Consider rolling a standard 6-sided die.

Possible outcomes: \{1, 2, 3, 4, 5, 6\} \quad \text{sample space} = \mathcal{S}

Let \(X\) be the number shown on a roll of the die. \(X\) is a r.v. \{1, 3\} is an event

\[
\begin{align*}
\Pr\{X = 1\} &= \frac{\text{#outcomes in event}}{\text{#outcomes in sample space}} = \frac{1}{6} \\
\Pr\{X > 4\} &= \frac{2}{6} = \frac{1}{3}
\end{align*}
\]

Relative frequency interpretation of probability

- Coin tosses \(\mathcal{S} = \{\text{H, T}\}\)
- Roulette \(\mathcal{S} = \{1, 2, \ldots, 36\}\)
- Cards

What is the probability of getting all heads on three flips of a fair coin?

Let \(X_i\) be the face on the \(i\)th flip

\[
\Pr\{X_1 = \text{H}, X_2 = \text{H}, X_3 = \text{H}\} = \Pr\{\{X_1 = \text{H}\} \cap \{X_2 = \text{H}\} \cap \{X_3 = \text{H}\}\} = \frac{1}{8}
\]

Sample space \(\mathcal{S} = \{\text{HHH, HHT, HTH, HTT, THT, THH, TTH, TTT}\}\)

\[
\Pr\{\text{HHH}\} = \frac{1}{8}
\]

Note that prob's must add to 1, e.g. \(\Pr\{X_1 = \text{H}\} \cup \{X = \text{T}\}\} = \frac{3}{2} = 1\)

\(\text{H and T are mutually exclusive}\)

The probability of a collection of mutually exclusive events is the sum of their probabilities: \(\Pr\{E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_n\} = \sum_{i=1}^{n} \Pr(E_i)\)

where the \(E_i\)'s are not excl.

Def: Probability set function: \((i)\) \(\Pr\{\emptyset\} = 0\) \((ii)\) \(\Pr\{\mathcal{S}\} = 1\) \((iii)\) \(\Pr\{\mathcal{S}\} = 1\)

Properties of probabilities:

1) \(\Pr(C^c) = 1 - \Pr(C)\)

2) \(\Pr(\emptyset) = 0\)

3) If \(C_1 \subseteq C_2\), \(\Pr(C_1) \leq \Pr(C_2)\)

4) \(0 \leq \Pr(E) \leq 1\)

5) \(\Pr(C_1 \cup C_2) = \Pr(C_1) + \Pr(C_2) - \Pr(C_1 \cap C_2)\)

\[
\begin{align*}
\Pr\{X = 1\} &= \frac{1}{6} \\
\Pr\{X > 4\} &= \frac{1}{3} \\
\Pr\{X_1 = \text{H}\} \cup \{X_2 = \text{H}\} &= \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6} \\
\Pr\{X_1 = \text{H}\} \cup \{X_2 = \text{H}\} &= \frac{1}{2} + \frac{1}{2} - \frac{1}{6} = \frac{5}{6}
\end{align*}
\]
Examples: Roll three dice.
what is the prob. that the sum is 4? \[ \frac{\binom{1}{2} \cdot \binom{2}{1}}{\binom{3}{3}} = \frac{1}{12} \]
what is the prob. the sum is even? \[ \frac{\binom{6}{3} \cdot \binom{3}{0}}{\binom{9}{3}} = \frac{4}{12} = \frac{1}{3} \]
what is the prob. exactly one 2 appears? \[ \binom{1}{2} \cdot \binom{2}{1} \cdot \frac{25}{25} \cdot \frac{25}{25} = \frac{25}{25} \cdot \frac{25}{25} = \frac{25}{72} \]
what is the prob. at least one 2 appears? \[ \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \cdot \frac{1}{36} \cdot \frac{1}{216} = \frac{9}{72} \]
Note: it will be easier to compute 1 - P(no 2's) once we learn how.

Permutations

Three people run a race. Assuming no ties, in how many different orderings can they finish?

(3 choices for first), (2 choices for second), (1 for third)

3 \cdot 2 \cdot 1 = 6

n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1

The number of different arrangements of \( n \) distinguishable objects is \( n! \).

Ex: How many different arrangements of a standard deck of cards? \( 52! \approx 8.1 \times 10^{67} \)

How many different blackjack hands (one card face down, one face up) can be dealt? \( \frac{52 \cdot 51}{(52-2)!} = 2652 \)

From a department of five faculty, how many ways can they choose a chair, a DGS, and a DDS, if no person has more than one duty?

5 \cdot 4 \cdot 3 = \frac{51}{(5-3)!} = 60

Identical objects:
A car dealer has four cars to display, two of which are identical.

How many different ways can he arrange them?

If unique, \( 4! = 24 \) arrangements
But now you can't tell two of them apart, so switching those two would look the same. Thus we divide \( 4! \) by 2.

\[ \frac{4!}{2!} = 12 \]

In general \( \frac{n!}{r! \cdot (n-r)!} \)

Combinations - when order does not matter

From five faculty, how many ways can they choose 3 to be on the admissions committee?

\[ \frac{\text{# arrangements of order matters}}{\text{# arrangements of the subset}} = \frac{\binom{5}{3}}{3!} = \frac{5!}{3!} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10 \]

\[ nC_r = \binom{n}{r} = \frac{n!}{(n-r)! \cdot r!} = \text{# ways to choose } r \text{ of } n \text{ objects when order does not matter} \]

Ex: How many different poker hands (5 cards) exist?

\[ \frac{\binom{52}{5}}{4715} = \frac{52 \cdot 51 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2} = 2,598,960 \]

What is the probability a poker hand contains three 3's and two 2's?

\[ \frac{\binom{13}{3} \cdot \binom{4}{2}}{5 \cdot 5 \cdot 4 \cdot 3 \cdot 2} = 2,598,960 \approx 0.0013 \approx \frac{1}{755} \]

What is the probability of getting a flush?

\[ \frac{\binom{13}{3} \cdot \binom{4}{5}}{5 \cdot 4 \cdot 3 \cdot 2} = 2,598,960 \approx 0.00197 \approx \frac{1}{500} \]
Conditional Probability

Ex: In Bingo, there are 75 numbers which are drawn randomly without replacement until someone wins.

i) What is the probability 23 comes up first? \( P(X_1=23) = \frac{1}{75} \)

ii) If 23 comes up first, what is the probability 24 is next? \( P(X_2=24 \mid X_1=23) = \frac{1}{74} \)

Flipping a fair coin. On one toss you get heads. What is the prob.
the next toss is also heads? \( P(H_2 \mid H_1) = P(H) = \frac{1}{2} \)

Def: \( P(E_2 \mid E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)} \quad \text{assuming } P(E_1) > 0 \)

Ex: \( P(H_2 \mid H_1) = \frac{P(H_2 \cap H_1)}{P(H_1)} = \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{1}{2} \)

In a particular collection of 30 graduate students, 15 are international students, 9 are female, and 6 are computer engineers. There are 3 international female students, of which one is CE. There are no domestic female CE students, and there is only one male
int'l CE student.

i) If a student is male, what is the prob they are CE? \( P(CE \mid F) = \frac{P(CE \cap F)}{P(F)} = \frac{\frac{1}{10}}{\frac{9}{10}} = \frac{1}{9} \)

ii) What is the probability a female student is CE? \( P(CE \mid F) = \frac{P(CE \cap F)}{P(F)} = \frac{\frac{3}{10}}{\frac{9}{10}} = \frac{1}{3} \)

iii) Among int'l students, what is the prob a female student is CE? \( P(CE \mid F, I) = \frac{P(CE \cap F \cap I)}{P(F \cap I)} = \frac{\frac{3}{30}}{\frac{5}{30}} = \frac{3}{5} \)

Properties: \( P(E_1E) = 1 \)
\( P(E_1E) = 0 \)
\( P(E_1 \cap E_2) = P(E_1 \cap E_2) P(E_1) \quad P(E_1 \cap E_2 \cap E_3) = P(E_2 \cap E_1 \cap E_3) P(E_1 \cap E_2) P(E_3) \)

Ex: In Bingo, what is the prob. 23 is drawn first and 24 second?
\( P(X_1=23 \cap X_2=24) = P(X_2=24 \mid X_1=23) P(X_1=23) = \frac{1}{74} \cdot \frac{1}{75} \)

What is the prob. a randomly selected grad student from the above group is a male int'l student?
\( P(M \mid I) = P(M \mid I) P(I) = \frac{1}{15} \cdot \frac{15}{30} = \frac{15}{450} = \frac{1}{30} \)

Law of total probability: If \( C_1, C_2, ..., C_k \) are a partition of \( S \), then \( P(E) = \sum_{i=1}^{k} P(E \mid C_i) P(C_i) \)

\[ C_i \cap C_j = \emptyset \quad \forall i \neq j \]
\[ \bigcup_{i=1}^{k} C_i = S \]

If time, do independence on page 15.
Bayes' Theorem

\[ P(C \mid E) = \frac{P(E \mid C)P(C)}{\sum_{C_i} P(E \mid C_i)P(C_i)} = \frac{P(C \mid E)P(C)}{\sum_{C_i} P(C_i \mid E)P(C_i)} = \frac{P(C, E)}{P(E)} \]

Ex: A jar contains 10 coins, half of which are fair and half that are weighted to show heads 70% of the time. If you draw a random coin, flip it and get heads, what is the probability it is a loaded coin?

\[ P(\text{loaded} \mid \text{H}) = \frac{P(\text{H} \mid \text{loaded})P(\text{loaded})}{P(\text{H} \mid \text{loaded})P(\text{loaded}) + P(\text{H} \mid \text{fair})P(\text{fair})} = \frac{\frac{7}{10} \times \frac{1}{2}}{\frac{7}{10} \times \frac{1}{2} + \frac{3}{10} \times \frac{1}{2}} = \frac{7}{12} \]

Mind-bending examples:

An early test for HIV antibodies was the ELISA test. The probability a person with HIV would test positive was 97.7%, and the probability of testing negative for a person without HIV was 92.6%. Approximately 0.26% of the population of North America has HIV.

Suppose a randomly selected person tests positive. What is the probability they have HIV?

\[ P(\text{HIV} \mid +) = \frac{P(+) \mid \text{HIV}P(\text{HIV})}{P(+) \mid \text{HIV}P(\text{HIV}) + P(+) \mid \text{HIV} \text{HIV})P(\text{HIV})} = \frac{(0.977)(0.0026)}{(0.977)(0.0026) + (0.074)(0.9974)} = 0.033 \]

A jar contains two coins—one fair, one with heads on both sides. You draw a random coin and look at one side and see that it is a head. What is the probability it is the 2-headed coin?

\[ P(\text{2H} \mid +) = \frac{P(\text{2H} \mid +)P(\text{2H})}{P(\text{H} \mid +)P(\text{2H}) + P(\text{1H} \mid +)P(\text{1H})} = \frac{\frac{1}{2} \frac{1}{4}}{\frac{1}{2} \frac{1}{4} + \frac{1}{2} \frac{3}{4}} = \frac{1}{3} \]

Monte Hall: There are three closed doors, A, B, C, behind one of which is a new car. You get to pick one door. The host then will open a door that does not have the car, and then offers you the chance to switch. Should you?

Suppose you pick A, and he opens B. Should you switch to C?

\[ P(\text{car in A \mid open B}) = \frac{P(\text{open B} \mid A)P(A)}{P(\text{open B} \mid A)P(A) + P(\text{open B} \mid B)P(B) + P(\text{open B} \mid C)P(C)} = \frac{\frac{1}{3} \frac{1}{3}}{\frac{1}{3} \frac{1}{3} + \frac{1}{3} \frac{1}{3} + \frac{1}{3} \frac{1}{3}} = \frac{1}{3} \]

\[ P(\text{C \mid open B}) = \frac{1 - \frac{1}{3}}{\frac{1}{3} \frac{1}{3} + \frac{1}{3} \frac{1}{3} + \frac{1}{3} \frac{1}{3}} = \frac{2}{3} \]
**Independence**

**Def:** $C_1$ and $C_2$ are independent if $P(C_2|C_1) = P(C_2)$

Then $P(C_1, C_2) = P(C_1)P(C_2)$.

By definition, if $P(C_i) = 1$ or $P(C_i) = 0$, $C_i$ is independent of all other events.

**Ex:** Roll three dice, what is the prob. at least one 2 appears?

$= 1 - P(\text{no } 2's) = 1 - \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{91}{216}$

*Note: the difference between mutually exclusive and independent.
Mutually exclusive events are dependent. $P(\text{hit}) = 0$*

---

**Random Variables**

**Def:** A **random variable** is a function of the sample space.

Perhaps the identity function

Book wants it to map to $\mathbb{R}$.

Roll one die $\Rightarrow$ identity

Roll two die $\Rightarrow$ sum

**Def:** A discrete RV is one that has a countably many possible values.

*Ex:* # heads in 5 flips

# flips until first head

Denote the range of RV $X$ by $\mathcal{X}$.

We typically describe $X$ by its **probability density function** (pdf)

(or pmf), a probability set function $f$ with domain $\mathcal{X}$.

*Note: $f(x) > 0 \forall x \in \mathcal{X}$*

**Ex:** Rolling a die. $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$. $f(x) = \frac{1}{6} \mathbb{I}_{\{x \in \mathcal{X}\}}$

Flipping three coins, $X = \# \text{heads}$. $\mathcal{X} = \{0, 1, 2, 3\}$. $f(x) = \begin{cases} \frac{1}{6} & \text{if } x = 0 \\ \frac{3}{8} & x = 1 \\ \frac{1}{2} & x = 2 \\ \frac{3}{8} & x = 3 \\ 0 & \text{otherwise} \end{cases}$

$= \left(\begin{array}{c} 3 \\ \frac{1}{6} \end{array}\right) = (3)\left(\frac{1}{6}\right)^3$

*Notational note: capital roman letter for RV, lower case for values.*

**Def:** The (cumulative) distribution function (cdf) $F(x) = P(X \leq x)$.

*Ex:* # heads on three flips

$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8} & 0 \leq x < 1 \\ \frac{1}{2} & 1 \leq x < 2 \\ \frac{7}{8} & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$

*Note: $0 \leq F(x) \leq 1 \forall x \in \mathbb{R}$

$F(x)$ is non-decreasing

For discrete RVs, $F(x)$ is a step function and is right-continuous:

$P(a < x \leq b) = F(b) - F(a)$
A continuous random variable takes values on a continuous interval (or union of intervals).

Let the domain of $X$, a continuous rv, be $\Omega$. If $A \subset \Omega$,

$$P(A) = \Pr(X \in A) = \int_A f(x) \, dx$$

where $f$ is the pdf or density.

Ex: Standard spinner. Let $X$ be the angle in radians of the result of a spin.

Then $0 \leq X < 2\pi$ with all values equally likely.

Since all values equally likely, $f$ is constant w.r.t. $X$.

To be a prob. it must integrate to 1, so $f(x) = \frac{1}{2\pi} I_{[0,2\pi]}$

What is $P(X < \pi)^2$?

$$\int_0^{\pi} \frac{1}{2\pi} \, dx = \frac{1}{\pi}$$

$P(X \leq \pi) = \frac{1}{2}$

$P(X = \pi) = 0$

Problem 1.64:

Ex: Let $f(x) = \frac{x^2}{18}$ on $(-3,3)$. What is $P(\mid X \mid < 1)$?

$$P(\mid X \mid < 1) = \int_{-1}^{1} f(x) \, dx = \int_{-1}^{1} \frac{x^2}{18} \, dx = \frac{1}{54} x^3 \bigg|_{-1}^{1} = \frac{1}{27}$$

Let $f(x) = \frac{x^2}{18}$. Then $P(X^2 < 9) = \int_{-3}^{3} f(x) \, dx = \int_{-3}^{3} \frac{x^2}{18} \, dx = \frac{1}{36} (x^3) \bigg|_{-3}^{3} = \frac{25}{36}$

CDF $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(w) \, dw$

$f(x) = F'(x)$ where $f$ is continuous.

(Note: $f$ is always the Radon-Nikodym derivative of $F$ wrt an appropriate dominatinig measure.)

Ex: $P(x) = \frac{x^2}{18} I_{[-3,3]}$

$$F(x) = \int_{-\infty}^{x} \frac{w^2}{18} I_{[-3,3]} \, dw = \frac{1}{36} \frac{(w^2)^3}{2} \bigg|_{-\infty}^{x} = \begin{cases} 0 & \text{if } x \leq -2 \\ \frac{1}{36} (x^2)^3 & \text{if } -2 < x < 4 \\ 1 & \text{if } x \geq 4 \end{cases}$$

What is the cdf for $Y = X^2$?

$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$$P(Y < y) = \begin{cases} \frac{1}{36} (\sqrt{y} + 3)^2 \leq \frac{25}{36} & \text{if } y \leq 0 \\ \frac{(\sqrt{y} + 3)^3}{36} & \text{if } 0 < y < 4 \\ \frac{1}{36} (\sqrt{y} + 3) \leq \frac{25}{36} & \text{if } 4 \leq y < 16 \\ 1 & \text{if } y \geq 16 \end{cases}$$
Expectation

\[ \mathbb{E}[X] = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f(x) \, dx & \text{if } X \text{ is continuous} \end{cases} \]

is the expected value or mean of \( X \).

If \( \mathbb{E}[X] = \infty \), we might say that it does not exist.

Ex: You flip a coin twice, \( X \) = # heads.

\[ \mathbb{E}[X] = (0)(\frac{1}{4}) + (1)(\frac{1}{4}) + (2)(\frac{1}{4}) = 1 \]

Ex: \( f(x) = \frac{1}{2} x \mathbb{I}[0 \leq x \leq 2] \)

\[ \mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{2} x \mathbb{I}[0 \leq x \leq 2] \, dx = \int_{0}^{2} \frac{1}{2} x^2 \, dx = \frac{1}{6} x^3 \bigg|_0^2 = \frac{4}{3} \]

Note: \( \mathbb{E}[aX+b] = a \mathbb{E}[X] + b \)

Law of the unconscious statistician: \( \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx \) or \( \sum_x g(x)f(x) \)

Moments

Def: The \( k^{th} \) moment of \( X \) is \( \mathbb{E}[X^k] \). \( \mathbb{E}[X], \mathbb{E}[X^2], \ldots \)

\( \mathbb{E}[X] \) is the mean.

\( \mathbb{E}[(X-\mathbb{E}[X])^2] \) is the variance.

skew, kurtosis,…

Denote \( \mathbb{E}[(X-\mathbb{E}[X])^2] = \sigma^2 \). \( \sigma \) is the standard deviation.

\( \sim 95\% \) within \( 2 \sigma \) \( \Rightarrow \) range of error \( 6 \sigma \) 

\[ \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - 2 \mathbb{E}[X] \mathbb{E}[X] + (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \]

Integral/Summation is linear

Moment Generating Function

Def: The MGF of \( X \) is \( M(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) \, dx \). Note it is a fn of \( t \). It may not exist.

We call it an MGF because the \( k^{th} \) moment is \( \left. \frac{d^{k} M(t)}{dt^{k}} \right|_{t=0} \)

Ex: \( f(x) = \frac{1}{2} x \mathbb{I}[0 \leq x \leq 2] \)

\[ M(t) = \mathbb{E}[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{2} x \mathbb{I}[0 \leq x \leq 2] \, dx = \int_{0}^{2} \frac{1}{2} x e^{tx} \, dx = \frac{x}{2} e^{tx} \bigg|_0^2 - \int_{0}^{2} \frac{1}{2} e^{tx} \, dx \\
= \frac{1}{2} e^{2t} - \frac{1}{2} e^{2t} \bigg|_0^2 = \frac{1}{2} e^{2t} - \frac{1}{2} e^{2t} + \frac{1}{2} e^{2t} = (\frac{1}{2} - \frac{1}{2}) e^{2t} + \frac{1}{2} e^{2t} \\
E[X] = M'(t) \bigg|_{t=0} = \frac{d}{dt} \left[ (\frac{1}{2} - \frac{1}{2}) e^{2t} + \frac{1}{2} e^{2t} \right] \bigg|_{t=0} = \left[ (\frac{1}{2} - \frac{1}{2}) e^{2t} + (\frac{1}{2} - \frac{1}{2}) e^{2t} \frac{2}{3} - \frac{1}{3} \right] \bigg|_{t=0} = \frac{(2t^2 - 2t + 1)e^{2t}}{3t^3} \bigg|_{t=0} = \frac{4}{3} e^{2t} \bigg|_{t=0} = \frac{4}{3} \]
Notes: MGF’s don’t always exist.
If they exist, they are unique, and completely determine the associated dist'n.
So if you can prove two RV’s have the same MGF’s, then they have the same dist'n.
Note that two RV’s could have the same moments, but have different MGF’s and different distributions.
Since many MGF’s don’t exist, we sometimes instead use the characteristic function \( \phi(t) = E[e^{itx}] \) which always exists.
Note: \( e^{itx} = \cos(tx) + i\sin(tx) \)

\[
E[x^k] = \frac{d^k}{dt^k} \phi(t) \bigg|_{t=0}
\]

**Chebyshev’s Inequality**

Let \( X \) be a RV with mean \( \mu \) and variance \( \sigma^2 < \infty \). Then for all \( k > 0 \),

\[
P( |X - \mu| > k\sigma) \leq \frac{1}{k^2}
\]