# Probability Theory Recap

Gioele La Manno

École Polytechnique Fédérale de Lausanne (EPFL)

School of Life Science (SV)

February 2025

EPFL - BMI - UPLAMANNO

# Contents

# Random Variables

# A New Type of Variable: Random Variables

- **"Regular" Variables:**
  - Deterministic: $y = f(x)$
  - One input $\rightarrow$ One output
  - Example: $y = 2x + 1$
- **Random Variables:**
  - Not a value, but a generator
  - Same input $\rightarrow$ Different outputs
  - Described by probability functions
- **Fundamental Shift:**
  - From *"What is the value?"*
  - To *"What is the probability?"*
- **Full Description Needs:**
  1. Support (possible values)
  2. pdf/pmf (likelihood)
  3. cdf (cumulative probability)
  4. inverse cdf (quantiles)

# A New Type of Variable: Random Variables

- **"Regular" Variables:**
  - Deterministic: $y = f(x)$
  - One input $\rightarrow$ One output
  - Example: $y = 2x + 1$
- **Random Variables:**
  - Not a value, but a generator
  - Same input $\rightarrow$ Different outputs
  - Described by probability functions

- **Fundamental Shift:**
  - From *"What is the value?"*
  - To *"What is the probability?"*
- **Full Description Needs:**
  1. Support (possible values)
  2. pdf/pmf (likelihood)
  3. cdf (cumulative probability)
  4. inverse cdf (quantiles)

# A New Type of Variable: Random Variables

- **"Regular" Variables:**
  - Deterministic: $y = f(x)$
  - One input $\rightarrow$ One output
  - Example: $y = 2x + 1$
- **Random Variables:**
  - Not a value, but a generator
  - Same input $\rightarrow$ Different outputs
  - Described by probability functions

- **Fundamental Shift:**
  - From *"What is the value?"*
  - To *"What is the probability?"*
- **Full Description Needs:**
  1. Support (possible values)
  2. pdf/pmf (likelihood)
  3. cdf (cumulative probability)
  4. inverse cdf (quantiles)

# A New Type of Variable: Random Variables

- **"Regular" Variables:**
  - Deterministic: $y = f(x)$
  - One input $\rightarrow$ One output
  - Example: $y = 2x + 1$
- **Random Variables:**
  - Not a value, but a generator
  - Same input $\rightarrow$ Different outputs
  - Described by probability functions

- **Fundamental Shift:**
  - From *"What is the value?"*
  - To *"What is the probability?"*
- **Full Description Needs:**
  1. Support (possible values)
  2. pdf/pmf (likelihood)
  3. cdf (cumulative probability)
  4. inverse cdf (quantiles)

# Support

- **Support of a Random Variable:**
  - **Definition:** The set of all possible values that a random variable $X$ can take.
  - **Notation:** $\text{Supp}(X) = \{x \in \mathbb{R} : P(X = x) > 0\}$ for discrete, or $\text{Supp}(X) = x \in [a, b]$ for continuous.
  - **Purpose:** Defines the "range" or domain where the random variable is valid.

# Types of Random Variables

## Types of Random Variables

- **Discrete Random Variable:**
  - Takes on a countable set of values (e.g., integers).
  - Defined by a probability mass function (PMF): $P(X = x_i)$.
  - *Example:* Number of heads in a series of coin flips.

- **Continuous Random Variable:**
  - Takes on an uncountable set of values, typically an interval.
  - Defined by a probability density function (PDF): $p(x)$ where $P(a \leq X \leq b) = \int_a^b p(x)\, dx$.
  - *Example:* Height of individuals in a population.

# Types of Random Variables

## Types of Random Variables

- **Discrete Random Variable:**
  - Takes on a countable set of values (e.g., integers).
  - Defined by a probability mass function (PMF): $P(X = x_i)$.
  - *Example:* Number of heads in a series of coin flips.

- **Continuous Random Variable:**
  - Takes on an uncountable set of values, typically an interval.
  - Defined by a probability density function (PDF): $p(x)$ where $P(a \leq X \leq b) = \int_a^b p(x)\,dx$.
  - *Example:* Height of individuals in a population.

# Functions associated to a Random Variable

# Probability Density Function (PDF)

- **Probability Density Function (PDF):**
  - Represents probability density at a point (continuous RVs)

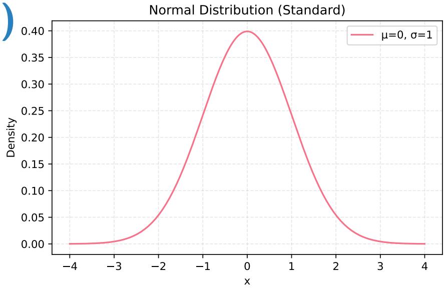
- **Properties:**
  - $p(x) \geq 0 \ \forall x$ *(non-negative)*
  - $\int_{-\infty}^{\infty} p(x)\, dx = 1$ *(normalized)*
- **Practical use:** Finding probability in an interval (e.g., $P(a \leq X \leq b)$)

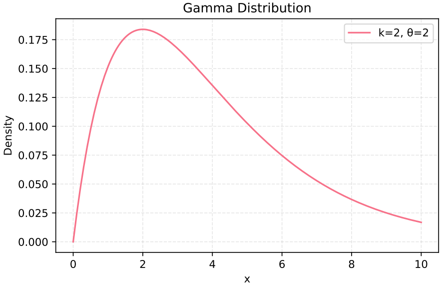
$$P(a \leq X \leq b) = \int_a^b p(x)\, dx$$

# Probability Density Function (PDF)

- **Probability Density Function (PDF):**
  - Represents probability density at a point (continuous RVs)
- **Properties:**
  - $p(x) \geq 0 \ \forall x$ *(non-negative)*
  - $\int_{-\infty}^{\infty} p(x)\, dx = 1$ *(normalized)*
- **Practical use:** Finding probability in an interval (e.g., $P(a \leq X \leq b)$)

$$P(a \leq X \leq b) = \int_a^b p(x)\, dx$$



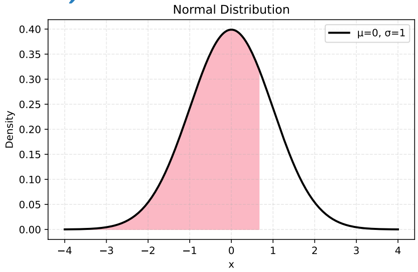
# Cumulative Distribution Function (CDF)

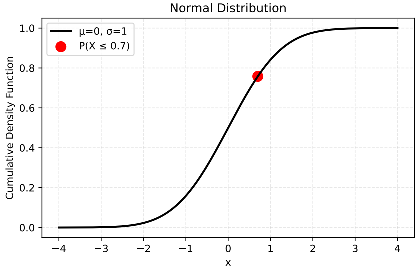
- **Cumulative Distribution Function (CDF):**
  - $F(x) = P(X \leq x) = \int_{-\infty}^{x} p(t)\, dt$
- **Properties:**
  - $0 \leq F(x) \leq 1$ *(probability bounds)*
  - $F(-\infty) = 0,\ F(\infty) = 1$
  - $F(x)$ is monotonically increasing
  - $F(x)$ is right-continuous

# Cumulative Distribution Function (CDF)

- **Cumulative Distribution Function (CDF):**
  - $F(x) = P(X \leq x) = \int_{-\infty}^{x} p(t)\, dt$

- **Properties:**
  - $0 \leq F(x) \leq 1$ *(probability bounds)*
  - $F(-\infty) = 0,\ F(\infty) = 1$
  - $F(x)$ is monotonically increasing
  - $F(x)$ is right-continuous



# Cumulative Distribution Function (CDF)

- **Practical Use:**
  - **Quantile Determination:** Identifies values at specified probabilities (e.g., median or percentiles).
  - **Statistical Testing:** CDFs are central in calculating p-values and test statistics in hypothesis testing.
  - **Distribution Transformations:** Allows mapping from one distribution to another, used in simulation and resampling methods.

# Other Useful Functions for Random Variables

- **Inverse CDF (Quantile Function)**
  Provides the value $x$ for a given probability $p$.

$$F^{-1}(p) = x \text{ such that } F(x) = p$$

  **Use:** Finding quantiles, and for setting probability thresholds.

- **Survival Function:**
  Definition: Probability that $X$ exceeds $x$

$$S(x) = P(X \geq x) = 1 - F(x)$$

  **Use:** Commonly used to compute p-values.

# Other Useful Functions for Random Variables

- **Inverse CDF (Quantile Function)**
  Provides the value $x$ for a given probability $p$.

  $$F^{-1}(p) = x \text{ such that } F(x) = p$$

  **Use:** Finding quantiles, and for setting probability thresholds.

- **Survival Function:**
  Definition: Probability that $X$ exceeds $x$

  $$S(x) = P(X \geq x) = 1 - F(x)$$

  **Use:** Commonly used to compute p-values.

# Other Useful Functions for Random Variables

- **Moment Generating Function (MGF)**
  Encodes the moments (mean, variance, etc.) of $X$.

$$M_X(t) = \mathbb{E}[e^{tX}]$$

  **Use:** Useful for calculating moments and deriving properties of distributions.

- **Characteristic Function (CF):**
  Complex-valued function that provides an alternative to the MGF for characterizing the distribution.

$$\phi_X(t) = \mathbb{E}[e^{itX}]$$

  **Use:** Particularly useful in Fourier transforms and convolution operations on distributions.

# Other Useful Functions for Random Variables

- **Moment Generating Function (MGF)**
  Encodes the moments (mean, variance, etc.) of $X$.

$$M_X(t) = \mathbb{E}[e^{tX}]$$

  **Use:** Useful for calculating moments and deriving properties of distributions.

- **Characteristic Function (CF):**
  Complex-valued function that provides an alternative to the MGF for characterizing the distribution.

$$\phi_X(t) = \mathbb{E}[e^{itX}]$$

  **Use:** Particularly useful in Fourier transforms and convolution operations on distributions.

# Related Functions for Discrete Random Variables

**Probability Mass Function (PMF):**

Definition: Function that gives the probability of each possible value of a discrete random variable $X$.

$$P(X = x_i) = p(x_i)$$

Properties:

- $p(x_i) \geq 0 \; \forall \; i$
- $\sum p(x_i) = 1$ (normalization)

**Examples:**

- PMF: Probability of rolling each side of a die.

# Related Functions for Discrete Random Variables

**Cumulative Mass Function (CMF):**

Definition: Function that gives the cumulative probability up to a certain value.

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

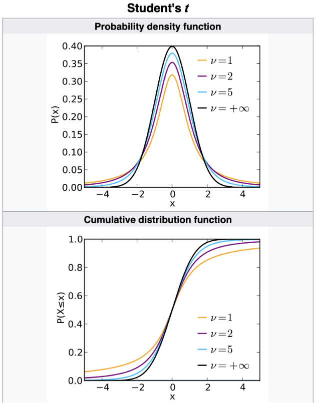
Properties:

- $F(x)$ is non-decreasing
- $0 \leq F(x) \leq 1$

**Examples:**

- Probability of rolling a number less than or equal to a certain value.

# Where to find them?




| | |
|---|---|
| Parameters | $\nu > 0$ degrees of freedom (real, almost always a positive integer) |
| Support | $x \in (-\infty, \infty)$ |
| PDF | $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi \nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ |
| CDF | $\frac{1}{2} + x\,\Gamma\left(\frac{\nu+1}{2}\right) \times \frac{{}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}$ ,<br>where ${}_2F_1( , ; ; )$ is the hypergeometric function |
| Mean | $0$ for $\nu > 1$ , otherwise undefined |
| Median | $0$ |
| Mode | $0$ |
| Variance | $\frac{\nu}{\nu-2}$ for $\nu > 2$ , ∞ for $1 < \nu \le 2$ , otherwise undefined |
| Skewness | $0$ for $\nu > 3$ , otherwise undefined |
| Excess kurtosis | $\frac{6}{\nu-4}$ for $\nu > 4$ , ∞ for $2 < \nu \le 4$ , otherwise undefined |
| Entropy | $\frac{\nu+1}{2}\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] + \ln\left[\sqrt{\nu}\,B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right]$ (nats)<br>where<br>$\psi()$ is the digamma function,<br>$B( , )$ is the beta function. |
| MGF | undefined |
| CF | $\frac{\left(\sqrt{\nu}\,\lvert t\rvert\right)^{\nu/2} K_{\nu/2}\left(\sqrt{\nu}\,\lvert t\rvert\right)}{\Gamma(\nu/2)\,2^{\nu/2-1}}$ for $\nu > 0$<br>• $K_\nu(x)$ is the modified Bessel function of the second kind[1] |
| Expected shortfall | $\mu + s\left(\frac{\nu + T^{-1}(1-p)^2 \times \tau\left(T^{-1}(1-p)^2\right)}{(\nu-1)(1-p)}\right)$<br>Where $T^{-1}( )$ is the inverse standardized Student *t* CDF, and $\tau( )$ is the standardized Student t PDF.[2] |

Figure: Summary of Probability Density Functions (Source: Wikipedia)

# Expectation of a Random Variable

# Expectation of a Random Variable

The **expectation** of a random variable $X$, denoted $E[X]$, is an operator defined as:

- **(Discrete)**

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

- **(Continuous)**

$$E[X] = \int_{-\infty}^{\infty} x\, p(x)\, dx$$

*Intuition*

- It is an operation that can be thought of as a weighted average of all possible values of $X$, with weights given by the probabilities.
- It provides the theoretical center or long-run average of the random variable.
- Repeting the experiment infinitely, the average outcome would converge to $E[X]$.

# Expectation of a Random Variable

The **expectation** of a random variable $X$, denoted $E[X]$, is an operator defined as:

- **(Discrete)**

$$E[X] = \sum_{x} x\, p(x)$$

- **(Continuous)**

$$E[X] = \int_{-\infty}^{\infty} x\, p(x)\, dx$$

*Intuition*

- It is an operation that can be thought of as a weighted average of all possible values of $X$, with weights given by the probabilities.
- It provides the theoretical center or long-run average of the random variable.
- Repeting the experiment infinitely, the average outcome would converge to $E[X]$.

# Properties of Expectation

- **Linearity:** $E[aX + bY] = aE[X] + bE[Y]$
- **Expectation of a Function:** $E[g(X)] = \sum_x g(x)\,p(x)$ or $\int g(x)\,p(x)\,dx$
- **Additivity for 2 RVs:** $E[X + Y] = E[X] + E[Y]$
- It can be defined with respect to marginal distribution (often implicit):

$$E_{p(x)}[X + Y] = E_{p(x)}[X] + Y$$

- **Expectation of a Product:** $E[XY] = E[X]E[Y]$ if $X$ and $Y$ are independent
- Avoid common naïve misconception:

$$E[X^2] = E[X \cdot X] \neq E[X]^2$$

# Properties of Expectation

- **Linearity:** $E[aX + bY] = aE[X] + bE[Y]$
- **Expectation of a Function:** $E[g(X)] = \sum_x g(x)\, p(x)$ or $\int g(x)\, p(x)\, dx$
- **Additivity for 2 RVs:** $E[X + Y] = E[X] + E[Y]$
- It can be defined with respect to marginal distribution (often implicit):

$$E_{p(x)}[X + Y] = E_{p(x)}[X] + Y$$

- **Expectation of a Product:** $E[XY] = E[X]E[Y]$ if $X$ and $Y$ are independent
- Avoid common naïve misconception:

$$E[X^2] = E[X \cdot X] \neq E[X]^2$$

# Properties of Expectation

- **Linearity:** $E[aX + bY] = aE[X] + bE[Y]$
- **Expectation of a Function:** $E[g(X)] = \sum_x g(x)\, p(x)$ or $\int g(x)\, p(x)\, dx$
- **Additivity for 2 RVs:** $E[X + Y] = E[X] + E[Y]$
- It can be defined with respect to marginal distribution (often implicit):

$$E_{p(x)}[X + Y] = E_{p(x)}[X] + Y$$

- **Expectation of a Product:** $E[XY] = E[X]E[Y]$ if $X$ and $Y$ are independent
- Avoid common naïve misconception:

$$E[X^2] = E[X \cdot X] \neq E[X]^2$$

# Properties of Expectation

- **Linearity:** $E[aX + bY] = aE[X] + bE[Y]$
- **Expectation of a Function:** $E[g(X)] = \sum_x g(x)\,p(x)$ or $\int g(x)\,p(x)\,dx$
- **Additivity for 2 RVs:** $E[X + Y] = E[X] + E[Y]$
- It can be defined with respect to marginal distribution (often implicit):

$$E_{p(x)}[X + Y] = E_{p(x)}[X] + Y$$

- **Expectation of a Product:** $E[XY] = E[X]E[Y]$ if $X$ and $Y$ are independent
- Avoid common naïve misconception:

$$E[X^2] = E[X \cdot X] \neq E[X]^2$$

# Moments of a Random Variable

# Moments of a Random Variable

Moments provide information about the shape and spread of the distribution of $X$. The $k$-th moment of a random variable $X$ is defined as the expectation of $X^k$, denoted $E[X^k]$:

$$E[X^k] = \int_{-\infty}^{\infty} x^k \, p(x) \, dx$$

**Moments:**

- **First Moment (Mean):** $E[X]$
- **Second Moment:** $E[X^2]$. (captures both the mean and the spread of $X$ around the mean, related to variance).

  *Do not confuse with Variance: Defined as* $Var(X) = E[X^2] - (E[X])^2$
- **Higher-Order Moments:**
  - Third Moment (related to Skewness): relates to the asymmetry of the distribution around the mean.
  - Fourth Moment (related to Kurtosis): relates to the "tailedness" of the distribution.

# Moments of a Random Variable

Moments provide information about the shape and spread of the distribution of $X$. The $k$-th moment of a random variable $X$ is defined as the expectation of $X^k$, denoted $E[X^k]$:

$$E[X^k] = \int_{-\infty}^{\infty} x^k \, p(x) \, dx$$

**Moments:**

- **First Moment (Mean):** $E[X]$
- **Second Moment:** $E[X^2]$. (captures both the mean and the spread of $X$ around the mean, related to variance).

  *Do not confuse with Variance: Defined as* $Var(X) = E[X^2] - (E[X])^2$
- **Higher-Order Moments:**
  - Third Moment (related to Skewness): relates to the asymmetry of the distribution around the mean.
  - Fourth Moment (related to Kurtosis): relates to the "tailedness" of the distribution.

# Moments of a Random Variable

Moments provide information about the shape and spread of the distribution of $X$. The $k$-th moment of a random variable $X$ is defined as the expectation of $X^k$, denoted $E[X^k]$:

$$E[X^k] = \int_{-\infty}^{\infty} x^k \, p(x) \, dx$$

**Moments:**

- **First Moment (Mean):** $E[X]$
- **Second Moment:** $E[X^2]$. (captures both the mean and the spread of $X$ around the mean, related to variance).

  *Do not confuse with Variance: Defined as* $Var(X) = E[X^2] - (E[X])^2$
- **Higher-Order Moments:**
  - Third Moment (related to Skewness): relates to the asymmetry of the distribution around the mean.
  - Fourth Moment (related to Kurtosis): relates to the "tailedness" of the distribution.

# Central Moments

**Central Moments:**

- Moments taken about the mean (rather than about zero) are called central moments. For example, the second central moment is variance:

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

- Central moments provide insight into the variability and shape of the distribution relative to its mean (Typically more useful).
- Connection to Empirical Estimators.

# Variance of a Random Variable

The **Variance** is the second central moment of a random variable. It is an operator denoted $\text{Var}(X)$ defined as:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

**Interpretation:**

- The Variance quantifies the spread or variability of a random variable $X$ around its mean.
- A higher variance indicates greater spread; a variance of zero implies $X$ is a constant.

**Key Properties:**

- **Non-negativity:** $\text{Var}(X) \geq 0$; variance cannot be negative.
- **Scaling:** For a constant $a$
$$\text{Var}(aX) = a^2\text{Var}(X)$$
- **Additivity for Independent Variables:** For independent random variables $X$ and $Y$:
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

# Variance of a Random Variable

The **Variance** is the second central moment of a random variable. It is an operator denoted $\text{Var}(X)$ defined as:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

**Interpretation:**

- The Variance quantifies the spread or variability of a random variable $X$ around its mean.
- A higher variance indicates greater spread; a variance of zero implies $X$ is a constant.

**Key Properties:**

- **Non-negativity:** $\text{Var}(X) \geq 0$; variance cannot be negative.
- **Scaling:** For a constant $a$

$$\text{Var}(aX) = a^2\text{Var}(X)$$

- **Additivity for Independent Variables:** For independent random variables $X$ and $Y$:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

# Variance of a Random Variable

The **Variance** is the second central moment of a random variable. It is an operator denoted $\text{Var}(X)$ defined as:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

**Interpretation:**

- The Variance quantifies the spread or variability of a random variable $X$ around its mean.
- A higher variance indicates greater spread; a variance of zero implies $X$ is a constant.

**Key Properties:**

- **Non-negativity:** $\text{Var}(X) \geq 0$; variance cannot be negative.
- **Scaling:** For a constant $a$

$$\text{Var}(aX) = a^2\text{Var}(X)$$

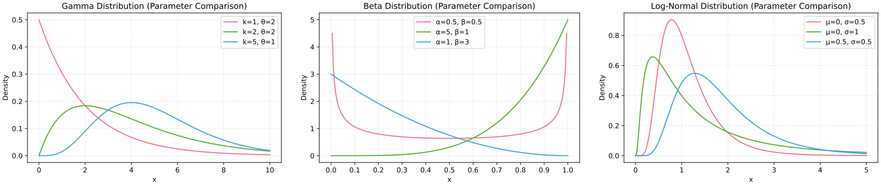
- **Additivity for Independent Variables:** For independent random variables $X$ and $Y$:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

# Parametrization of Distributions

# Analytical Parametrization of Probability Distributions

- Many probability distributions can be expressed analytically as a function of specific parameters, often denoted as $\theta$.
- These parameters $\theta$ control the form of the PDF (they are sometimes chosen to intuitively link them with location, shape, spread of the distribution, but **they do not have to be!**)
- In general, we will indicate a pdf with $p_\theta(X)$ to emphasize that it is a function of the form $f(x, \theta)$.



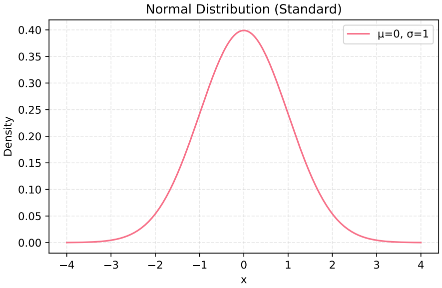
# Common Distributions in Biodata

# Normal Distribution

- **PDF:**

$$p_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- **Key Properties:**
  - Symmetric around mean $\mu$
  - 68% of data within $1\sigma$
  - 95% within $2\sigma$
  - Sum of normal RVs is normal
- **Biological Examples:**
  - Height distributions in populations
  - Measurement errors in experiments
  - Gene expression in large populations

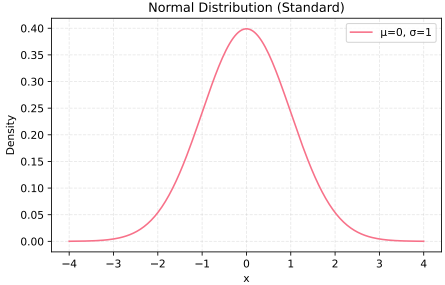


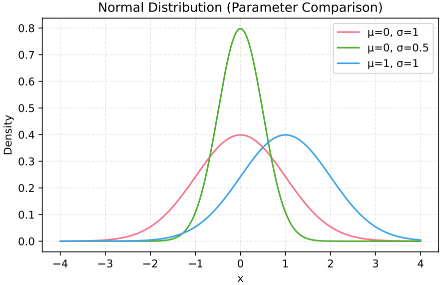
# Normal Distribution

- **PDF:**

$$p_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- **Key Properties:**
  - Symmetric around mean $\mu$
  - 68% of data within $1\sigma$
  - 95% within $2\sigma$
  - Sum of normal RVs is normal
- **Biological Examples:**
  - Height distributions in populations
  - Measurement errors in experiments
  - Gene expression in large populations



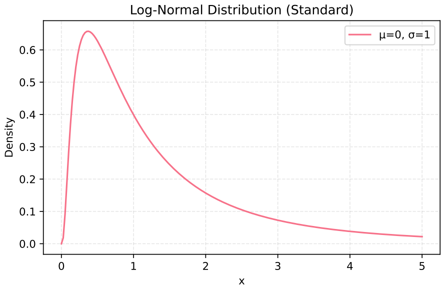



# Log-Normal Distribution

- If $Y = \ln(X)$ is normal, then $X$ is log-normal
- **PDF:**

$$p_{\mu,\sigma^2}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

- **Properties:**
  - Right-skewed
  - Only positive values
  - Multiplicative effects become additive on log scale
- **Biological Applications:**
  - RNA expression levels (normalized)
  - Protein concentrations
  - Species abundance distributions

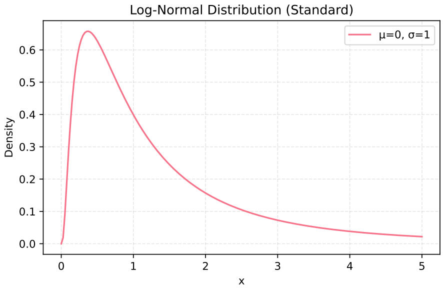


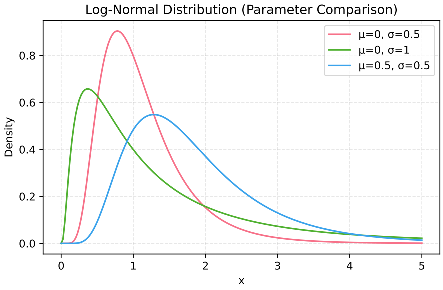
# Log-Normal Distribution

- If $Y = \ln(X)$ is normal, then $X$ is log-normal
- **PDF:**

$$p_{\mu,\sigma^2}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

- **Properties:**
  - Right-skewed
  - Only positive values
  - Multiplicative effects become additive on log scale
- **Biological Applications:**
  - RNA expression levels (normalized)
  - Protein concentrations
  - Species abundance distributions



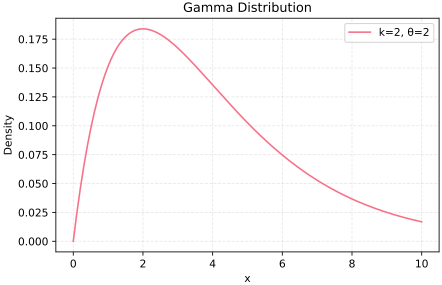
# Gamma Distribution

- **PDF:**

$$p_{k,\theta}(x) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k\Gamma(k)}$$

where $k$ is shape, $\theta$ is scale, and $\Gamma(k)$ is the gamma function

- **Properties:**
  - Always positive
  - Flexible shape (can be exponential when $k = 1$)
  - Sum of exponential RVs follows gamma
- **Biological Applications:**
  - Waiting times between cellular events
  - Duration of biological processes
  - Gene expression burst sizes



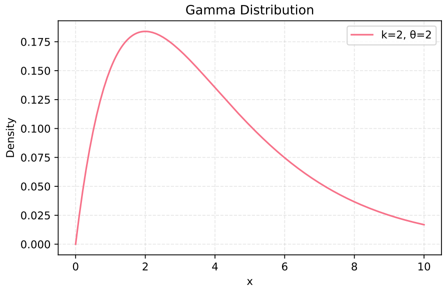
# Gamma Distribution

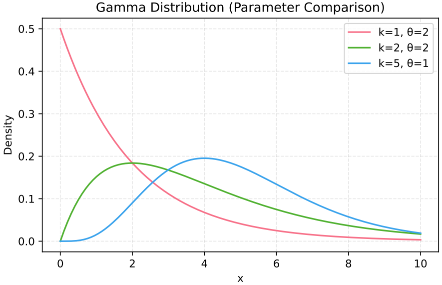
- **PDF:**

$$p_{k,\theta}(x) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k \Gamma(k)}$$

  where $k$ is shape, $\theta$ is scale, and $\Gamma(k)$ is the gamma function

- **Properties:**
  - Always positive
  - Flexible shape (can be exponential when $k = 1$)
  - Sum of exponential RVs follows gamma
- **Biological Applications:**
  - Waiting times between cellular events
  - Duration of biological processes
  - Gene expression burst sizes



# Beta Distribution

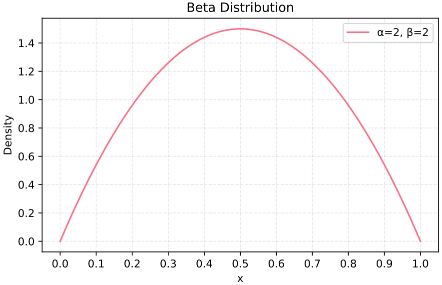
- **PDF:**

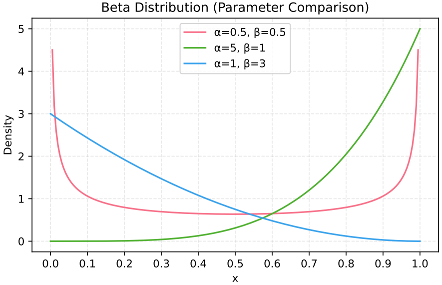
$$p_{\alpha,\beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

  where $B(\alpha,\beta)$ is the beta function

- **Properties:**
  - Defined on [0,1]
  - Conjugate prior to Bernoulli
  - Flexible shapes based on $\alpha, \beta$

- **Biological Applications:**
  - Allele frequencies
  - Gene expression proportions
  - Success probabilities in experiments



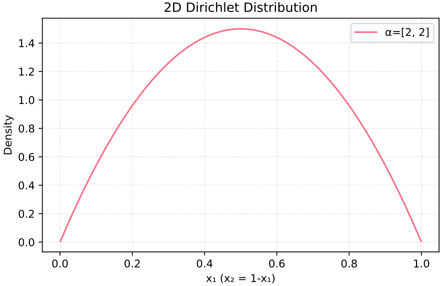
# Dirichlet Distribution

- **PDF:**

$$p_\alpha(x) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

  where $\sum x_i = 1$ and $B(\alpha)$ is the multivariate beta function

- **Properties:**
  - Multivariate generalization of Beta
  - Each component in (0,1)
  - Components sum to 1
- **Biological Applications:**
  - Cell type proportions
  - Species composition in microbiome
  - Multi-allele frequencies



# Dirichlet Distribution

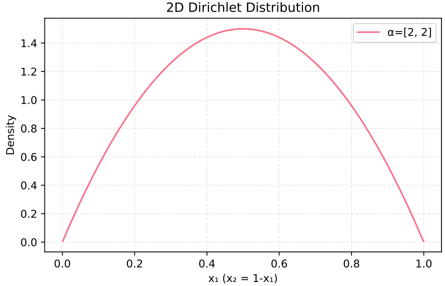
- **PDF:**

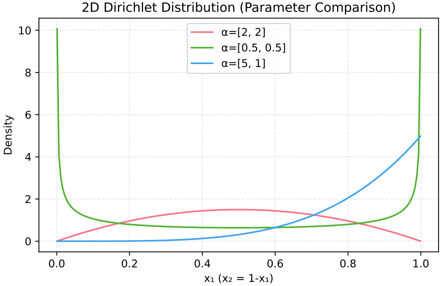
$$p_\alpha(x) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

  where $\sum x_i = 1$ and $B(\alpha)$ is the multivariate beta function

- **Properties:**
  - Multivariate generalization of Beta
  - Each component in (0,1)
  - Components sum to 1

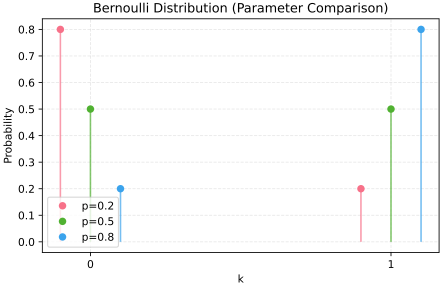
- **Biological Applications:**
  - Cell type proportions
  - Species composition in microbiome
  - Multi-allele frequencies



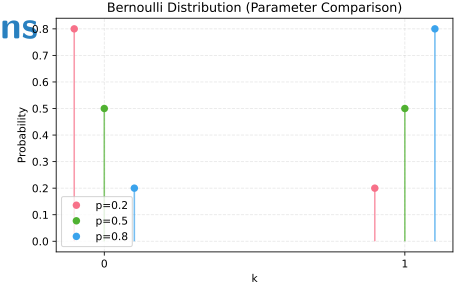
# Bernoulli and Binomial Distributions

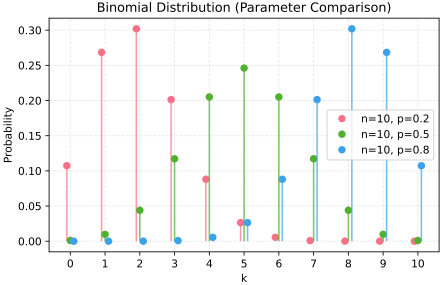
- **Bernoulli** (single trial):
  - PMF: $P(X = k) = p^k(1-p)^{1-k}$, $k \in \{0, 1\}$
  - Mean: $p$, Variance: $p(1-p)$
  - *Example:* Single mutation event
- **Binomial** ($n$ independent trials):
  - PMF: $P(X = k) = \binom{n}{k} p^k(1-p)^{n-k}$
  - Mean: $np$, Variance: $np(1-p)$
  - *Examples:*
    - Number of mutated cells in a population
    - Success count in $n$ identical experiments



# Bernoulli and Binomial Distributions

- **Bernoulli** (single trial):
  - PMF: $P(X = k) = p^k(1-p)^{1-k}$, $k \in \{0, 1\}$
  - Mean: $p$, Variance: $p(1-p)$
  - *Example:* Single mutation event
- **Binomial** ($n$ independent trials):
  - PMF: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - Mean: $np$, Variance: $np(1-p)$
  - *Examples:*
    - Number of mutated cells in a population
    - Success count in $n$ identical experiments



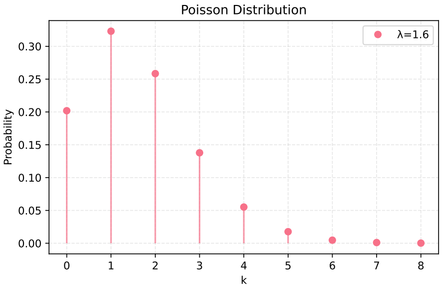
# Poisson Distribution

- **PMF:**

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

  where $\lambda$ is both mean and variance

- **Properties:**
  - Models rare events
  - Sum of Poisson RVs is Poisson
  - Limit of Binomial as $n \to \infty$, $p \to 0$
- **Biological Applications:**
  - Mutations in DNA sequences
  - Number of cells in a small sample
  - RNA-seq read counts



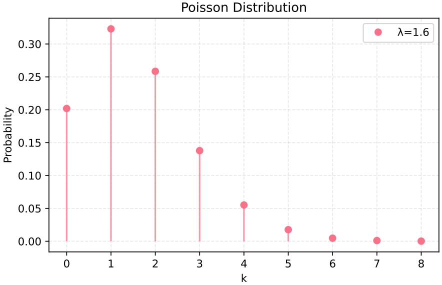
# Poisson Distribution

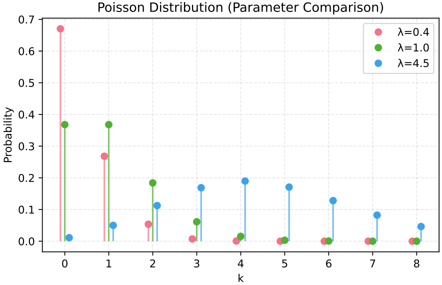
- **PMF:**

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

where $\lambda$ is both mean and variance

- **Properties:**
  - Models rare events
  - Sum of Poisson RVs is Poisson
  - Limit of Binomial as $n \to \infty$, $p \to 0$
- **Biological Applications:**
  - Mutations in DNA sequences
  - Number of cells in a small sample
  - RNA-seq read counts



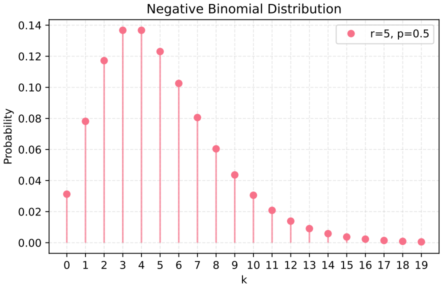
# Negative Binomial Distribution

- **PMF:**

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

where $r$ is number of successes, $p$ is probability

- **Properties:**
  - More dispersed than Poisson
  - Mean: $\frac{pr}{1-p}$, Variance: $\frac{pr}{(1-p)^2}$
  - Handles overdispersion in count data
- **Biological Applications:**
  - RNA-seq differential expression
  - Microbial abundance counts
  - Cellular heterogeneity modeling



# Negative Binomial Distribution

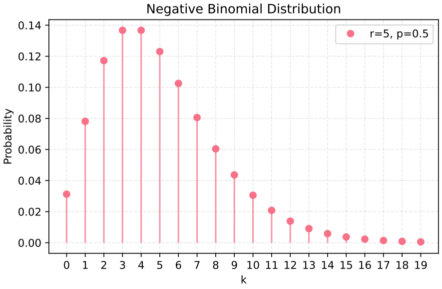
- **PMF:**

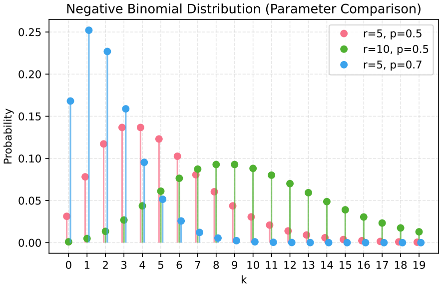
$$P(X = k) = \binom{k + r - 1}{k} p^r (1 - p)^k$$

where $r$ is number of successes, $p$ is probability

- **Properties:**
  - More dispersed than Poisson
  - Mean: $\frac{pr}{1-p}$, Variance: $\frac{pr}{(1-p)^2}$
  - Handles overdispersion in count data

- **Biological Applications:**
  - RNA-seq differential expression
  - Microbial abundance counts
  - Cellular heterogeneity modeling



# NegBinom - Alternative Param.

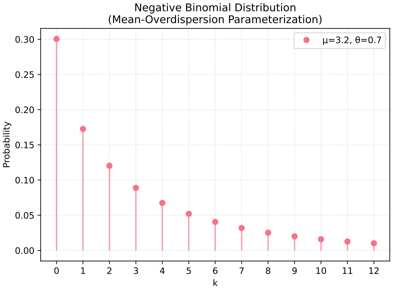
- The NB distribution can be alternatively parameterized in terms of a mean $\mu$ and the parameter $\theta$.

$$P(X = k) = \binom{k+\theta-1}{k} \left(\frac{\theta}{\theta+\mu}\right)^{\theta} \left(\frac{\mu}{\theta+\mu}\right)^{k}$$

$$E[X] = \mu, \quad \text{Var}(X) = \mu + \frac{\mu^2}{\theta}$$

- The Poisson distribution is a special case when $(\theta \to \infty)$. So one can think about $r = \frac{1}{\theta}$ as an the overdispersion parameter

  *Example:* RNA-seq read counts



# NegBinom - Alternative Param.

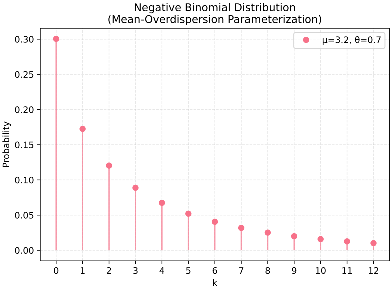
- The NB distribution can be alternatively parameterized in terms of a mean $\mu$ and the parameter $\theta$.

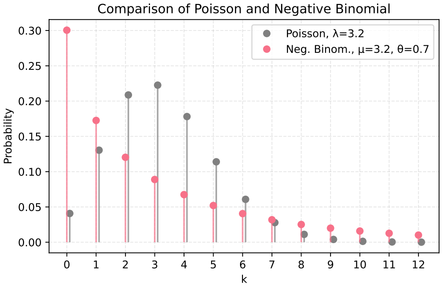
$$P(X = k) = \binom{k + \theta - 1}{k} \left(\frac{\theta}{\theta + \mu}\right)^{\theta} \left(\frac{\mu}{\theta + \mu}\right)^{k}$$

$$E[X] = \mu, \quad \mathrm{Var}(X) = \mu + \frac{\mu^2}{\theta}$$

- The Poisson distribution is a special case when $(\theta \to \infty)$. So one can think about $r = \frac{1}{\theta}$ as an the overdispersion parameter

  *Example:* RNA-seq read counts



# Operations on Random Variables

# Operations on Random Variables: Introduction

- **Fundamental Operations:**
  - **Convolution:** Summing independent random variables
  - **Mixture Distributions:** Random selection from multiple distributions
  - **Transformations:** Applying functions to random variables
- **Why It Matters:**
  - Essential for modeling combined effects
  - Critical in statistical inference and data analysis
- **Applications:**
  - Signal processing
  - Financial risk assessment
  - Biological systems modeling

# Convolution of Random Variables

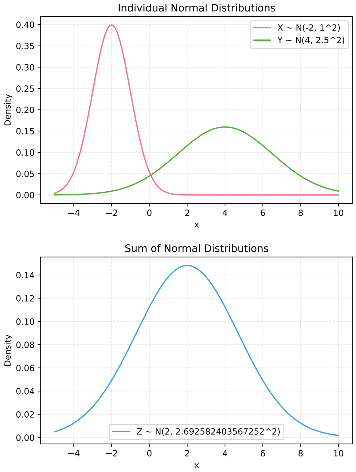
- **Definition:**
  - The convolution of two independent random variables $X$ and $Y$ results in a new random variable $Z = X + Y$.
  - The PDF of $Z$ is given by:

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

- **Interpretation:**
  - This integral accounts for all combinations of $x$ and $y$ that sum to $z$, weighted by their probabilities.
- **Properties:**
  - **Mean:** $E[Z] = E[X] + E[Y]$
  - **Variance:** $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$ (if $X$ and $Y$ are independent)
  - **Distribution Shape:** The shape of $f_Z(z)$ may be smoother and differ from $f_X(x)$ and $f_Y(y)$

# Convolution of Normals

- **Examples:**
  - **Normal Variables:** $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ imply $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- **Key Point:** Convolution models the *aggregate effect* of independent random variables
- **Common Mistake:** Assuming the convolution of two distributions is the same type as the original distributions

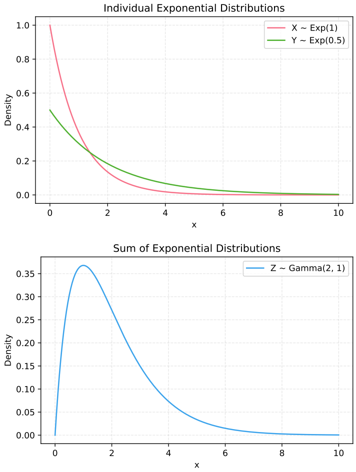


# Convolution of Exponential Variables

- **Examples:**
  - **Exponential Variables:** $X_i \sim \text{Exp}(\lambda)$, the sum $Z = \sum_{i=1}^{k} X_i$ follows a Gamma distribution with shape $k$ and rate $\lambda$:

$$Z \sim \text{Gamma}(k, \lambda)$$

- **Key Point:** Convolution of exponential variables models the total waiting time for $k$ events in a Poisson process



# Mixture Distributions

- **Definition:** A mixture distribution arises when a random variable $X$ is drawn from one of several distributions $f_i(x)$ with probabilities $w_i$ (where $\sum_i w_i = 1$):

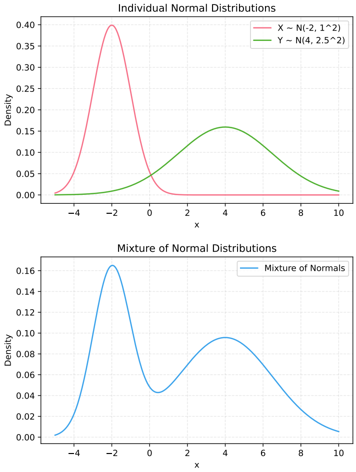
$$f_X(x) = \sum_i w_i f_i(x)$$

- **Properties:**
  - **Mean:** $E[X] = \sum_i w_i E[X_i]$
  - **Variance:** $\text{Var}(X) = \sum_i w_i \left(\text{Var}(X_i) + [E[X_i] - E[X]]^2\right)$
  - **Distribution Shape:** Can be multimodal
- **Contrast with Convolution:**
  - **Mixture:** Randomly selects one distribution to sample from
  - **Convolution:** Sums values from all distributions
- **Common Mistake:** Confusing mixtures with convolutions

# Mixture Distributions: Examples

- **Examples:**
  - Consider a mixture of two normal distributions: $N(-2, 1^2)$ and $N(4, 2.5^2)$ with weights $w = 0.4$ and $1 - w = 0.6$
  - **Step 1:** Draw a Bernoulli random variable $B \sim \text{Bernoulli}(0.4)$
  - **Step 2:** If $B = 0$, draw from $N(-2, 1^2)$; if $B = 1$, draw from $N(4, 2.5^2)$
  - **Notation:**

$$X \sim \begin{cases} N(-2, 1^2) & \text{if } B = 0 \\ N(4, 2.5^2) & \text{if } B = 1 \end{cases}$$



# Transformation of Random Variables

- **Definition:** If $X$ is a random variable with PDF $f_X(x)$ and $Y = g(X)$ is a function of $X$, then the PDF of $Y$ is:

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

  where $g^{-1}(y)$ is the inverse function of $g(x)$.
- **Properties:**
  - **Change of Variables:** Requires Jacobian adjustment
  - **Moments:** Moments of $Y$ depend on $g(x)$ and moments of $X$
- **Examples:**
  - **Log Transformation:** $Y = \ln(X)$, useful for skewed data
  - **Exponential Transformation:** $Y = e^X$, converts normal to log-normal
- **Key Difference:**
  - Transformation modifies a single variable's distribution
  - Unlike convolution and mixtures, no combining of variables

# Summary and Key Differences

- **Convolution:**
  - Sum of independent variables
  - Models cumulative effects
  - Requires convolution integral
- **Mixture Distributions:**
  - Random selection among distributions
  - Models heterogeneous populations
  - PDF is a weighted sum
- **Transformations:**
  - Function applied to a variable
  - Alters the shape of the distribution
  - Requires change of variables formula
- **Final Thought:** Understanding these differences is crucial for accurate statistical modeling