Indian Statistical Institute

# PROBABILITY - I

*Notes by Utkarsh Bharadwaj*

*Fall 2025*

**Prof. Abhay Gopal Bhatt**

Indian Statistical Institute, Delhi

# Contents

# 1 Foundations of Probability

## 1.1 Basic Concepts

**Random Experiment & Sample Space**

- **Random Experiment:** A procedure in which repeated trials may result in different outcomes.
- **Sample Space ($\Omega$):** The collection of all possible outcomes.

**Examples of Sample Spaces**

- **Coin Toss:** $\Omega = \{H, T\}$
- **Die Toss:** $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Travel Time:** If measuring time from residence to class, $\Omega = [0, 15]$ (Continuous interval).

**Events**

An **event** $A$ is any subset of the sample space ($A \subseteq \Omega$).

- **Mutually Exclusive:** Events $A$ and $B$ are disjoint if $A \cap B = \emptyset$.
- **Partition:** $A$ and $A^c$ form a partition of $\Omega$, meaning $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$.

## 1.2 Axioms of Probability (Kolmogorov)

A probability measure $P$ is a function satisfying three axioms:

**The Three Axioms**

1. **Non-negativity:** $P(A) \geq 0$ for any event $A$.
2. **Normalization:** $P(\Omega) = 1$.
3. **Countable Additivity:** If $A_1, A_2, \ldots$ are pairwise mutually exclusive, then:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

# 2 Derived Properties

From the axioms, we derive several standard properties.

**Basic Properties**

1. **Empty Set:** $P(\emptyset) = 0$.

2. **Finite Additivity:** For mutually exclusive $A_1, \dots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

3. **Complement Rule:** $P(A^c) = 1 - P(A)$.
4. **Bounds:** $0 \le P(A) \le 1$.
5. **Inclusion-Exclusion (2 Events):**

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

## 2.1 Continuity of Probability

Probability measures behave continuously with respect to limits of sets.

**Continuity from Below**

If $A_1 \subseteq A_2 \subseteq A_3 \dots$ is an increasing sequence of events, and $A = \bigcup_{i=1}^{\infty} A_i$, then:

$$P(A) = \lim_{n\to\infty} P(A_n)$$

*Proof.* Define disjoint sets $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, ..., $B_n = A_n \setminus A_{n-1}$. Then $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ and $A_n = \bigcup_{i=1}^{n} B_i$. Using countable additivity:

$$P(A) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n)$$

□

**Continuity from Above**

If $A_1 \supseteq A_2 \supseteq A_3 \dots$ is a decreasing sequence, and $A = \bigcap_{i=1}^{\infty} A_i$, then:

$$P(A) = \lim_{n\to\infty} P(A_n)$$

# 3 Counting Techniques

If the sample space $\Omega$ is finite and all outcomes are **Equally Likely**:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}}$$

**Combinatorics Formulas**

- **Permutations (Order matters):** ${}^nP_k = \frac{n!}{(n-k)!}$
- **Combinations (Order doesn't matter):** $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

**Cricket Team Selection**

A squad has 15 players: 7 Batsmen, 6 Bowlers, 2 All-rounders. We need a team of 5 Batsmen, 5 Bowlers, and 1 All-rounder.

$$\text{Ways} = \binom{7}{5} \times \binom{6}{5} \times \binom{2}{1} = 21 \times 6 \times 2 = 252$$

## 4 Conditional Probability

**Definition**

For events $A$ and $B$ with $P(B) > 0$:

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

**Multiplication Rule**

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

**Defective Components (Total Probability)**

A factory has two assembly lines.

- Line 1: 10 components (2 defective). $P(L_1) = 0.4, P(D|L_1) = 0.2$.
- Line 2: 15 components (1 defective). $P(L_2) = 0.6, P(D|L_2) = 1/15$.

Probability a randomly chosen component is defective ($D$):

$$\begin{aligned} P(D) &= P(D|L_1)P(L_1) + P(D|L_2)P(L_2) \\ &= (0.2)(0.4) + \left(\frac{1}{15}\right)(0.6) \\ &= 0.08 + 0.04 = 0.12 \end{aligned}$$

## 5 Bayes' Theorem

Let $A_1, \ldots, A_n$ be a partition of $\Omega$ (mutually exclusive and exhaustive). For any event $B$:

**Bayes' Formula**

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

**Rare Disease Testing**

- Incidence of Disease ($D$): $P(D) = 0.001$ (Rare).
- Test Accuracy (True Positive): $P(+|D) = 0.99$.
- False Positive Rate: $P(+|D^c) = 0.02$.

If a person tests positive, what is the probability they actually have the disease?

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} \\ &= \frac{0.99(0.001)}{0.99(0.001) + 0.02(0.999)} \\ &= \frac{0.00099}{0.00099 + 0.01998} \approx 0.047 \end{aligned}$$

**Result:** Even with a 99% accurate test, the probability is only **4.7%** because the disease is so rare.

# 6 Independence

Events $A$ and $B$ are independent if and only if:

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

This is equivalent to $P(A|B) = P(A)$.

**Die Toss Independence**

Consider a fair die. $\Omega = \{1..6\}$.

- $A = \{1, 2, 3\}$ (Low), $P(A) = 1/2$.
- $B = \{2, 4, 6\}$ (Even), $P(B) = 1/2$.
- $C = \{2, 3, 4, 5\}$, $P(C) = 2/3$.

**Check A and B:** $A \cap B = \{2\}$, so $P(A \cap B) = 1/6$. $P(A)P(B) = 1/4$. $1/6 \neq 1/4$, so **Not Independent**.

**Check B and C:** $B \cap C = \{2, 4\}$, so $P(B \cap C) = 2/6 = 1/3$. $P(B)P(C) = (1/2)(2/3) = 1/3$. Matches! So $B$ and $C$ are **Independent**.

**Mutual Independence**

Let $A_1, A_2, \ldots, A_n$ be $n$ events. These events are mutually independent if for any sub-collection $2 \leq k \leq n$:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$$

**Key Point**

Any set of $K$ events must be mutually independent in themselves. The total number of cases to check are:

$$\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = \sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$$

# 7 Bernoulli Trials

**Bernoulli Trial**

An experiment where there are only two outcomes. One results in **Success** $(S)$ and one results in **Failure** $(F)$.

## Examples

1. Consider any experiment with only two possible outcomes: $S$ (Success) and $F$ (Failure), where $P(S) = p$ for $0 < p < 1$. Consider repeated trials (independent events). The trial continues until the $1^{st}$ $S$.

$$P\Big(\underbrace{F \cdots F}_{n} S\Big) = (1-p)^n p$$

2. Let $A_i =$ Component $i$ is working.

$$P(A_i) = p \quad \text{and} \quad P(A_i^c) = 1 - p \quad \forall\, i = 1, 2, 3$$

The system **fails** if: $A_1^c \cap (A_2^c \cup A_3^c)$

The system **works** (Event $W$) if: $A_1 \cup (A_2 \cap A_3)$

$$P(W) = P(A_1) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3) = p + p^2 - p^3$$

3. **Random Variables on Bernoulli Trials:** $\Omega = \{S, F\}$. Define $X : \Omega \to \mathbb{R}$. Any such random variable is called a **Bernoulli random variable with parameter** $p$,

denoted $X \sim \text{Ber}(p)$.

$$X(\omega) = \begin{cases} 0 & \text{if } \omega \in F \\ 1 & \text{if } \omega \in S \end{cases}$$

Note: $S = \{\omega : X(\omega) = 1\}$ and $P(S) = P(\omega : X(\omega) = 1) = p$.

4. Let $X \sim \text{Ber}(p)$. Define $Y = 1 - X$.

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega \in F \\ 0 & \text{if } \omega \in S \end{cases}$$

Then $Y \sim \text{Ber}(1 - p)$.

# 8 Probability Mass Function (PMF)

**PMF Definition**

Define $p : \mathbb{R} \to [0, 1]$ such that:

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

$p$ is called the **Probability Mass Function** (pmf) of $X$.

**Note**: $p(x) = P(X = x)$.

### Inverse Images

Let $X : \Omega \to \mathbb{R}$. Define for any $B \in \mathbb{R}$:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

For the distribution $Q$:

$$Q(B) = P(X^{-1}(B))$$

# 9 Common Distributions

## 9.1 Binomial Distribution

We perform an experiment with $n$ repeated Bernoulli trials. Total outcomes: $2^n$. Let $X =$ number of $S$'s in $n$ independent repeated Bernoulli trials with $P(S) = p$. $X$ takes values $0, 1, 2, \ldots, n$.

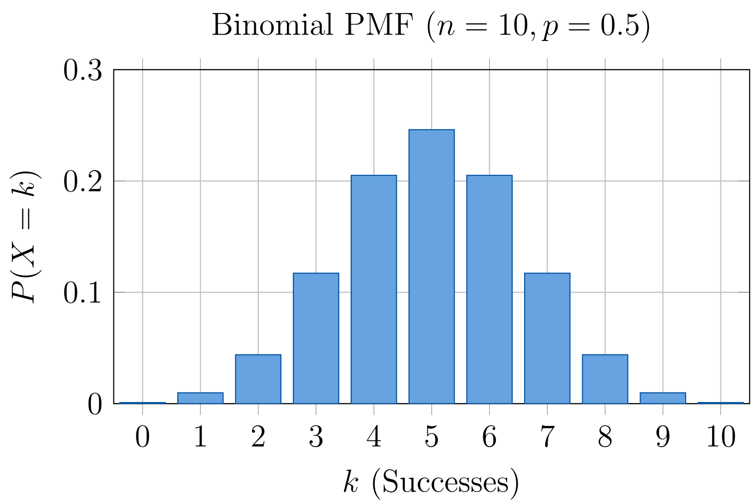



Figure 1: Probability Mass Function of a Binomial Distribution

**Binomial PMF**

$$p(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

## 9.2 Geometric Distribution

Repeat Bernoulli trials until you get the $1^{st}$ Success ($S$). Let $X$ = number of failures **before** the $1^{st}$ $S$.

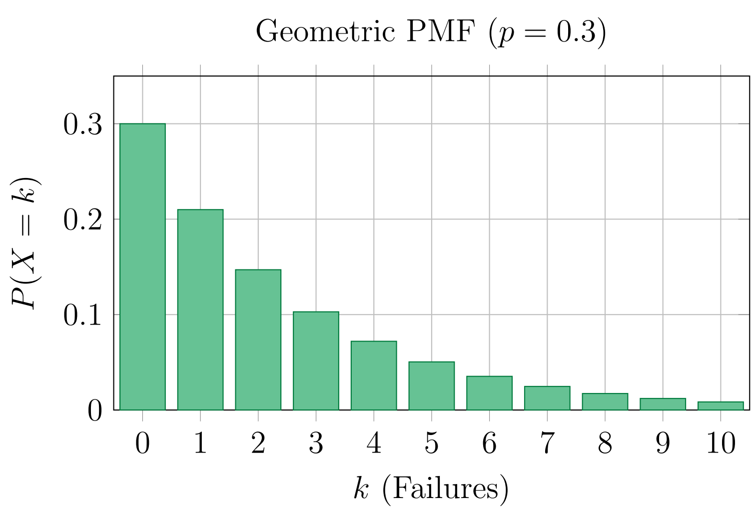



Figure 2: Probability Mass Function of a Geometric Distribution

**Geometric PMF (Failures)**

$$p(k) = P(X = k) = \begin{cases} p(1-p)^k & \text{for } k = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

*Variation:* Let $Y$ = number of trials **until** the $1^{st}$ $S$. Thus $Y = X + 1$.

$$p_Y(k) = \begin{cases} p(1-p)^{k-1} & \text{for } k = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

### 9.3 Negative Binomial Distribution

Repeat Bernoulli trials independently until you get $r$ successes. Let $X$ = number of failures before the $r^{th}$ success.

$$X \sim \text{NB}(r, p)$$

**Negative Binomial PMF**

$$P(X = k) = \begin{cases} \binom{k+r-1}{r-1} p^r (1-p)^k & \text{for } k = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

### 9.4 Hypergeometric Distribution

Consider an urn with $r$ red balls and $g$ green balls. Draw one ball, record color, do **not** replace it. Repeat. This is a repeated trial *without replacement*.

## 10 Poisson Approximation of Binomial

**Theorem**

Let $\lambda > 0$, $n \geq \lambda$, and $p = \frac{\lambda}{n}$. Let

$$A_k = \{\text{k successes in independent Ber(p) trials}\}$$

Then,

$$\lim_{n \to \infty} P(A_k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

*Proof.*

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$
\begin{aligned}
&= \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \left[\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}\right] \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \left[1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right] \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k}
\end{aligned}
$$

Taking the limit as $n \to \infty$:

- $\lim_{n\to\infty}(1 - \frac{r}{n}) = 1$
- $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^{-k} = 1$
- $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^{n} = e^{-\lambda}$

$$\therefore P(A_k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

□

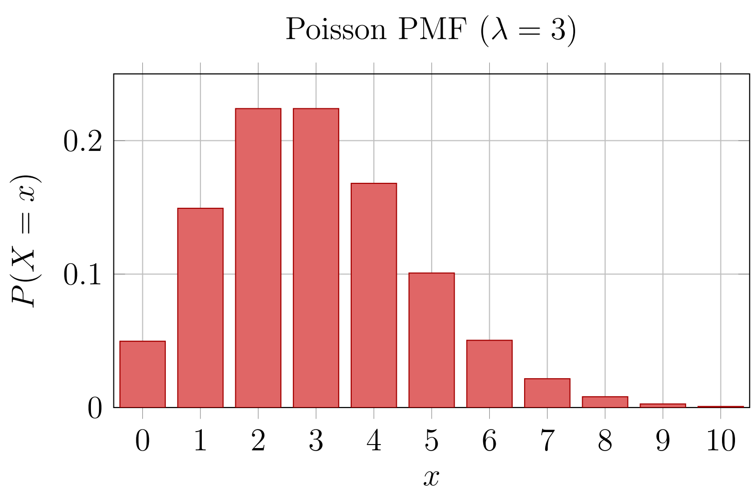



Figure 3: Probability Mass Function of a Poisson Distribution

**Poisson Distribution**

A random variable $Y$ is called a Poisson Random Variable, $Y \sim \text{Poi}(\lambda)$, if:

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

# 11 Expected Values of Discrete R.V.

Let $X$ be a discrete random variable with pmf $p(x)$.

$$E(X) = \sum_x xp(x) = \sum_x xP(X = x)$$

## 11.1 Examples

**1. Discrete Uniform**

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p(x)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

$$E(X) = \sum xp(x) = 3.5$$

Generally, for $Y$ uniform on $\{1, \ldots, N\}$: $E[Y] = \frac{N+1}{2}$.

**2. Arbitrary Distribution**

| $x$ | -2 | -1 | 0 | 2 | 4 |
|---|---|---|---|---|---|
| $p(x)$ | $\frac{1}{7}$ | $\frac{2}{7}$ | $\frac{1}{7}$ | $\frac{2}{7}$ | $\frac{1}{7}$ |

$$E[X] = (-2)\frac{1}{7} + (-1)\frac{2}{7} + (0)\frac{1}{7} + (2)\frac{2}{7} + (4)\frac{1}{7} = \frac{4}{7}$$

## 11.2 Summary of Expectations

1. **Bernoulli:** $X \sim \text{Ber}(p) \implies E(X) = p$
2. **Binomial:** $X \sim \text{Bin}(n, p) \implies E(X) = np$
3. **Poisson:** $X \sim \text{Poi}(\lambda) \implies E(X) = \lambda$

**Expectation of a Function**

$$E[h(X)] = \sum_x h(x)p(x)$$

**Linearity of Expectation**

Let $a, b \in \mathbb{R}$. Then:

$$E[aX + b] = aE(X) + b$$

**Exercise:** Let $f(x) = \sum_{n=0}^{N} a_n x^n$. Show that $E[f(X)] = \sum_{n=0}^{N} a_n E(X^n)$.
This result follows from the **Linearity of Expectation**. The expectation operator $E[\cdot]$ satisfies two key properties for any random variables and constants:

- **Additivity:** $E[X+Y] = E[X] + E[Y]$
- **Homogeneity:** $E[cX] = cE[X]$ for any constant $c$.

We proceed with the derivation:

$$
\begin{aligned}
E[f(X)] &= E\left[\sum_{n=0}^{N} a_n X^n\right] \\
&= \sum_{n=0}^{N} E\left[a_n X^n\right] && \text{(Using Additivity)} \\
&= \sum_{n=0}^{N} a_n E\left[X^n\right] && \text{(Using Homogeneity)}
\end{aligned}
$$

Thus, we have shown that:

$$E[f(X)] = \sum_{n=0}^{N} a_n E[X^n]$$

## 12 Existence of Expectation

Not all random variables have a finite expected value.

**Divergent Expectation (St. Petersburg Paradox)**

Suppose $Z$ is a random variable with pmf:

$$p(z) = \begin{cases} \frac{1}{2^n} & \text{if } z = 2^n \quad (n = 1, 2, \dots) \\ 0 & \text{otherwise} \end{cases}$$

Calculating the expectation:

$$
\begin{aligned}
E(Z) &= \sum_{z} z p(z) \\
&= \sum_{n=1}^{\infty} 2^n p(2^n) \\
&= \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} \\
&= \sum_{n=1}^{\infty} 1 = \infty
\end{aligned}
$$

**Conclusion:** The expectation of $Z$ does not exist (it diverges).

**Undefined Expectation**

Let $T$ take values $(-1)^n 2^n$ for $n = 1, 2, \ldots$ with probability $\frac{1}{2^n}$ respectively.

$$p_T(t) = \begin{cases} \frac{1}{2^n} & \text{if } t = (-1)^n 2^n \\ 0 & \text{otherwise} \end{cases}$$

$$E(T) = \sum_{n=1}^{\infty} (-1)^n 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} (-1)^n = -1 + 1 - 1 + 1 \ldots$$

This series does not converge. Hence, the expectation value of $T$ does not exist.

**Remarks on Existence**

1. If a random variable takes only **finitely** many values, then the expectation will always exist.
2. If $X$ takes **infinitely** many values but is bounded ($0 \leq X \leq k$), then:

$$E(X) = \sum_x xp(x) \leq \sum_x kp(x) = k \sum p(x) = k$$

   The expectation exists.
3. If $X$ takes arbitrarily large values with positive probability, and the distribution has "heavy tails," the expectation may not exist.

# 13 Application: Expected Profit

**Store Inventory Problem**

A store purchases three items at Rs. 500 each. They are sold for Rs. 1000 per piece. Unsold items are bought back by the company for Rs. 200/piece.
Let $X$ = number of pieces sold. The probabilities are:

$$p(0) = 0.1, \quad p(1) = 0.2, \quad p(2) = 0.3, \quad p(3) = 0.4$$

**Solution:** Let $Y$ denote the actual profit. The cost is fixed at $3 \times 500 = 1500$. Revenue comes from sales ($1000X$) and buyback ($200 \times (3 - X)$).
Alternatively, formulated directly as profit:

$$Y = 500X - 300(3 - X)$$

$$Y = 800X - 900$$

To find expected profit $E(Y)$:

$$E(Y) = 800E(X) - 900$$

First, find $E(X)$:

$$E(X) = \sum_{x=0}^{3} xp(x) = 0(0.1) + 1(0.2) + 2(0.3) + 3(0.4) = 2.0$$

Therefore:

$$E(Y) = 800(2) - 900 = 1600 - 900 = 700$$

The expected profit is Rs. 700.

# 14 Variance and Standard Deviation

**Variance Definition**

Let $X$ be a discrete random variable with mean $\mu = E(X)$ and pmf $p(x)$. The **Variance of X** is:

$$V(X) = E[(X - \mu)^2]$$

The **Standard Deviation** is $\sigma_X = \sqrt{V(X)}$.

Visualizing Variance (Spread)

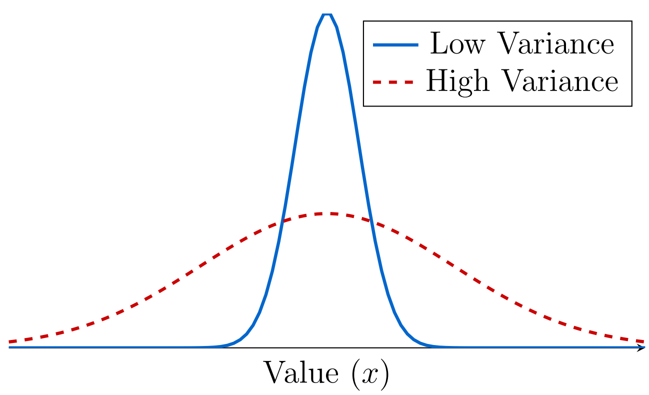



Figure 4: Variance measures how spread out the data is around the mean.

## 14.1 Calculation Example

**Calculating Variance**

Consider the following distribution:

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $p(x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

1. Calculate Mean ($E(X)$):

$$E(X) = 1(0.4) + 2(0.3) + 3(0.2) + 4(0.1) = 2$$

2. Calculate Variance ($V(X) = E[(X-\mu)^2]$):

$$\begin{aligned} V(X) &= (1-2)^2(0.4) + (2-2)^2(0.3) + (3-2)^2(0.2) + (4-2)^2(0.1) \\ &= (-1)^2(0.4) + 0 + (1)^2(0.2) + (2)^2(0.1) \\ &= 0.4 + 0 + 0.2 + 0.4 = 1 \end{aligned}$$

**Properties of Variance**

1. $E(X-\mu) = E(X) - \mu = 0$ (The mean of deviations is zero).
2. $E|X-\mu|$ is the Mean Absolute Deviation.
3. **Shortcut Formula:**

$$\begin{aligned} V(X) = E[(X-\mu)^2] &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

4. **Linear Transformation:**

$$V(aX+b) = a^2 V(X)$$

*Proof:*

$$\begin{aligned} V(aX+b) &= E[((aX+b) - E(aX+b))^2] \\ &= E[(aX + b - a\mu - b)^2] \\ &= E[(a(X-\mu))^2] \\ &= a^2 E[(X-\mu)^2] = a^2 V(X) \end{aligned}$$

# 15 Variance of Common Distributions

## 15.1 Bernoulli Distribution

Let $X \sim \text{Ber}(p)$. We know $E(X) = p$.

$$E(X^2) = 0^2 \cdot p(0) + 1^2 \cdot p(1) = 0 + p = p$$

$$V(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$$

## 15.2 Binomial Distribution

Let $X \sim \text{Bin}(n, p)$. We use the technique of factorial moments: $E[X(X-1)]$.

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1)\frac{n(n-1)(n-2)!}{x(x-1)(x-2)![(n-2)-(x-2)]!} p^2 p^{x-2}(1-p)^{(n-2)-(x-2)} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)![(n-2)-(x-2)]!} p^{x-2}(1-p)^{(n-2)-(x-2)}
\end{aligned}$$

The summation part is the sum of probabilities for a Binomial distribution with parameters $(n-2, p)$, which sums to 1.

$$E[X(X-1)] = n(n-1)p^2$$

Now, we relate this to Variance:

$$\begin{aligned}
E[X(X-1)] = E(X^2 - X) &= E(X^2) - E(X) \\
\implies E(X^2) &= n(n-1)p^2 + np \\
\implies E(X^2) &= n^2p^2 - np^2 + np
\end{aligned}$$

Finally, using $V(X) = E(X^2) - [E(X)]^2$:

$$\begin{aligned}
V(X) &= (n^2p^2 - np^2 + np) - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p)
\end{aligned}$$

# 16 Moments

$$E(X) = \sum_x xp(x)$$

$$V(X) = E((X-\mu)^2) = E(X^2) - (E(X))^2$$

**Existence**

$E(X)$ exists if and only if $E[|X|] < \infty$.

$$E[|X|] = \sum_x |x|p(x) < \infty$$

- **Raw Moment:** The $k^{th}$ raw moment of $X$ is $E[X^k]$.
- **Central Moment:** The $k^{th}$ central moment of $X$ is $E[(X-\mu_x)^k]$.

**Note:** If the $k^{th}$ central moment exists, then the $k^{th}$ raw moment exists.

# 17 Moment Generating Functions (MGF)

**Moment Generating Function**

The MGF of a random variable $X$ is defined as:

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} p(x)$$

**Property:** $M_X(0) = E[e^0] = 1$.

## 17.1 Examples

**1. Poisson Distribution**

Let $X \sim \text{Poi}(\lambda)$.

$$\begin{aligned}
M_X(t) = E[e^{tX}] &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
&= e^{-\lambda} \cdot e^{\lambda e^t} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}$$

**2. Binomial Distribution**

Let $Y \sim \text{Bin}(n, p)$.

$$\begin{aligned}
M_Y(t) &= \sum_{y=0}^{n} e^{ty} \binom{n}{y} p^y (1-p)^{n-y} \\
&= \sum_{y=0}^{n} \binom{n}{y} (pe^t)^y (1-p)^{n-y} \\
&= (pe^t + 1 - p)^n
\end{aligned}$$

## 17.2 Generating Moments from MGF

**Derivatives of MGF**

If $M_X(t)$ exists for $t \in (-\delta, \delta)$, then the $k^{th}$ moment of $X$ is given by the $k^{th}$ derivative of the MGF at $t = 0$:

$$E[X^k] = M_X^{(k)}(0)$$

**Verification using Binomial Distribution:**

$$M_Y(t) = (1 - p + pe^t)^n$$

1. First Derivative:

$$M_Y^{(1)}(t) = n(1 - p + pe^t)^{n-1} \cdot pe^t$$

At $t = 0$: $M_Y^{(1)}(0) = n(1)^{n-1}p = np = E(Y)$.

2. Second Derivative:

$$M_Y^{(2)}(t) = n(n-1)(1 - p + pe^t)^{n-2}(pe^t)^2 + n(1 - p + pe^t)^{n-1}pe^t$$

At $t = 0$:

$$M_Y^{(2)}(0) = n(n-1)p^2 + np = n^2p^2 - np^2 + np = E(Y^2)$$

**Proof Outline (Taylor Series)**

Using the expansion $e^{tx} = \sum \frac{(tx)^n}{n!}$:

$$M_X(t) = E\left[1 + tX + \frac{t^2X^2}{2!} + \dots\right] = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \dots$$

Differentiating $k$ times and setting $t = 0$ isolates the term $E(X^k)$.

# 18 Properties of MGF & Independence

- **Scaling:** $M_{aX}(t) = M_X(at)$.
- **Sum of Independent Vars:** If $X$ and $Y$ are independent:

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

**Uniqueness Theorem**

The MGF uniquely determines the distribution. If two variables have the same MGF, they have the same distribution.

**Application to Sums:** Let $X_1, \dots, X_n$ be independent.

1. If $X_i \sim \text{Poi}(\lambda_i)$, then $\sum X_i \sim \text{Poi}(\sum \lambda_i)$.
2. If $X_i \sim \text{Bin}(n_i, p)$, then $\sum X_i \sim \text{Bin}(\sum n_i, p)$.
3. If $X_i \sim \text{Geo}(p)$, then $\sum X_i \sim \text{NB}(n, p)$.

# 19 Covariance and Correlation

### Covariance

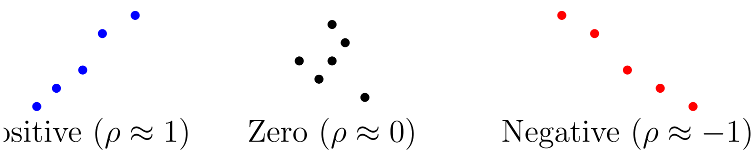
$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)] = E[XY] - E(X)E(Y)$$

- If $X, Y$ are independent, $E(XY) = E(X)E(Y) \implies Cov(X,Y) = 0$.
- $Cov(X,X) = Var(X)$.
- **Linearity:** $Cov(aX+b, cY+d) = ac \cdot Cov(X,Y)$.

### Correlation Coefficient

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

**Visualizing Correlation ($\rho$)**




### Bounds of Correlation

1. $-1 \leq \rho(X,Y) \leq 1$.
2. If $Y = aX + b$ $(a \neq 0)$, then $\rho(X,Y) = \pm 1$ (sign depends on $a$).
3. Conversely, if $\rho(X,Y) = \pm 1$, then $Y$ is a linear function of $X$.

*Proof that $\rho \geq -1$.* Let $Z_1 = \frac{X-\mu_X}{\sigma_X}$ and $Z_2 = \frac{Y-\mu_Y}{\sigma_Y}$. Consider the variance of their sum (which must be non-negative):

$$\begin{aligned} 0 \leq V(Z_1+Z_2) &= E[(Z_1+Z_2)^2] \quad \text{(since means are 0)} \\ &= E[Z_1^2 + Z_2^2 + 2Z_1Z_2] \\ &= 1 + 1 + 2E[Z_1Z_2] \\ &= 2 + 2\rho(X,Y) \end{aligned}$$

$$2\rho(X,Y) \geq -2 \implies \rho(X,Y) \geq -1$$

□

# 20 Variance of Sums

**General Variance Formula**

For any random variables $X_1, \ldots, X_n$:

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

If $X_1, \ldots, X_n$ are **independent**, then $Cov(X_i, X_j) = 0$ for $i \neq j$, so:

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$$

# 21 Continuous Random Variable

**Probability Density Function (PDF)**

A function $f : \mathbb{R} \to \mathbb{R}$ is called a PDF if:

1. $f(x) \geq 0,\ \forall x$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$

**Continuous Probability**

A Continuous Random Variable $X$ has pdf $f(x)$ if for any $a \leq b$:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

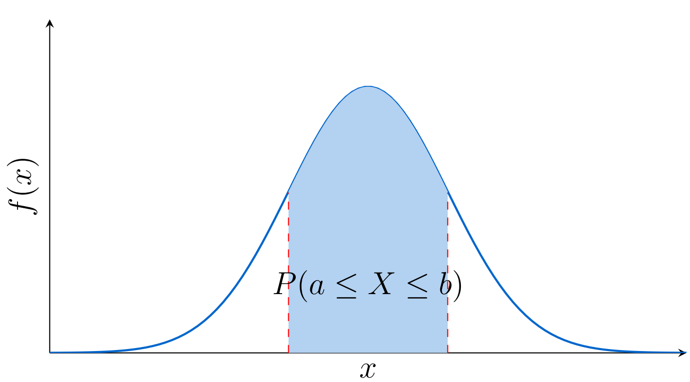



Figure 5: Probability corresponds to the area under the curve.

**Remarks**

- A continuous random variable takes values in an interval.
- The probability of a single point is zero:

$$P(X = a) = \int_a^a f(x)dx = 0$$

- Consequently, strict inequalities do not change the probability:

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b)$$

### 21.1 Example: Imperfection on a Tyre

Let $X$ = angle measure clockwise to the location of an imperfection. $X \in [0, 360]$.

$$f(x) = \begin{cases} \frac{1}{360} & \text{if } 0 \leq x \leq 360 \\ 0 & \text{otherwise} \end{cases}$$

**1. Probability imperfection is between $90^\circ$ & $180^\circ$:**

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360} dx = \frac{1}{360} [x]_{90}^{180} = \frac{90}{360} = \frac{1}{4}$$

**2. Probability imperfection is within $90^\circ$ of the reference line ($0^\circ$):** This corresponds to the union of intervals $[0, 90]$ and $[270, 360]$.

$$P = \int_0^{90} f(x)dx + \int_{270}^{360} f(x)dx = \frac{90}{360} + \frac{90}{360} = \frac{1}{2}$$

## 22 Uniform Distribution

**Uniform Distribution**

$X$ has a **Uniform distribution** on $[A, B]$, denoted $X \sim U[A, B]$, if:

$$f(x) = \begin{cases} \frac{1}{B-A} & \text{if } A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

**Triangular Distribution Example**

Let $Y$ be a random variable with pdf:

$$f(y) = \begin{cases} |y| & \text{if } -1 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \implies f(y) = \begin{cases} -y & \text{if } -1 \leq y \leq 0 \\ y & \text{if } 0 \leq y \leq 1 \end{cases}$$

**Find** $P(-1/3 < Y \leq 2/3)$**:**

$$\begin{aligned} P &= \int_{-1/3}^{0} (-y)dy + \int_{0}^{2/3} (y)dy \\ &= \left[-\frac{y^2}{2}\right]_{-1/3}^{0} + \left[\frac{y^2}{2}\right]_{0}^{2/3} \\ &= \left(0 - (-\frac{1}{18})\right) + \left(\frac{4}{18} - 0\right) = \frac{1}{18} + \frac{4}{18} = \frac{5}{18} \end{aligned}$$

# 23 Cumulative Distribution Function (CDF)

**CDF Definition**

The cumulative distribution function (cdf) of $X$ is:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

## 23.1 Properties of CDF

1. **Monotonically Increasing:** $x_1 \leq x_2 \implies F(x_1) \leq F(x_2)$.
2. **Limits:**
$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1$$
3. **Right Continuous:** $F(x) = \lim_{h\to 0^+} F(x+h)$. (If $X$ is continuous, $F$ is continuous everywhere).

**Fundamental Theorem of Calculus**

For a continuous random variable, wherever $F$ is differentiable:

$$f(x) = \frac{d}{dx}F(x) = F'(x)$$

## 23.2 Visualizing Uniform CDF

Let $X \sim U[1, 4]$. The CDF is linear between $A$ and $B$.

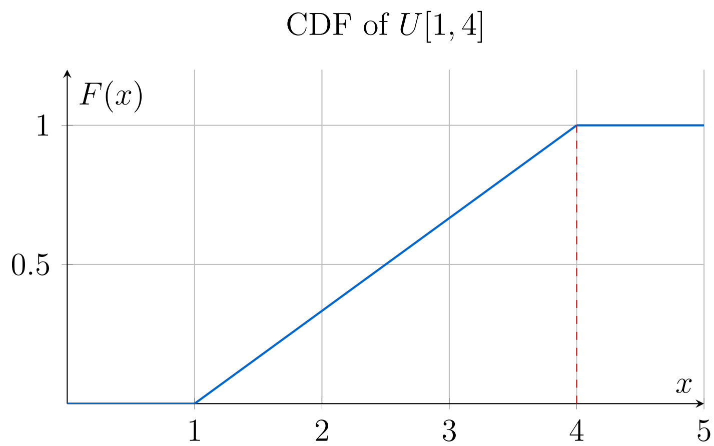



Figure 6: The slope represents the constant probability density.

### 23.3 CDF of Discrete Variable (Bernoulli)

Consider $X \sim \text{Ber}(p)$.

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

*Observation:* The CDF of a discrete random variable is a **step function**.

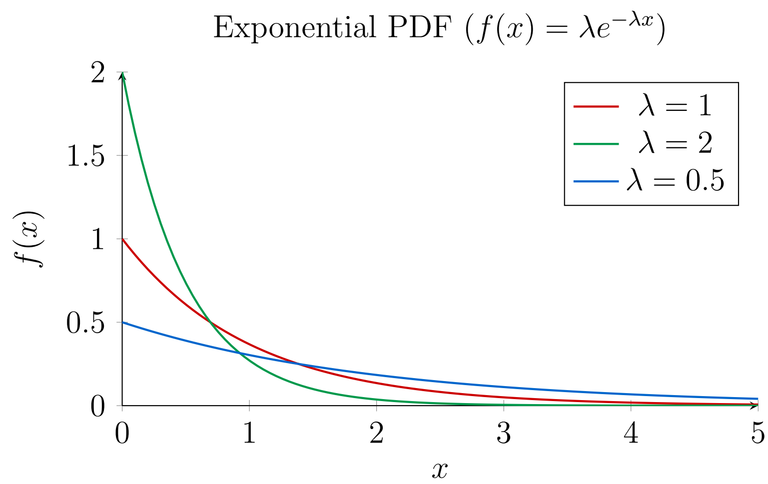
## 24 Exponential Distribution

[Image of Exponential Distribution PDF]

**Exponential Distribution**

Let $\lambda > 0$. $X$ has an exponential distribution with parameter $\lambda$ if:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$




### 24.1 Derivation of Exponential CDF

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} \lambda e^{-\lambda y} dy$$

$$= \lambda \left[ \frac{e^{-\lambda y}}{-\lambda} \right]_0^x = -(e^{-\lambda x} - e^0)$$

$$\boxed{F(x) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0}$$

**Radioactive Decay**

Let $N_0$ be the number of radioactive particles at time 0. The proportion remaining at time $t$ follows an exponential decay:

$$\frac{N_t}{N_0} = e^{-\lambda t}$$

# 25 CDF Properties and Transformations

### 25.1 Discrete Random Variables

**Discrete CDF Calculation**

Given a discrete random variable $X$ with probability mass function $p(x)$:

| $x$ | $-0.5$ | $0.25$ | $3.3$ |
|---|---|---|---|
| $p(x)$ | 0.1 | 0.7 | 0.2 |

Find the Cumulative Distribution Function $F(x)$.
**Solution:**

- For $x < -0.5$: $F(x) = 0$

- At $x = -0.5$: $F(-0.5) = P(X \leq -0.5) = p(-0.5) = 0.1$
- At $x = 0.25$: $F(0.25) = \sum_{x \leq 0.25} p(x) = 0.1 + 0.7 = 0.8$
- At $x = 3.3$: $F(3.3) = 0.1 + 0.7 + 0.2 = 1$
- For $x \geq 3.3$: $F(x) = 1$

**Relation between PMF and CDF**

In the discrete case, the probability at a specific point is the magnitude of the "jump" in the CDF:

$$p(x) = F(x) - F(x^-)$$

where $F(x^-) = \lim_{y \uparrow x} F(y)$.

### 25.2 Continuous Random Variables

Let $X$ be a continuous random variable with CDF $F(x)$ and PDF $f(x)$.

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

$$f(x) = F'(x) \quad \text{where } F'(x) \text{ exists.}$$

## 26 Transformations of Random Variables

### 26.1 Square Transformation ($Y = X^2$)

Let $X$ be a continuous random variable with PDF $f_X(x)$. Let $Y = X^2$. We wish to find the PDF of $Y$.

For $y > 0$:

$$\begin{aligned} G(y) = P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Differentiating with respect to $y$ (using the Chain Rule):

$$G'(y) = f_X(\sqrt{y}) \cdot \frac{d}{dy}(\sqrt{y}) - f_X(-\sqrt{y}) \cdot \frac{d}{dy}(-\sqrt{y})$$

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

**PDF of Square Transformation**

The PDF of $Y = X^2$ is:

$$f_Y(y) = \begin{cases} \dfrac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} & \text{if } y > 0 \\ 0 & \text{if } y \leq 0 \end{cases}$$

## 26.2 General Transformation Theorem

Let $X$ be a continuous random variable with PDF $f_X(x)$. Let $\phi$ be a differentiable, strictly monotone function on an interval $I$. Define $Y = \phi(X)$.

**Change of Variable Technique**

Let $y \in \phi(I)$. The PDF of $Y$ is given by:

$$g(y) = f_X(\phi^{-1}(y)) \left| \frac{d}{dy}\phi^{-1}(y) \right|$$

Alternatively written as: $g(y)dy = f(x)dx$.

*Proof.* Let $F$ and $G$ be the CDFs of $X$ and $Y$.

**Case 1: $\phi$ is strictly increasing.**

$$G(y) = P(Y \leq y) = P(\phi(X) \leq y) = P(X \leq \phi^{-1}(y)) = F(\phi^{-1}(y))$$

Differentiating: $G'(y) = f(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y)$.

**Case 2: $\phi$ is strictly decreasing.**

$$G(y) = P(Y \leq y) = P(\phi(X) \leq y) = P(X \geq \phi^{-1}(y)) = 1 - F(\phi^{-1}(y))$$

Differentiating: $G'(y) = -f(\phi^{-1}(y))\frac{d}{dy}\phi^{-1}(y)$.

Since the derivative of a decreasing function is negative, $-\frac{d}{dy}\phi^{-1}(y) = \left|\frac{d}{dy}\phi^{-1}(y)\right|$. Thus, both cases satisfy the theorem. □

## 26.3 Examples of Transformations

**Uniform to Exponential**

Let $X \sim U(0,1)$, so $f_X(x) = 1$ for $0 \leq x \leq 1$. Let $Y = -\frac{1}{\lambda}\ln(1-X)$ for $\lambda > 0$.

Find the distribution of $Y$.

$$\begin{aligned} G(y) = P(Y \le y) = P\left(-\frac{1}{\lambda}\ln(1-X) \le y\right) \\ &= P(\ln(1-X) \ge -\lambda y) \\ &= P(1-X \ge e^{-\lambda y}) \\ &= P(X \le 1-e^{-\lambda y}) \\ &= F_X(1-e^{-\lambda y}) = 1-e^{-\lambda y} \end{aligned}$$

Thus, $Y \sim \mathrm{Exp}(\lambda)$.

**Power Transformation (Weibull-like)**

Let $X \sim \mathrm{Exp}(\lambda)$ with pdf $f(x) = \lambda e^{-\lambda x}$ for $x > 0$. Let $\beta > 0$ and $Y = X^{1/\beta}$. Inverse function: $x = y^\beta \implies \frac{dx}{dy} = \beta y^{\beta-1}$. Using the General Transformation Theorem:

$$g(y) = f(y^\beta)\left|\beta y^{\beta-1}\right| = \lambda e^{-\lambda y^\beta}\beta y^{\beta-1}$$

for $y > 0$.

### 26.4 Linear Transformations

**Linear Transformation**

Let $X$ have PDF $f(x)$. Let $Y = aX + b$ where $a \neq 0$. Then the PDF of $Y$ is:

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right) \quad \text{for } -\infty < y < \infty$$

## 27 Percentiles

**Percentile Definition**

Let $X$ be a continuous random variable with cdf $F$. Let $0 < p < 1$. The $(100p)^{th}$ percentile of the distribution of $X$ is the value $x_p$ that satisfies:

$$F(x_p) = P(X \le x_p) = p$$

## Examples

### 1. Uniform Distribution

Let $X \sim \mathcal{U}(A, B)$.

$$F(x) = \frac{x - A}{B - A} \quad \text{for } A \leq x < B$$

Find the $73^{rd}$ percentile ($p = 0.73$).

$$\frac{x - A}{B - A} = 0.73 \implies x = A + 0.73(B - A)$$

### 2. Exponential Distribution

Let $X \sim \exp(\lambda)$.

$$F(x) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0$$

Find the $90^{th}$ percentile ($p = 0.90$).

$$\begin{aligned} 1 - e^{-\lambda x} &= 0.9 \\ e^{-\lambda x} &= 0.1 \\ -\lambda x &= \ln(0.1) \\ x &= -\frac{1}{\lambda} \ln(0.1) \end{aligned}$$

### 3. Piecewise CDF

Consider the CDF:

$$F(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1-x^2}{2} & -1 \leq x < 0 \\ \frac{1+x^2}{2} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

Find the $35^{th}$ percentile. Since $p = 0.35$ and $F(0) = 0.5$, we know $x$ must be in the range $(-1, 0)$.

$$\begin{aligned} \frac{1 - x^2}{2} &= 0.35 \\ 1 - x^2 &= 0.7 \\ x^2 &= 0.3 \end{aligned}$$

Since $x \in (-1, 0)$, we take the negative root: $x = -\sqrt{0.3}$.

# 28 Memoryless Property

**Definition**

A random variable $X$ is memoryless if for all $s, t > 0$:

$$P(X > s + t \mid X > s) = P(X > t)$$

## Continuous Case: Exponential

Suppose $X \sim \exp(\lambda)$. We know $P(X \geq t) = e^{-\lambda t}$.

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(\{X > s + t\} \cap \{X > s\})}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t) \end{aligned}$$

**Note:** The Exponential distribution is the *only* continuous random variable that satisfies this property.

## Discrete Case: Geometric

**Proof for Geometric Distribution**

Let $X \sim \text{Geo}(p)$ with pmf $P(X = k) = (1 - p)^k p$. The probability of survival is $P(X > n) = (1 - p)^{n+1}$ (depending on definition of support, here assuming support $\{0, 1, \dots\}$ implies tail probability is $(1 - p)^{n+1}$ or similar form $(1 - p)^n$).
Using the summation form provided:

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{\sum_{n=s+t}^{\infty} (1 - p)^n p}{\sum_{n=s}^{\infty} (1 - p)^n p} \\ &= \frac{(1 - p)^{s+t}}{(1 - p)^s} \quad \text{(Sum of Geometric Series)} \\ &= (1 - p)^t = P(X > t) \end{aligned}$$

# 29 Cauchy Distribution

**Deriving Cauchy from Uniform**

Let $X \sim \text{Unif}(-\frac{\pi}{2}, \frac{\pi}{2})$ and let $Y = \tan X$. $Y$ takes values in $(-\infty, \infty)$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\tan X \leq y) \\ &= P(X \leq \tan^{-1}(y)) \\ &= \frac{\tan^{-1}(y) - (-\frac{\pi}{2})}{\frac{\pi}{2} - (-\frac{\pi}{2})} \\ &= \frac{\tan^{-1}(y) + \frac{\pi}{2}}{\pi} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(y) \end{aligned}$$

Differentiating to find the PDF:

$$\boxed{f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{\pi(1+y^2)}}$$

# 30 Expectation and Variance (Continuous)

**Formulas**

Let $X$ be continuous with pdf $f(x)$.

- $E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$
- $V(X) = E[X^2] - (E[X])^2$

**Note:** $E[h(X)]$ exists iff $E[|h(X)|] < \infty$.

## Examples

**1. Uniform Distribution** $X \sim \mathcal{U}(A, B)$

$$\begin{aligned} E(X) &= \int_A^B x\frac{1}{B-A}dx = \frac{1}{B-A}\left[\frac{x^2}{2}\right]_A^B = \frac{A+B}{2} \\ E(X^2) &= \int_A^B x^2\frac{1}{B-A}dx = \frac{B^3 - A^3}{3(B-A)} = \frac{B^2 + A^2 + AB}{3} \\ V(X) &= \frac{B^2 + A^2 + AB}{3} - \left(\frac{A+B}{2}\right)^2 = \frac{(B-A)^2}{12} \end{aligned}$$

**2. Exponential Distribution** $X \sim \exp(\lambda)$

$$E(X) = \lambda\int_0^\infty xe^{-\lambda x}dx = \frac{1}{\lambda}$$

$$E(X^2) = \lambda \int_0^\infty x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2}$$
$$V(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

**Note:** The expectation of the **Cauchy Distribution** does not exist (the integral diverges).

# 31 Moment Generating Functions (MGF)

$$M(t) = E[e^{tX}]$$

- **Uniform** $X \sim \mathcal{U}(A, B)$:

$$M_X(t) = \int_A^B \frac{e^{tx}}{B - A} dx = \frac{e^{tB} - e^{tA}}{t(B - A)}$$

- **Exponential** $Y \sim \exp(\lambda)$ (for $t < \lambda$):

$$M_Y(t) = \int_0^\infty e^{ty} \lambda e^{-\lambda y} dy = \lambda \int_0^\infty e^{-(\lambda - t)y} dy = \frac{\lambda}{\lambda - t}$$

# 32 Normal Distribution

**Normal PDF**

A random variable $X$ has a Normal distribution, $X \sim \mathcal{N}(\mu, \sigma^2)$, if its pdf is:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

## 32.1 Standard Normal Distribution

Let $Z = \frac{X-\mu}{\sigma}$. Then $Z \sim \mathcal{N}(0, 1)$ with pdf:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

**Proof that PDF integrates to 1**

Let $I = \int_{-\infty}^{\infty} e^{-x^2/2} dx$. Then:

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx\right) \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy\right) = \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy$$

Switching to polar coordinates ($x = r\cos\theta, y = r\sin\theta, dxdy = rdrd\theta$):

$$I^2 = \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}} rdr = (2\pi)\cdot(1) = 2\pi$$

Thus $I = \sqrt{2\pi}$, and $\frac{1}{\sqrt{2\pi}} I = 1$.

## 32.2 Standard Normal Probability

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

where $\Phi(z)$ is the CDF of the Standard Normal.

## 32.3 MGF of Normal Distribution

We calculate $M_X(t) = E[e^{tX}]$ for $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{ tx - \frac{x^2 - 2x\mu + \mu^2}{2\sigma^2} \right\} dx \\
&= \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{ -\frac{x^2 - 2x(\mu + t\sigma^2)}{2\sigma^2} \right\} dx
\end{aligned}$$

Complete the square in the exponent for $x$ around the mean $(\mu + t\sigma^2)$:

$$\begin{aligned}
&= \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(x-(\mu+t\sigma^2))^2 - (\mu+t\sigma^2)^2}{2\sigma^2} \right\} dx \\
&= e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{(\mu+t\sigma^2)^2}{2\sigma^2}} \underbrace{\left[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-(\mu+t\sigma^2))^2}{2\sigma^2}} dx \right]}_{\text{Integrates to 1 (PDF of } N(\mu+t\sigma^2,\sigma^2))} \\
&= \exp\left\{ \frac{-\mu^2 + (\mu^2 + 2\mu t\sigma^2 + t^2\sigma^4)}{2\sigma^2} \right\} \\
&= \exp\left\{ \frac{2\mu t\sigma^2 + t^2\sigma^4}{2\sigma^2} \right\} = \exp\left\{ \mu t + \frac{1}{2}\sigma^2 t^2 \right\}
\end{aligned}$$

$$\boxed{M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}}$$

### Verifying Moments from MGF

$$\begin{aligned}
M'(t) &= (\mu + \sigma^2 t)M(t) \implies M'(0) = \mu \cdot 1 = \mu = E[X] \\
M''(t) &= \sigma^2 M(t) + (\mu + \sigma^2 t)M'(t) \\
M''(0) &= \sigma^2(1) + \mu(\mu) = \sigma^2 + \mu^2 = E[X^2] \\
V(X) &= E[X^2] - (E[X])^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2
\end{aligned}$$

# 33 Gamma Distribution

**Gamma Function**

For $\alpha > 0$, the Gamma function $\Gamma(\alpha)$ is defined by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

## Properties

1. For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

   *Proof.* We know that

   $$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

   Now, apply integration by parts to get the following

   $$\begin{aligned}
   \Gamma(\alpha) &= x^{\alpha-1} \int e^{-x} dx \Big|_0^\infty + (\alpha - 1) \int_0^\infty x^{\alpha-2} e^{-x} dx \\
   &= \underbrace{x^{\alpha-1} \int e^{-x} dx \Big|_0^\infty}_{0} + (\alpha - 1) \underbrace{\int_0^\infty x^{\alpha-2} e^{-x} dx}_{\Gamma(\alpha-1)}
   \end{aligned}$$

   $$\Gamma(\alpha) = (\alpha - 1)\, \Gamma(\alpha - 1)$$

   □

2. For any positive integer $n$, $\Gamma(n) = (n - 1)!$.

   *Proof.* Consider the previous proof for $\alpha = n$ where n is a positive integer. So, we have

   $$\begin{aligned}
   \Gamma(n) &= (n - 1) \int_0^\infty x^{n-2} e^{-x}\, dx \\
   &= (n - 1)(n - 2) \int_0^\infty x^{n-2} e^{-x}\, dx
   \end{aligned}$$

$$\begin{aligned} &\cdots\cdots \\ &= (n-1)(n-2)\cdots 1 \times \int_0^\infty x^1 e^{-x}\, dx \\ \Gamma(n) &= (n-1)! \end{aligned}$$

$\square$

3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.
4. $\Gamma\left(\dfrac{n}{2}\right) = \dfrac{\sqrt{\pi}}{2^{n-1}} \dfrac{(n-1)!}{\left(\frac{n-1}{2}\right)!}$ where n is an odd integer

   *Proof.* Consider $n = 2k+1$ where $k \in \mathbb{N} \cup \{0\}$. Therefore, we can proof this property using mathematical induction. Hence, we have the following result

$$\Gamma\left(\frac{2k+1}{2}\right) = \frac{\sqrt{\pi}}{2^{2k}} \frac{(2k)!}{(k)!}$$

   (a) **Step 1:** Check for $k = 0$, we have

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

   which is true

   (b) **Step 2:** Assume that the expression holds for some arbitrary $m$ such that $0 < m < k$. Hence, we have

$$\Gamma\left(\frac{2m+1}{2}\right) = \frac{\sqrt{\pi}}{2^{2m}} \frac{(2m)!}{(m)!}$$

   (c) **Step 3:** Verify the result for $m+1$. So, we have

$$\begin{aligned} \Gamma\left(\frac{2m+3}{2}\right) &= \frac{\sqrt{\pi}}{2^{2m}.2^2} \frac{(2m+2)!}{(m+1)!} \\ &= \frac{(2m+2)(2m+1)}{4(m+1)} \frac{\sqrt{\pi}}{2^{2m}} \frac{(2m)!}{(m)!} \\ &= \frac{2m+1}{2} \frac{\sqrt{\pi}}{2^{2m}} \frac{(2m)!}{(m)!} \\ \Gamma\left(\frac{2m+3}{2}\right) &= \frac{2m+1}{2}\, \Gamma\left(\frac{2m+1}{2}\right) \end{aligned}$$

This result follows directly from property 1. Therefore,

$$\boxed{\Gamma\left(\frac{n}{2}\right) = \frac{\sqrt{\pi}}{2^{n-1}}\frac{(n-1)!}{\left(\frac{n-1}{2}\right)!}}$$

□

**Gamma PDF**

A continuous random variable $X$ has a Gamma distribution with parameters $\alpha, \beta > 0$, denoted $X \sim \Gamma(\alpha, \beta)$, if its pdf is:

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

## Cumulative Distribution Function of Gamma Random Variable

There is no simple expression for the distribution function corresponding to $\Gamma(\alpha, \beta)$ except when $\alpha = m$ is a positive integer. In this case we can integrate $\Gamma(\alpha, \beta)$ by parts to obtain the cdf for $x > 0$. Hence, consider the following:

$$\begin{aligned} I(m) &= \int_0^x \frac{1}{\beta^m (m-1)!} x^{m-1} e^{-x/\beta} dx \\ &= \frac{1}{\beta^m (m-1)!}\left[ -\beta x^{m-1} e^{-x/\beta}\Big|_0^x + \beta(m-1)\int_0^x x^{m-2} e^{-x/\beta} dx \right] \\ &= -\frac{x^{m-1}}{\beta^{m-1}(m-1)!} e^{-x/\beta} + \underbrace{\int_0^x \frac{1}{\beta^{m-1}(m-2)!} x^{m-2} e^{-x/\beta} dx}_{I(m-1)} \end{aligned}$$

Hence, we have

$$I(m) - I(m-1) = -\frac{1}{(m-1)!}\left(\frac{x}{\beta}\right)^{m-1} e^{-x/\beta}$$

It's a noteworthy point that $I(1) = 1 - e^{-x/\beta}$. So, we have the following

$$\boxed{I(m) = 1 - e^{-x/\beta}\sum_{k=0}^{m-1}\frac{(x/\beta)^k}{k!}}$$

## Moment Generating Function of a Gamma Random Variable

Let $X \sim \Gamma(\alpha, \beta)$. We know that $M_X(t) = E[e^{tx}]$. Therefore, we have

$$M_X(t) = \int_0^\infty e^{tx}\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx$$

$$M_X(t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \underbrace{\int_0^\infty x^{\alpha-1} e^{-(1/\beta - t)x} dx}_{I(\alpha)}$$

Note that for this function to return any meaningful value, $1/\beta > t$. Let $u = (1/\beta - t)x$. Then, $dx = \dfrac{du}{1/\beta - t}$. When $x = 0, u = 0$ and as $x \to \infty, u \to \infty$.

Substituting these into the integral:

$$\begin{aligned}
I(\alpha) &= \int_0^\infty \left( \frac{u}{1/\beta - t} \right)^{\alpha-1} e^{-u} \frac{du}{1/\beta - t} \\
&= \frac{1}{(1/\beta - t)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u} du \\
&= \frac{\Gamma(\alpha)}{(1/\beta - t)^\alpha}
\end{aligned}$$

Substituting $I(\alpha)$ back into the equation for $M_X(t)$:

$$\begin{aligned}
M_X(t) &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{(1/\beta - t)^\alpha} \\
&= \left( \frac{1}{\beta(1/\beta - t)} \right)^\alpha = \left( \frac{1}{1 - \beta t} \right)^\alpha
\end{aligned}$$

$$\boxed{M_X(t) = (1 - \beta t)^{-\alpha} \quad \text{for } t < \frac{1}{\beta}}$$

## Visualizing the Gamma Distribution

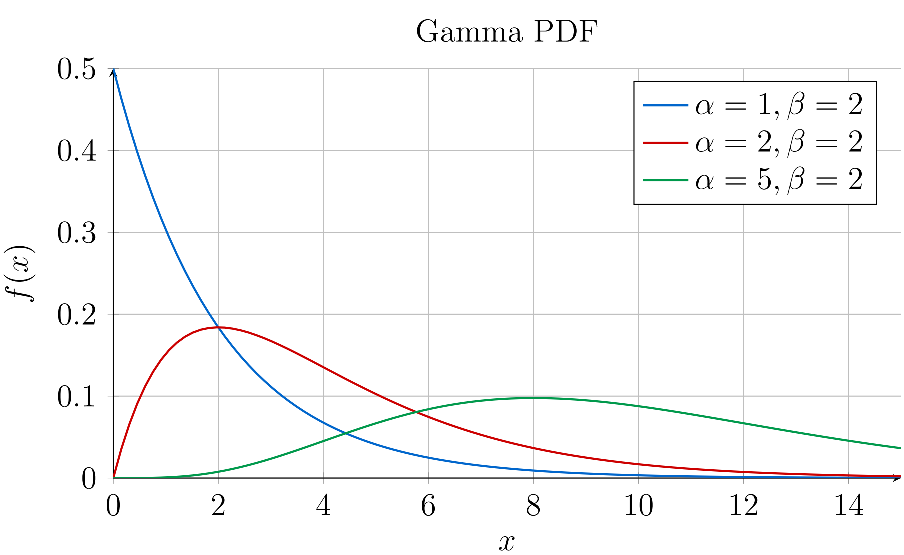



Figure 7: Effect of shape ($\alpha$) parameters on the Gamma distribution ($\beta = 2$).

**Relation to Exponential**

If $\alpha = 1$ and $\beta = \frac{1}{\lambda}$, then:

$$f(x; 1, 1/\lambda) = \lambda e^{-\lambda x} \quad \text{for } x > 0$$

This is the **Exponential Distribution**. It is important to note that while the Gamma Distribution uses parameters $(\alpha, \beta)$, the exponential distribution uses $\lambda$.

# 34 Jointly Distributed Continuous R.V.

**Joint PDF**

Let $X, Y$ be continuous random variables. A joint probability density function $f(x, y)$ satisfies $f(x, y) \geq 0$ and:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

The probability of a region is:

$$P(a \leq X \leq b\ ,\ c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dx\, dy$$

**Fast Food Example**

Let $X$ = proportion of time drive-through is busy, $Y$ = proportion of time inside counter is busy.

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that neither is busy for more than $1/4$ of the time.

$$\begin{aligned} P\left(X \leq \frac{1}{4}, Y \leq \frac{1}{4}\right) &= \int_0^{\frac{1}{4}} \int_0^{\frac{1}{4}} \frac{6}{5}(x + y^2)\, dx\, dy \\ &= \frac{6}{5} \int_0^{\frac{1}{4}} \left[\frac{x^2}{2} + y^2 x\right]_{x=0}^{1/4} dy \\ &= \frac{6}{5} \int_0^{\frac{1}{4}} \left(\frac{1}{32} + \frac{y^2}{4}\right) dy \\ &= \frac{6}{5} \left[\frac{y}{32} + \frac{y^3}{12}\right]_0^{1/4} \approx 0.0190 \end{aligned}$$

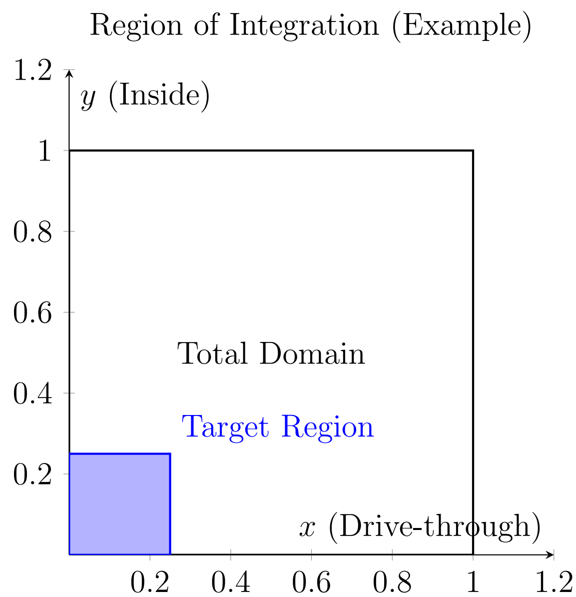



Figure 8: The probability is the volume of the PDF over the shaded blue region.

## 35 Marginal and Conditional Distributions

**Marginal PDF**

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

**Conditional PDF**

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \quad \text{and} \quad f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

## 36 Independence

**Independence Definition**

Two continuous random variables $X$ and $Y$ are independent if for every $(x, y)$:

$$f(x,y) = f_X(x)f_Y(y)$$

**Exercise: Minimum of Exponentials**

Let $X_1, \ldots, X_n$ be independent lifetimes with $X_i \sim \exp(\lambda)$. Find the distribution of $Y = \min(X_1, \ldots, X_n)$.

**Solution:** First, find the survival function of the minimum:

$$\begin{aligned} P(Y > t) &= P(X_1 > t, \dots, X_n > t) \\ &= P(X_1 > t) \cdots P(X_n > t) \quad \text{(Independence)} \\ &= e^{-\lambda t} \cdots e^{-\lambda t} = e^{-n\lambda t} \end{aligned}$$

Thus, the CDF is $F_Y(t) = 1 - e^{-n\lambda t}$. Differentiating gives the PDF:

$$f_Y(t) = n\lambda e^{-n\lambda t}$$

**Conclusion:** The minimum of $n$ i.i.d. exponentials is $\exp(n\lambda)$.

## 37 Expectation of Functions of Two Variables

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y)\, dx\, dy$$

**Cost of Mixed Nuts**

Let $X, Y, Z$ be weights of Almonds, Cashews, Peanuts. $X + Y + Z = 1$. Costs: A=250, C=200, P=300. Cost function:

$$\begin{aligned} h(X, Y) &= 250X + 200Y + 300(1 - X - Y) \\ &= 300 - 50X - 100Y \end{aligned}$$

Expected Cost:

$$E[h(X, Y)] = 300 - 50E[X] - 100E[Y]$$

## 38 Sums of Random Variables (Convolution)

If $X$ and $Y$ are independent, the PDF of $Z = X + Y$ is the convolution of their PDFs:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

If $X, Y \geq 0$, the limits become $\int_0^z$.

**Sum of Independent Gammas**

Let $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$ be independent. Then:

$$X + Y \sim \Gamma(\alpha_1 + \alpha_2, \beta)$$

# 39 Beta Distribution

**Beta PDF**

The Beta distribution with parameters $\alpha_1, \alpha_2 > 0$ has pdf:

$$f(x) = \begin{cases} \frac{1}{B(\alpha_1,\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$.

**Moments of Beta**

If $X \sim \text{Beta}(\alpha_1, \alpha_2)$:

$$E[X] = \frac{\alpha_1}{\alpha_1 + \alpha_2}$$

$$V[X] = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}$$

# 40 Distribution of Quotients

Let $X$ and $Y$ be continuous random variables. Let $Z = Y/X$.

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_{X,Y}(x, xz)\, dx$$

**Ratio of Independent Gammas**

Let $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$ be independent. Let $Z = Y/X$. The pdf of $Z$ is:

$$f_Z(z) = \frac{1}{B(\alpha_1, \alpha_2)} z^{\alpha_2 - 1}(1+z)^{-(\alpha_1+\alpha_2)} \quad \text{for } z > 0$$

(This is known as the Beta Prime distribution).

*Proof.* Using the quotient formula for independent variables ($f_{X,Y} = f_X f_Y$):

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} x \cdot \left(\frac{x^{\alpha_1-1}e^{-x/\beta}}{\beta^{\alpha_1}\Gamma(\alpha_1)}\right)\left(\frac{(xz)^{\alpha_2-1}e^{-xz/\beta}}{\beta^{\alpha_2}\Gamma(\alpha_2)}\right) dx \\ &= \frac{z^{\alpha_2-1}}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^{\infty} x^{(\alpha_1+\alpha_2)-1} e^{-\frac{x}{\beta}(1+z)} dx \end{aligned}$$

The integral is the kernel of a Gamma function. Using $\int_0^\infty t^{k-1}e^{-ct}dt = \frac{\Gamma(k)}{c^k}$:

$$\text{Integral} = \frac{\Gamma(\alpha_1+\alpha_2)}{[\frac{1}{\beta}(1+z)]^{\alpha_1+\alpha_2}} = \frac{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}}{(1+z)^{\alpha_1+\alpha_2}}$$

Substituting this back simplifies to the result. □

# 41 General Linear Transformations

Let $X_1, \dots, X_n$ have joint pdf $f_X$. Let $Y = AX$, where $A$ is an invertible $n \times n$ matrix.

$$f_Y(y) = \frac{1}{|\det(A)|} f_X(A^{-1}y)$$

**Sum of Exponentials (Ordered)**

Let $X_i \sim \exp(\lambda)$ be i.i.d. Let $S_k = \sum_{i=1}^k X_i$. The transformation matrix $A$ is lower triangular with 1s on the diagonal, so $\det(A) = 1$. The inverse transformation is $x_k = s_k - s_{k-1}$ (with $s_0 = 0$).
The joint PDF of $S$ is:

$$f_S(s_1, \dots, s_n) = \begin{cases} \lambda^n e^{-\lambda s_n} & 0 \le s_1 \le \dots \le s_n \\ 0 & \text{otherwise} \end{cases}$$

# 42 MGF of Jointly Distributed R.V.

**Joint MGF**

Let $X$ and $Y$ be jointly distributed random variables. The joint moment generating function for $(X, Y)$ is:

$$M_{X,Y}(s,t) = E[e^{sX+tY}]$$

**Uniqueness Theorem**

1. If $M_X(t) = M_Y(t)$ for all $t$ in a neighborhood of 0, then $X$ and $Y$ have the same distribution.
2. If $M_{X,W}(s,t) = M_{Y,Z}(s,t)$ for all $s,t$ in a rectangle containing $(0,0)$, then $(X,W)$ and $(Y,Z)$ have the same joint distribution.

## Example: Linear Transformation

Let $X \sim \mathcal{N}(0,1)$ and $Y = cX + d$. We compute the MGF of $Y$:

$$M_Y(t) = E[e^{t(cX+d)}] = e^{td}E[e^{(tc)X}] = e^{td}e^{\frac{1}{2}(tc)^2} = \exp\left\{dt + \frac{1}{2}c^2t^2\right\}$$

This confirms $Y \sim \mathcal{N}(d, c^2)$.

# 43 Independence via MGF

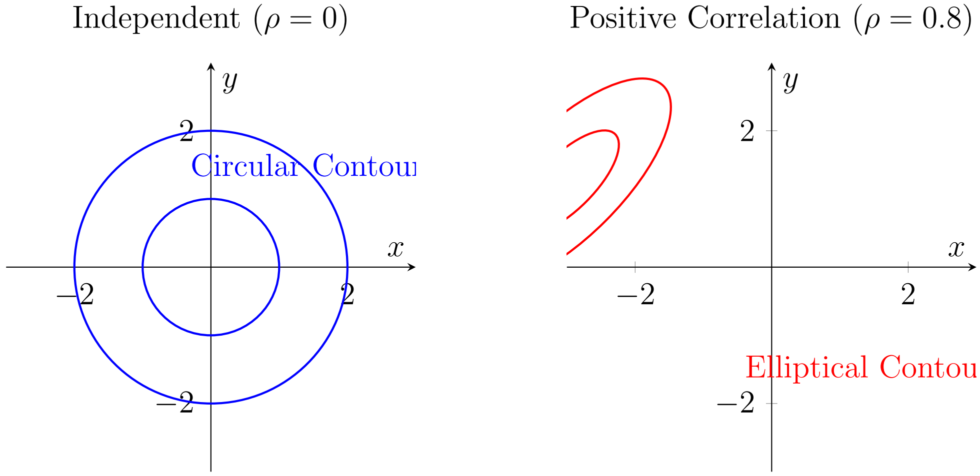



Figure 9: Top-down view (Contours). Left: Independence implies circular symmetry. Right: Correlation stretches the distribution along the diagonal.

**Independence Condition**

Let $X$ and $Y$ be random variables with joint MGF $M_{X,Y}(s,t)$. Then $X$ and $Y$ are independent **if and only if**:

$$M_{X,Y}(s,t) = M_X(s)M_Y(t)$$

for all $s, t$ in some neighborhood of zero.

*Proof.* **Forward ($\implies$):** If $X$ and $Y$ are independent, then:

$$\begin{aligned} M_{X,Y}(s,t) &= E[e^{sX+tY}] = E[e^{sX}e^{tY}] \\ &= E[e^{sX}]E[e^{tY}] \quad \text{(by Independence)} \\ &= M_X(s)M_Y(t) \end{aligned}$$

**Converse ($\impliedby$):** Suppose $M_{X,Y}(s,t) = M_X(s)M_Y(t)$. Let $\hat{X}$ and $\hat{Y}$ be independent random variables such that $\hat{X}$ has the same distribution as $X$, and $\hat{Y}$ has the same distribution as $Y$. Then the joint MGF of $(\hat{X}, \hat{Y})$ is:

$$M_{\hat{X},\hat{Y}}(s,t) = M_{\hat{X}}(s)M_{\hat{Y}}(t) = M_X(s)M_Y(t) = M_{X,Y}(s,t)$$

By the Uniqueness Theorem, $(X,Y)$ and $(\hat{X},\hat{Y})$ have the same joint distribution. Since $\hat{X},\hat{Y}$ are independent, $X, Y$ must also be independent. $\square$

### Example: Sum of Independent Normals

Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent.

$$\begin{aligned}
M_{aX+bY}(t) &= E[e^{t(aX+bY)}] = E[e^{(at)X}e^{(bt)Y}] \\
&= M_X(at)M_Y(bt) \\
&= \exp\left\{\mu_1(at) + \frac{\sigma_1^2(at)^2}{2}\right\} \exp\left\{\mu_2(bt) + \frac{\sigma_2^2(bt)^2}{2}\right\} \\
&= \exp\left\{(a\mu_1 + b\mu_2)t + \frac{1}{2}(a^2\sigma_1^2 + b^2\sigma_2^2)t^2\right\}
\end{aligned}$$

This is the MGF of a Normal distribution. Thus:

$$aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

**Generalization**

If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ are independent, then:

$$Y = \sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

# 44 Bivariate Normal Distribution

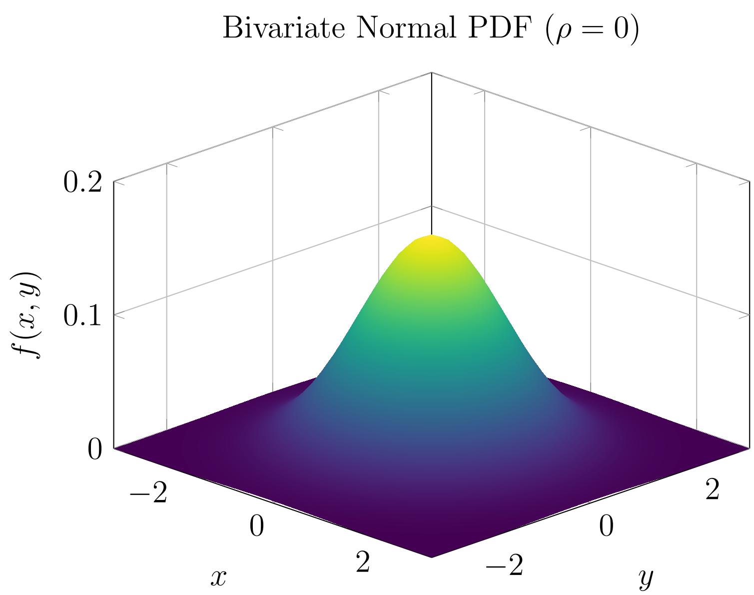



Figure 10: The joint PDF of two independent standard normal variables looks like a 3D bell.

**Bivariate Normal Definition**

A pair of random variables $(X, Y)$ is called **Bivariate Normal** if any linear combination $aX + bY$ is normally distributed for all $a, b \in \mathbb{R}$.

**Remarks**

1. **Degenerate Case:** If $aX + bY = c$ (a constant), it is considered Normal with variance 0.
2. If $(X, Y)$ is bivariate normal, then the marginals $X$ (taking $b = 0$) and $Y$ (taking $a = 0$) are individually Normal.
3. **Caution:** If $X$ and $Y$ are individually Normal, it does **not** imply $(X, Y)$ is Bivariate Normal (unless they are independent or have a specific dependence structure).

**Uniqueness of Parameters**

Suppose $(X, Y)$ and $(Z, W)$ are both bivariate normal sets with matching means ($E[X] = E[Z]$, etc.), matching variances ($V(X) = V(Z)$, etc.), and matching covariance ($Cov(X, Y) = Cov(Z, W)$). Then $(X, Y)$ and $(Z, W)$ have the same joint distribution.

*Proof.* Let $s, t \in \mathbb{R}$. Let $L_1 = sX + tY$ and $L_2 = sZ + tW$. Since both are linear combinations of bivariate normals, $L_1$ and $L_2$ are univariate Normal. Their means and variances depend only on individual means, variances, and covariance (which are identical). Thus, $L_1 \sim L_2$.

$$\begin{aligned}
M_{X,Y}(s,t) &= E[e^{sX+tY}] = E[e^{L_1}] = M_{L_1}(1) \\
&= M_{L_2}(1) = E[e^{sZ+tW}] \\
&= M_{Z,W}(s,t)
\end{aligned}$$

Since the joint MGFs are identical, the distributions are identical. □

**Independence of Bivariate Normals**

Let $(X, Y)$ be Bivariate Normal. Then $X$ and $Y$ are independent **if and only if** they are uncorrelated (i.e., $Cov(X, Y) = 0$).

*Proof.* ($\implies$) Independence always implies covariance is 0.

($\impliedby$) Suppose $Cov(X, Y) = 0$. We compute the joint MGF. The variance of $sX + tY$ is $s^2\sigma_1^2 + t^2\sigma_2^2 + 2st(0)$.

$$\begin{aligned}
M_{X,Y}(s,t) &= M_{sX+tY}(1) \\
&= \exp\left\{1(s\mu_1 + t\mu_2) + \frac{1}{2}(1)^2(s^2\sigma_1^2 + t^2\sigma_2^2)\right\} \\
&= \exp\left\{s\mu_1 + \frac{1}{2}s^2\sigma_1^2\right\} \cdot \exp\left\{t\mu_2 + \frac{1}{2}t^2\sigma_2^2\right\} \\
&= M_X(s) \cdot M_Y(t)
\end{aligned}$$

Since the joint MGF factors into the product of marginal MGFs, $X$ and $Y$ are independent. $\square$