcome that I am expecting to have?
- For example, if I bet the same amount of money on roulette and play it for a long-term period, how much do I expect to make?

# Expected Value: Question

- Expectation:

$$E[X] = \sum_{k \in \mathbb{R}} k\, P(X = k)$$

- If $X$ maps to $\{1, 2, 6\}$ and

$$P(X = 1) = 1/3 \quad , \quad P(X = 2) = 1/2 \quad , \quad P(X = 6) = 1/6$$

then $E[X] = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} + 6 \cdot \frac{1}{6} = 2.33\ldots$

# Expectations of Common Random Variables

- **Uniform on** $\{a, a+1, \ldots, b\}$**:** $E[X] = \frac{a+b}{2}$
- **Bernoulli:** $E[X] = (1-p) \cdot 0 + p \cdot 1 = p$
- **Binomial:** $E[X] = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = np$
- **Geometric:** $E[X] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p = \frac{1}{p}$
- **Poisson:** $E[X] = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}}{k!} \lambda^k = \lambda$

# Properties of Expectation

- **Linearity of Expectation:** If $a$ and $b$ are any real values, then the expectation of $aX + b$ is:

$$E[aX + b] = a \cdot E[X] + b$$

- **Expectation of Expectation:** Applying the expectation operator more than once has no effect. $E[E[X]] = E[X]$ since $E[X]$ is already a constant.

# Variance

- **Definition**: Variance measures how far we expect a random variable to be from its average:

$$\mathrm{var}(X) = E[(X - E[X])^2] = \sum_k (k - E[X])^2 \cdot P(X = k)$$

- An equivalent definition is

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

- **Definition**: we generally define the $\mathbf{n^{th}}$ **moment** of $X$ as $E[X^n]$, the expected value of the random variable $X^n$.

# Variance of Common Random Variables

- **Bernoulli:** $\text{var}[X] = p(1-p)$
- **Binomial:** $\text{var}[X] = np(1-p)$
- **Geometric:** $\text{var}[X] = \frac{1-p}{p^2}$
- **Uniform:** $\text{var}[X] = \frac{(b-a+1)^2-1}{12}$
- **Poisson:** $\text{var}[X] = \lambda$

# Standard Deviation

- The term **standard deviation** simply refers to the positive square root of the variance, which always exists and is also positive:
$$\text{std}(X) = \sqrt{\text{var}(X)}$$
- The standard deviation is **also** a measure of dispersion around the mean.
- One reason that people like to report standard deviations instead of variances is that **the units are the same as** $X$.
- $\text{var}(X) = E[(X - E[X])^2]$ vs. $\text{std}(X) = \sqrt{E[(X - E[X])^2]}$
- Example 1: If $X$ is height in feet, then $\text{var}(X)$ has units in square feet while $\text{std}(X)$ again has units in feet.

# Functions of Random Variables

# Functions of Random Variables

- If $X$ is a random variable and $f : \mathbb{R} \to \mathbb{R}$ then

$$Y = f(X)$$

is also a random variable with PMF:

$$P(Y = k) = P(f(X) = k) = \sum_{o \in \Omega \text{ with } f(X(o))=k} P(o)$$

- Example, let $X$ represent an outcome from a 6 sided fair die where

$$P(X = i) = 1/6, \forall i.$$

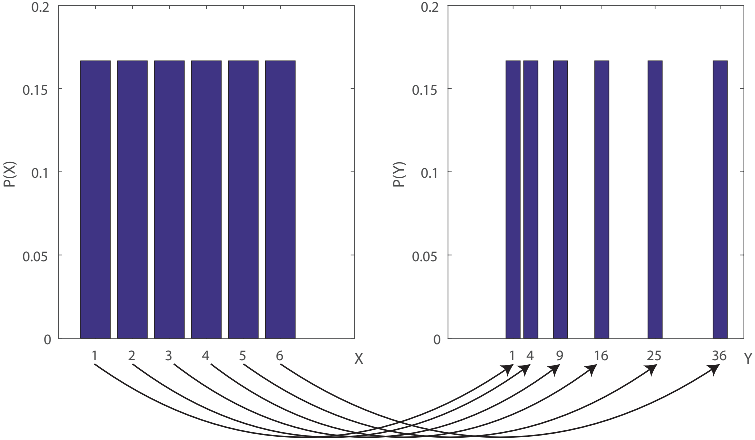
Suppose that you will receive money that is the square of the outcome, and we define a r.v. $Y$ as the amount of money.

- This function $Y = f(X)$ can be expressed as

$$Y = X^2$$

# Functions of Random Variables

- Note that $Y$ is also a random variable, whose PMF looks like.



# Functions of Random Variables

- Expectation of $Y = f(X)$:

$$E[Y] = \sum_y yP(Y = y) = \sum_y yP(X = f^{-1}(y))$$

$$= \sum_x f(x)P(X = x)$$

- For the previous example,

$$E[X] = 1 \times 1/6 + 2 \times 1/6 + \cdots + 6 \times 1/6 = 3.5$$

$$E[Y] = 1^2 \times 1/6 + 2^2 \times 1/6 + \cdots + 6^2 \times 1/6 = \$15.2$$

# Functions of Random Variables

- Variance of $Y$:

$$var[Y] = E[(Y - E[Y])^2] = \sum_{k \in \{1,4,\ldots,36\}} (k - E[Y])^2 \cdot P(Y = k)$$

$$= \sum_{k \in \{1,4,\ldots,36\}} (k - E[Y])^2 \cdot P(X = f^{-1}(k))$$

$$= (1 - E[Y])^2 P(X = \sqrt{1}) + (4 - E[Y])^2 P(X = \sqrt{4}) + \cdots$$

- For the previous example,

$$E[X] = 1 \times 1/6 + 2 \times 1/6 + \cdots + 6 \times 1/6 = 3.5$$

$$E[Y] = 1^2 \times 1/6 + 2^2 \times 1/6 + \cdots + 6^2 \times 1/6 = \$15.2$$

- Then, the variance for $Y$ is

$$var[Y] = (1^2 - 15.2)^2 \times 1/6 + (2^2 - 15.2)^2 \times 1/6 + \cdots$$

$$+ (6^2 - 15.2)^2 \times 1/6 = 149.1$$

# Example: Linear function

- If $Y = aX + b$, then $E[Y] = aE[X] + b$ and $\text{var}[Y] = a^2\text{var}[X]$

$$
\begin{aligned}
\text{var}[Y] &= \text{var}[aX + b] \\
&= \sum_k (ak + b - E[aX + b])^2 P(X = k) \\
&= \sum_k (ak + b - aE[X] - b)^2 P(X = k) \\
&= \sum_k (ak - aE[X])^2 P(X = k) \\
&= a^2 \sum_k (k - E[X])^2 P(X = k) \\
&= a^2 \text{var}[X].
\end{aligned}
$$

# Multiple Random Variables

# Multiple Random Variables

- Consider two random variables, $X$ and $Y$ associated with the same experiment.
- For $x, y \in \mathbb{R}$, we can define events of the form

$$\{X = x, Y = y\} = \{X = x\} \cap \{Y = y\}$$

- The probabilities of these events give the **joint PMF** of $X$ and $Y$:

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

- Useful for describing **multiple properties** over the outcome space of a single experiment, e.g., pick a random student and let $X$ be their height and $Y$ be their weight.

# Tabular Representation of Joint PMFs

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | $Y=1$ | $Y=2$ | $Y=3$ | $Y=4$ |
| $X=1$ | 0.1 | 0.1 | 0 | 0.2 |
| $X=2$ | 0.05 | 0.05 | 0.1 | 0 |
| $X=3$ | 0 | 0.1 | 0.2 | 0.1 |

- e.g., $P(X=2, Y=3) = ?$, $P(X=3, Y=1) = ?$, ...
- Given the joint PMF, can we compute $P(X=x)$ and $P(Y=y)$?

$$p_X(x) = P(X=x) = \sum_y P(X=x, Y=y)$$

$$p_Y(y) = P(Y=y) = \sum_x P(X=x, Y=y)$$

- If we start with the joint PMF of $X$ and $Y$, we say $p_X(x)$ is the **marginal PMF** of $X$ and $p_Y(y)$ is the **marginal PMF** of $Y$.

# Computing Marginals from the Joint Distribution

- Suppose $Y$ takes the values $y_1, y_2, \ldots, y_N$, then

$$\{Y = y_1\}, \{Y = y_2\}, \ldots, \{Y = y_N\}$$

  form partitions of $\Omega_Y$.
- Hence, $\{X = x\}$ can be partitioned into

$$\{X = x\} \cap \{Y = y_1\}, \{X = x\} \cap \{Y = y_2\}, \ldots, \{X = x\} \cap \{Y = y_N\}$$

- Therefore,

$$\begin{aligned} P(X = x) &= P(\{X = x\}) \\ &= P(\{X = x\} \cap \{Y = y_1\}) + P(\{X = x\} \cap \{Y = y_2\}) \\ &\qquad \ldots + P(\{X = x\} \cap \{Y = y_N\}) \\ &= \sum_y P(\{X = x\} \cap \{Y = y\}) = \sum_y P(X = x, Y = y) \end{aligned}$$

# Marginal PMFs

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0.2 |
| 2 | 0.05 | 0.05 | 0.1 | 0 |
| 3 | 0 | 0.1 | 0.2 | 0.1 |

| X | P(X) |
|---|---|
| 1 | 0.4 |
| 2 | 0.2 |
| 3 | 0.4 |

# Marginal PMFs

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0.2 |
| 2 | 0.05 | 0.05 | 0.1 | 0 |
| 3 | 0 | 0.1 | 0.2 | 0.1 |

| Y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| P(Y) | 0.15 | 0.25 | 0.3 | 0.3 |

# Example 1

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0 |
| 2 | 0 | 0.05 | 0.1 | 0.05 |
| 3 | 0.1 | 0.2 | 0.2 | 0.1 |

What's the value of $P(X = 2, Y = 3)$?

A: 0

B: 0.1

C: 0.05

D: 0.2

E: 1

# Example 1

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0 |
| 2 | 0 | 0.05 | 0.1 | 0.05 |
| 3 | 0.1 | 0.2 | 0.2 | 0.1 |

What's the value of $P(X = 2, Y = 3)$?

A: 0

B: 0.1

C: 0.05

D: 0.2

E: 1

Answer is B.

# Example 2

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0 |
| 2 | 0 | 0.05 | 0.1 | 0.05 |
| 3 | 0.1 | 0.2 | 0.2 | 0.1 |

What's the value of $P(X = 3)$?

A: 0.1

B: 0.4

C: 0.05

D: 0.6

E: 1

# Example 2

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0 |
| 2 | 0 | 0.05 | 0.1 | 0.05 |
| 3 | 0.1 | 0.2 | 0.2 | 0.1 |

What's the value of $P(X = 3)$?

A: 0.1

B: 0.4

C: 0.05

D: 0.6

E: 1

Answer is D.

# Conditional PMFs

- Conditional PMF of $X$ given $Y$:

$$P(X = i | Y = j) = P(\{X = i\} | \{Y = j\}) \ .$$

- Compute $P(X|Y)$ using the definition of conditional probability:

$$P(X = i | Y = j) = \frac{P(X = i, Y = j)}{P(Y = j)}$$

  since for any two events $A, B$ we have $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

- The conditional probability $P(X = i | Y = j)$ is the joint probability $P(X = i, Y = j)$ normalized by the marginal $P(Y = j)$.
- An equivalent definition of independence is $X$ and $Y$ are independent if

$$\text{for all } i, j, \quad P(X = i | Y = j) = P(X = i)$$

# Conditional PMFs

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0.2 |
| 2 | 0.05 | 0.05 | 0.1 | 0 |
| 3 | 0 | 0.1 | 0.2 | 0.1 |

| Y | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| P(Y) | 0.15 | 0.25 | 0.3 | 0.3 |

| $P(X\|Y)$ | | | | |
|---|---|---|---|---|
| X \Y | 1 | 2 | 3 | 4 |
| 1 | 0.66 | 0.4 | 0 | 0.66 |
| 2 | 0.33 | 0.2 | 0.33 | 0 |
| 3 | 0 | 0.4 | 0.66 | 0.33 |

# Functions of Two Random Variables

Given two random variables $X$ and $Y$ and a function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$Z = f(X, Y)$$

is a new random variable.

For example, pick random students and let $X$ be their height and $Y$ be their weight. If we define $Z$ as the Body Mass Index (BMI) where

$$\text{BMI} = \text{weight (lb)}/(\text{height (in)})^2 \times 703.$$

That is,

$$Z = f(X, Y) = Y/X^2 \times 703.$$

Then, $Z$ is also a random variable.

# Functions of Two Random Variables

The PMF of $Z$ can be expressed as

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

For example, let us define a new random variable $Z = X \times Y$ where the joint PMF of X and Y is

| P(X,Y) | | | | |
|---|---|---|---|---|
| X\Y | 1 | 2 | 3 | 4 |
| 1 | 0.1 | 0.1 | 0 | 0 |
| 2 | 0 | 0.05 | 0.1 | 0.05 |
| 3 | 0.1 | 0.2 | 0.2 | 0.1 |

Then, the PMF of $Z$ looks like

| $Z$ | 1 | 2 | 3 | 4 | 6 | 8 | 9 | 12 |
|---|---|---|---|---|---|---|---|---|
| $P(Z)$ | 0.1 | 0.1 | 0.1 | 0.05 | 0.3 | 0.05 | 0.2 | 0.1 |

# Expectation and Variance of Two Random Variables

- The expected value and variance of Z can be respectively computed as

$$E[Z] = \sum_z zP(Z = z) = \sum_{x,y} f(x, y)P(X = x, Y = y)$$

$$= \sum_x \sum_y f(x, y)P(X = x, Y = y)$$

$$= \sum_y \sum_x f(x, y)P(X = x, Y = y)$$

and

$$var(Z) = E[Z^2] - E[Z]^2.$$

- If $X$ and $Y$ are **independent**, for all $x, y$

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

# Linearity of Expectation

- **Lemma:** Given two random variables $X$, $Y$, and $Z = X + Y$ then
$$E[Z] = E[X + Y] = E[X] + E[Y]$$
- **Lemma:** If $X$ and $Y$ are independent then
$$E[XY] = E[X]E[Y]$$
$$var(X + Y) = var(X) + var(Y)$$

# Multiple Random Variables

- Given random variables $X_1, X_2, \ldots, X_N$ and a function $f : \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \to \mathbb{R}$,

$$Z = f(X_1, X_2, \ldots, X_N)$$

  is a new random variable.
- Linearity of Expectation: If $Z = \sum_{i=1}^{N} X_i$,

$$E[Z] = E\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E[X_i]$$

- Independence: If $X_1, \ldots, X_N$ are independent,

$$P_{X_1,\cdots,X_N}(x_i, \cdots, x_N) = \prod_{i=1}^{N} P_{X_i}(x_i) \ .$$

- Linearity of Variance: If $Z = \sum_{i=1}^{N} X_i$ and all $X_i$ are independent,

$$var[Z] = var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} var[X_i]$$