Selected Topics in Mathematics of Learning

High-Dimensional Statistics

Lecturer: Marius Yamakou

Winter Semester 2024/25
Department of Data Science, FAU

October 15, 2024

General motivation and perspectives

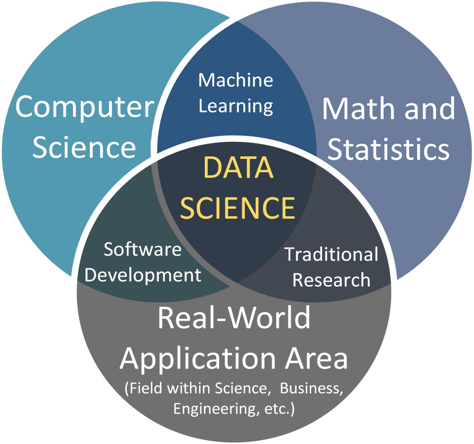


Figure: https://www.usu.edu/math/datascience/

Syllabus

- Review: Probability and Statistics
- Motivation: Why high-dimensional statistics?
- Concentration inequalities
- Sparse linear models
- Random matrices and covariance estimation
- Covariance estimation and thresholding
- Inverse Covariance estimation
- Principal component analysis in high dimensions
- Reproducing kernel Hilbert spaces
- Review Session

Syllabus

- Review: Probability and Statistics
- Motivation: Why high-dimensional statistics?
- Concentration inequalities
- Sparse linear models
- Random matrices and covariance estimation
- Covariance estimation and thresholding
- Inverse Covariance estimation
- Principal component analysis in high dimensions
- Reproducing kernel Hilbert spaces
- Review Session

Relevant Literature

- MW: High-Dimensional Statistics: A Non-Asymptotic Viewpoint, by Martin J. Wainwright
- RV: High-Dimensional Probability, by Roman Vershynin
- BG: Statistics for High-Dimensional Data: Methods, Theory and Applications, by Peter Bühlmann and Sara van de Geer
- BL: Covariance regularization by thresholding, by Peter Bickel and Elizaveta Levina
- RBLZ: Sparse permutation invariant covariance estimation, by Adam Rothmann, Peter Bickel, Elizaveta Levina, and Ji Zhu

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.
- Highlight the role of combinatorics in probability.
- Differentiate between independent and dependent events.

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.
- Highlight the role of combinatorics in probability.
- Differentiate between independent and dependent events.
- Cover key discrete and continuous distributions.
- Review statistical measures like mean, variance, and standard deviation.

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.
- Highlight the role of combinatorics in probability.
- Differentiate between independent and dependent events.
- Cover key discrete and continuous distributions.
- Review statistical measures like mean, variance, and standard deviation.
- Introduce estimators and their properties, such as bias and consistency.

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.
- Highlight the role of combinatorics in probability.
- Differentiate between independent and dependent events.
- Cover key discrete and continuous distributions.
- Review statistical measures like mean, variance, and standard deviation.
- Introduce estimators and their properties, such as bias and consistency.
- Explain different types of convergence (in probability and in distribution).
- Discuss the law of large numbers and the central limit theorem.

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.
- Highlight the role of combinatorics in probability.
- Differentiate between independent and dependent events.
- Cover key discrete and continuous distributions.
- Review statistical measures like mean, variance, and standard deviation.
- Introduce estimators and their properties, such as bias and consistency.
- Explain different types of convergence (in probability and in distribution).
- Discuss the law of large numbers and the central limit theorem.
- Cover basic elements of linear algebra relevant to statistics.
- Introduce the multivariate normal distribution and its properties.

Part 0

Review of some basic concepts in probability and statistics

Objectives:

- Recall basic concepts in probability theory, including measurable functions, random variables, and distributions.
- Explain probability mass functions and probability density functions.
- Highlight the role of combinatorics in probability.
- Differentiate between independent and dependent events.
- Cover key discrete and continuous distributions.
- Review statistical measures like mean, variance, and standard deviation.
- Introduce estimators and their properties, such as bias and consistency.
- Explain different types of convergence (in probability and in distribution).
- Discuss the law of large numbers and the central limit theorem.
- Cover basic elements of linear algebra relevant to statistics.
- Introduce the multivariate normal distribution and its properties.
- Prepare students for advanced topics in statistics and machine learning.

Outline

1. Basics: Probability space

An experiment refers to any process that can be repeated under the same conditions and leads to a well-defined set of possible outcomes. Each performance of an experiment is called a **trial**, and the possible results from each trial, the **outcomes**. An experiment can be deterministic and follow a regular pattern (boring!) or random and thus difficult to predict.

1. Basics: Probability space

An experiment refers to any process that can be repeated under the same conditions and leads to a well-defined set of possible outcomes. Each performance of an experiment is called a **trial**, and the possible results from each trial, the **outcomes**. An experiment can be deterministic and follow a regular pattern (boring!) or random and thus difficult to predict.

The principal concept of probability theory is the **Probability Space**, which can be denoted mathematically as an ordered triple $(\Omega, \mathcal{F}, P)$, representing a structured framework for modeling uncertainty.

1. Basics: Probability space

An experiment refers to any process that can be repeated under the same conditions and leads to a well-defined set of possible outcomes. Each performance of an experiment is called a **trial**, and the possible results from each trial, the **outcomes**. An experiment can be deterministic and follow a regular pattern (boring!) or random and thus difficult to predict.

The principal concept of probability theory is the **Probability Space**, which can be denoted mathematically as an ordered triple $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \boldsymbol{P})$, representing a structured framework for modeling uncertainty.

1. **Sample Space $\boldsymbol{\Omega}$:** The set of all possible outcomes of a random experiment. *Examples:* For a single coin toss, $\Omega = \{$Head(H), Tail(T)$\}$. For tossing two coins, $\Omega = \{$HH, HT, TH, TT$\}$.

1. Basics: Probability space

An experiment refers to any process that can be repeated under the same conditions and leads to a well-defined set of possible outcomes. Each performance of an experiment is called a **trial**, and the possible results from each trial, the **outcomes**. An experiment can be deterministic and follow a regular pattern (boring!) or random and thus difficult to predict.

The principal concept of probability theory is the **Probability Space**, which can be denoted mathematically as an ordered triple $\boldsymbol{(\Omega, \mathcal{F}, P)}$, representing a structured framework for modeling uncertainty.

1. **Sample Space $\boldsymbol{\Omega}$:** The set of all possible outcomes of a random experiment. *Examples:* For a single coin toss, $\Omega = \{$Head(H), Tail(T)$\}$. For tossing two coins, $\Omega = \{$HH, HT, TH, TT$\}$.

2. **Sigma-Algebra $\boldsymbol{\mathcal{F}}$:** A collection of subsets of Ω that includes the sample space itself and is closed under complementation and countable unions. *Examples:* For a single coin toss, $\mathcal{F} = \{\emptyset, \Omega, \{\text{H}\}, \{\text{T}\}\}$. For tossing two coins, $\mathcal{F} = \{\emptyset, \Omega, \{\text{HH}\}, \{\text{HT}\}, \{\text{TH}\}, \{\text{TT}\}, \{\text{H}\}, \{\text{T}\}\}$.

1. Basics: Probability space

3. **Probability Function P:** A function that assigns a probability to each event in $\mathcal{F}$, satisfying:

- $P(A) \geq 0$ for all $A \in \mathcal{F}$
- $P(\Omega) = 1$
- For disjoint events A and B: $P(A \cup B) = P(A) + P(B)$

1. Basics: Probability space

3. **Probability Function P:** A function that assigns a probability to each event in $\mathcal{F}$, satisfying:

- $P(A) \geq 0$ for all $A \in \mathcal{F}$
- $P(\Omega) = 1$
- For disjoint events A and B: $P(A \cup B) = P(A) + P(B)$

Examples

1. **Single Die Toss**:
- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Let $A = \{\text{even outcomes}\} = \{2, 4, 6\}$
- $P(A) = \frac{3}{6} = \frac{1}{2}$

1. Basics: Probability space

3. **Probability Function P:** A function that assigns a probability to each event in $\mathcal{F}$, satisfying:

- $P(A) \geq 0$ for all $A \in \mathcal{F}$
- $P(\Omega) = 1$
- For disjoint events A and B: $P(A \cup B) = P(A) + P(B)$

disjoint are never occur at the same time or mutually excluisve

Examples

1. **Single Die Toss**:
- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Let $A = \{\text{even outcomes}\} = \{2, 4, 6\}$
- $P(A) = \frac{3}{6} = \frac{1}{2}$

2. **Two Coin Tosses**:
- $\Omega = \{\text{HH, HT, TH, TT}\}$
- Let $B = \{\text{at least one Heads}\} = \{\text{HH, HT, TH}\}$
- $P(B) = \frac{3}{4}$

1. Basics: Measurable functions and random variables

A Borel Set: A Borel set $\mathcal{B}$ is any set that can be formed from open intervals (or open sets) through the operations of countable unions, countable intersections, and relative complements.

1. Basics: Measurable functions and random variables

A Borel Set: A Borel set $\mathcal{B}$ is any set that can be formed from open intervals (or open sets) through the operations of countable unions, countable intersections, and relative complements.

Measurable Function: A function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable if for every Borel set $B \in \mathcal{B}$ we have $X^{-1}(B) \in \mathcal{F}$. This means that the pre-image of any Borel set under X must be a measurable set in the sigma-algebra $\mathcal{F}$.

1. Basics: Measurable functions and random variables

A Borel Set: A Borel set $\mathcal{B}$ is any set that can be formed from open intervals (or open sets) through the operations of countable unions, countable intersections, and relative complements.

Measurable Function: A function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable if for every Borel set $B \in \mathcal{B}$ we have $X^{-1}(B) \in \mathcal{F}$. This means that the pre-image of any Borel set under X must be a measurable set in the sigma-algebra $\mathcal{F}$.

Borel σ-Algebra: The collection of all Borel sets forms a σ-algebra known as the Borel σ-algebra, denoted by $\mathcal{B}$. This σ-algebra is crucial for defining measures, such as the Borel measure, which is used to assign probabilities to events in probability theory.

1. Basics: Measurable functions and random variables

A Borel Set: A Borel set $\mathcal{B}$ is any set that can be formed from open intervals (or open sets) through the operations of countable unions, countable intersections, and relative complements.

Measurable Function: A function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable if for every Borel set $B \in \mathcal{B}$ we have $X^{-1}(B) \in \mathcal{F}$. This means that the pre-image of any Borel set under X must be a measurable set in the sigma-algebra $\mathcal{F}$.

Borel σ-Algebra: The collection of all Borel sets forms a σ-algebra known as the Borel σ-algebra, denoted by $\mathcal{B}$. This σ-algebra is crucial for defining measures, such as the Borel measure, which is used to assign probabilities to events in probability theory.

In probability theory, Borel sets are significant because they provide a way to define events on the real line (or in higher dimensions) mathematically rigorously.

1. Basics: Measurable functions and random variables

Random Variable: A random variable is a measurable function that maps outcomes from a sample space Ω to the real unit interval $[0, 1]$ (for probability measures):

$$X : \Omega \to [0, 1]$$

where X represents a quantitative outcome of a random experiment.

1. Basics: Measurable functions and random variables

Random Variable: A random variable is a measurable function that maps outcomes from a sample space Ω to the real unit interval $[0, 1]$ (for probability measures):

$$X : \Omega \to [0, 1]$$

where X represents a quantitative outcome of a random experiment.

Discrete Random Variable: A discrete random variable is one where the set of possible outcomes is discrete (i.e., countable).

1. Basics: Measurable functions and random variables

Random Variable: A random variable is a measurable function that maps outcomes from a sample space Ω to the real unit interval $[0, 1]$ (for probability measures):

$$X : \Omega \to [0, 1]$$

where X represents a quantitative outcome of a random experiment.

Discrete Random Variable: A discrete random variable is one where the set of possible outcomes is discrete (i.e., countable).

Example: Consider a random experiment where we roll a fair six-sided die:
- Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the associated sigma-algebra $\mathcal{F}$.
- Define the random variable X that maps the outcomes to probabilities:

$$X(i) = \frac{i}{6} \quad \text{for } i \in \Omega$$

Thus, X transforms the outcome of the die roll into its probability representation in the unit interval $[0, 1]$.

2. Discrete distributions and first two moments

Probability Mass Function: The probability mass function (PMF) of a discrete random variable X, denoted by p_X, is defined as $p_X(x) = P(X = x)$, where $P(X = x)$ represents the probability that X takes the value x in the probability space $(\Omega, \mathcal{F}, P)$.

2. Discrete distributions and first two moments

Probability Mass Function: The probability mass function (PMF) of a discrete random variable X, denoted by p_X, is defined as $p_X(x) = P(X = x)$, where $P(X = x)$ represents the probability that X takes the value x in the probability space $(\Omega, \mathcal{F}, P)$.

Expected Value: Given a discrete random variable X that takes values in a set $A = \{x_1, x_2, x_3, \dots\}$, the expected value of X is denoted $\mathbb{E}(X)$ or μ_X. It is calculated by multiplying each possible value of X by its probability:

$$\mathbb{E}(X) = \sum_{x \in A} x \cdot P(X = x) = \sum_{x \in A} x \cdot p_X(x).$$

2. Discrete distributions and first two moments

Probability Mass Function: The probability mass function (PMF) of a discrete random variable X, denoted by p_X, is defined as $p_X(x) = P(X = x)$, where $P(X = x)$ represents the probability that X takes the value x in the probability space $(\Omega, \mathcal{F}, P)$.

Expected Value: Given a discrete random variable X that takes values in a set $A = \{x_1, x_2, x_3, \dots\}$, the expected value of X is denoted $\mathbb{E}(X)$ or μ_X. It is calculated by multiplying each possible value of X by its probability:

$$\mathbb{E}(X) = \sum_{x \in A} x \cdot P(X = x) = \sum_{x \in A} x \cdot p_X(x).$$

The expected value of $g(X)$ is calculated as:

$$\mathbb{E}[g(X)] = \sum_{x \in A} g(x) \cdot P(X = x) = \sum_{x \in A} g(x) \cdot p_X(x).$$

2. Discrete distributions and first two moments

Variance: The variance of a discrete random variable X measures the spread or dispersion of its values around the expected value. It is denoted by **Var(X)** and is calculated as:

$$Var(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right],$$

where $\mathbb{E}(X)$ is the expected value of X. Variance represents the average of the squared differences between X and its mean, $\mathbb{E}(X)$.

2. Discrete distributions and first two moments

Variance: The variance of a discrete random variable X measures the spread or dispersion of its values around the expected value. It is denoted by **Var(X)** and is calculated as:

$$Var(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right],$$

where $\mathbb{E}(X)$ is the expected value of X. Variance represents the average of the squared differences between X and its mean, $\mathbb{E}(X)$.

Standard Deviation: The standard deviation of X, denoted σ_X, is the square root of the variance:

$$\sigma_X = \sqrt{Var(X)}.$$

Standard deviation provides a measure of dispersion in the same units as the random variable X, making it easier to interpret than the variance.

2. Discrete distributions: Bernoulli (p)

Bernoulli Distribution: A random variable X has a Bernoulli distribution with parameter p, denoted $X \sim \text{Bernoulli}(p)$, if its PMF is given by:

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

2. Discrete distributions: Bernoulli (p)

Bernoulli Distribution: A random variable X has a Bernoulli distribution with parameter p, denoted $X \sim \text{Bernoulli}(p)$, if its PMF is given by:

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

In other words, we are saying X is a binary random variable with support $\mathcal{F} = \{0, 1\}$, parameter space $\Omega = \{p | 0 < p < 1\}$, and PMF $f(x) = p^x(1-p)^{1-x}$, $x \in \mathcal{F}$.

2. Discrete distributions: Bernoulli (p)

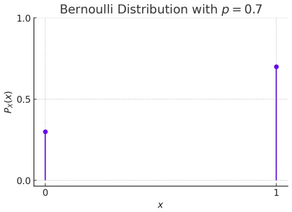
Bernoulli Distribution: A random variable X has a Bernoulli distribution with parameter p, denoted $X \sim$ Bernoulli(p), if its PMF is given by:

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

In other words, we are saying X is a binary random variable with support $\mathcal{F} = \{0, 1\}$, parameter space $\Omega = \{p | 0 < p < 1\}$, and PMF $f(x) = p^x(1-p)^{1-x}$, $x \in \mathcal{F}$.

Properties:

- $\mathbb{E}(X) = \sum_{x \in \mathcal{F} = \{0,1\}} x \cdot P(X = x)$
  $= 0 \cdot (1-p) + 1 \cdot (p) = p$
- $[\mathbb{E}(X)]^2 = p^2$
- $\mathbb{E}(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot (p) = p$
- $Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$
  $= p(1-p)$



2. Discrete distributions: Binomial (n, p)

Binomial Distribution: A random variable X has a binomial distribution with parameters n and p, denoted $X \sim \text{Binomial}(n, p)$, if its PMF is given by:

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, 2, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient and the factorial operator ! is defined as $n! := n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1$ with $0! = 1$.

2. Discrete distributions: Binomial (n, p)

Binomial Distribution: A random variable X has a binomial distribution with parameters n and p, denoted $X \sim \text{Binomial}(n, p)$, if its PMF is given by:

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, 2, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the binomial coefficient and the factorial operator ! is defined as $n! := n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1$ with $0! = 1$.

In other words, the random variable X represents the number of successes in n independent Bernoulli trials with success probability p. In this case, the support is $\mathcal{F} = \{0, 1, \ldots, n\}$, and the parameter space is $\Omega = \{(n, p) | n \in \mathbb{Z}^+, 0 < p < 1\}$.

2. Discrete distributions: Binomial (n, p)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{n} x \cdot P(X = x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$

2. Discrete distributions: Binomial (n, p)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{n} x \cdot P(X = x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$ Using the identity:

  $x \cdot \binom{n}{x} = n \cdot \binom{n-1}{x-1}$, we get: $\mathbb{E}(X) = \sum_{x=1}^{n} n \cdot \binom{n-1}{x-1} p^x (1-p)^{n-x}$

2. Discrete distributions: Binomial (n, p)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{n} x \cdot P(X = x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$ Using the identity: $x \cdot \binom{n}{x} = n \cdot \binom{n-1}{x-1}$, we get: $\mathbb{E}(X) = \sum_{x=1}^{n} n \cdot \binom{n-1}{x-1} p^x (1-p)^{n-x}$

  Changing the index of summation to $k = x - 1$, we get

$$\mathbb{E}(X) = n \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{(n-1)-k}$$

  This can be expressed as: $\mathbb{E}(X) = n \cdot p \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}$

2. Discrete distributions: Binomial (n, p)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{n} x \cdot P(X = x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$ Using the identity:

  $x \cdot \binom{n}{x} = n \cdot \binom{n-1}{x-1}$, we get: $\mathbb{E}(X) = \sum_{x=1}^{n} n \cdot \binom{n-1}{x-1} p^x (1-p)^{n-x}$

  Changing the index of summation to $k = x - 1$, we get

  $$\mathbb{E}(X) = n \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{(n-1)-k}$$

  This can be expressed as: $\mathbb{E}(X) = n \cdot p \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}$

  Using the binomial theorem, we have:

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

  Thus, we conclude: $\mathbb{E}(X) = n \cdot p \cdot 1 = n \cdot p$

2. Discrete distributions: Binomial (n, p)

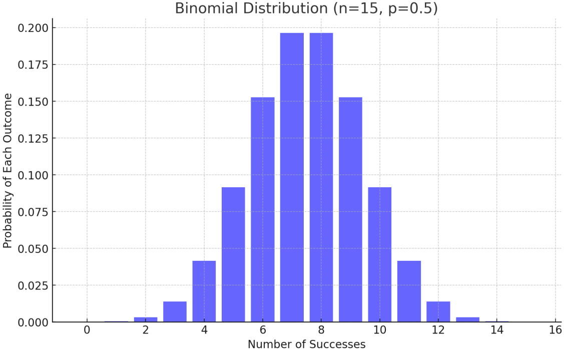
Properties:

- $\mathbb{E}(X) = n \cdot p$
- $\mathbb{E}(X^2) \stackrel{?}{=} n \cdot p \cdot (1-p) + (n \cdot p)^2$
- $Var(X) \stackrel{?}{=} n \cdot p \cdot (1-p)$

2. Discrete distributions: Binomial (n, p)

Properties:

- $\mathbb{E}(X) = n \cdot p$
- $\mathbb{E}(X^2) \overset{?}{=} n \cdot p \cdot (1-p) + (n \cdot p)^2$
- $Var(X) \overset{?}{=} n \cdot p \cdot (1-p)$



2. Discrete distributions: Binomial (n, p)

Example: Suppose a coin is flipped 3 times. Let X represent the number of heads (successes) in these 3 flips. Assuming the coin is fair, the probability of getting a head on any flip is $p = 0.5$. Since each flip is independent, X follows a binomial distribution with parameters $n = 3$ (number of flips) and $p = 0.5$ (probability of heads). This is denoted as:

$$X \sim \text{Binomial}(3, 0.5)$$

2. Discrete distributions: Binomial (n, p)

Example: Suppose a coin is flipped 3 times. Let X represent the number of heads (successes) in these 3 flips. Assuming the coin is fair, the probability of getting a head on any flip is $p = 0.5$. Since each flip is independent, X follows a binomial distribution with parameters $n = 3$ (number of flips) and $p = 0.5$ (probability of heads). This is denoted as:

$$X \sim \text{Binomial}(3, 0.5)$$

The probability of getting exactly x heads in 3 flips is given by the binomial PMF:

$$P(X = x) = \binom{3}{x}(0.5)^x(0.5)^{3-x}, \quad x = 0, 1, 2, 3$$

Let's calculate the probabilities for different values of x:

- $P(X = 0) = \binom{3}{0}(0.5)^0(0.5)^3 = 1 \times 0.125 = 0.125$
- $P(X = 1) = \binom{3}{1}(0.5)^1(0.5)^2 = 3 \times 0.125 = 0.375$
- $P(X = 2) = \binom{3}{2}(0.5)^2(0.5)^1 = 3 \times 0.125 = 0.375$
- $P(X = 3) = \binom{3}{3}(0.5)^3(0.5)^0 = 1 \times 0.125 = 0.125$

2. Discrete distributions: Binomial (n, p)

Example: Suppose a coin is flipped 3 times. Let X represent the number of heads (successes) in these 3 flips. Assuming the coin is fair, the probability of getting a head on any flip is $p = 0.5$. Since each flip is independent, X follows a binomial distribution with parameters $n = 3$ (number of flips) and $p = 0.5$ (probability of heads). This is denoted as:

$$X \sim \text{Binomial}(3, 0.5)$$

The probability of getting exactly x heads in 3 flips is given by the binomial PMF:

$$P(X = x) = \binom{3}{x}(0.5)^x(0.5)^{3-x}, \quad x = 0, 1, 2, 3$$

Let's calculate the probabilities for different values of x:

- $P(X = 0) = \binom{3}{0}(0.5)^0(0.5)^3 = 1 \times 0.125 = 0.125$
- $P(X = 1) = \binom{3}{1}(0.5)^1(0.5)^2 = 3 \times 0.125 = 0.375$
- $P(X = 2) = \binom{3}{2}(0.5)^2(0.5)^1 = 3 \times 0.125 = 0.375$
- $P(X = 3) = \binom{3}{3}(0.5)^3(0.5)^0 = 1 \times 0.125 = 0.125$

In this case, the expected value $\mathbb{E}(X)$ and variance $Var(X)$ can be computed as:

$$\mathbb{E}(X) = n \cdot p = 3 \cdot 0.5 = 1.5$$

$$Var(X) = n \cdot p \cdot (1 - p) = 3 \cdot 0.5 \cdot 0.5 = 0.75$$

2. Discrete distributions: Poisson (λ)

Poisson Distribution: A random variable X has a Poisson distribution with parameter λ, denoted $X \sim \text{Poisson}(\lambda)$, if its PMF is given by:

$$p_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{if } x \in \{0, 1, 2, \dots\}, \\ 0 & \text{otherwise.} \end{cases}$$

2. Discrete distributions: Poisson (λ)

Poisson Distribution: A random variable X has a Poisson distribution with parameter λ, denoted $X \sim \text{Poisson}(\lambda)$, if its PMF is given by:

$$p_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{if } x \in \{0, 1, 2, \dots\}, \\ 0 & \text{otherwise.} \end{cases}$$

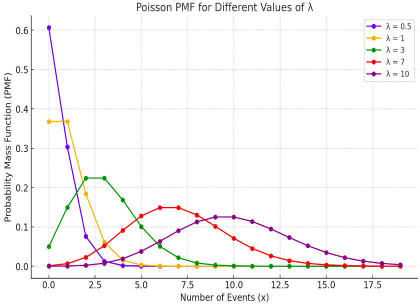
In other words, the Poisson distribution models the probability of a given number of events occurring in a fixed interval of time or space, under the assumption that events occur independently and at a constant rate λ.

2. Discrete distributions: Poisson (λ)

Poisson Distribution: A random variable X has a Poisson distribution with parameter λ, denoted $X \sim \text{Poisson}(\lambda)$, if its PMF is given by:

$$p_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{if } x \in \{0, 1, 2, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the Poisson distribution models the probability of a given number of events occurring in a fixed interval of time or space, under the assumption that events occur independently and at a constant rate λ.



Precisely, λ represents the **average** number of events that occur in an interval, and x represents the **actual** number of events observed in that interval.

2. Discrete distributions: Poisson (λ)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$

2. Discrete distributions: Poisson (λ)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$

  Notice that $x \cdot \frac{\lambda^x}{x!}$ simplifies to: $x \cdot \frac{\lambda^x}{x!} = \frac{\lambda^x}{(x-1)!}$

2. Discrete distributions: Poisson (λ)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$

  Notice that $x \cdot \frac{\lambda^x}{x!}$ simplifies to: $x \cdot \frac{\lambda^x}{x!} = \frac{\lambda^x}{(x-1)!}$

  Thus, we rewrite the sum as: $\mathbb{E}(X) = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}$

2. Discrete distributions: Poisson (λ)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$

  Notice that $x \cdot \frac{\lambda^x}{x!}$ simplifies to: $x \cdot \frac{\lambda^x}{x!} = \frac{\lambda^x}{(x-1)!}$

  Thus, we rewrite the sum as: $\mathbb{E}(X) = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}$

  Shift index to zero using the substitution $y = x - 1$ (so when $x = 1$, $y = 0$):

$$\mathbb{E}(X) = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$$

2. Discrete distributions: Poisson (λ)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$

  Notice that $x \cdot \frac{\lambda^x}{x!}$ simplifies to: $x \cdot \frac{\lambda^x}{x!} = \frac{\lambda^x}{(x-1)!}$

  Thus, we rewrite the sum as: $\mathbb{E}(X) = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}$

  Shift index to zero using the substitution $y = x - 1$ (so when $x = 1$, $y = 0$):

$$\mathbb{E}(X) = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$$

  Recognize the power series expansion $\sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{\lambda}$

2. Discrete distributions: Poisson (λ)

Properties:

- $\mathbb{E}(X) = \sum_{x=0}^{\infty} x \cdot P(X = x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$

  Notice that $x \cdot \frac{\lambda^x}{x!}$ simplifies to: $x \cdot \frac{\lambda^x}{x!} = \frac{\lambda^x}{(x-1)!}$

  Thus, we rewrite the sum as: $\mathbb{E}(X) = \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}$

  Shift index to zero using the substitution $y = x - 1$ (so when $x = 1$, $y = 0$):

$$\mathbb{E}(X) = \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$$

  Recognize the power series expansion $\sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{\lambda}$

  Hence, the expected value becomes: $\mathbb{E}(X) = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda$

- $\mathbb{E}(X^2) \stackrel{?}{=} \lambda + \lambda^2$
- $Var(X) \stackrel{?}{=} \mathbb{E}[(X - \mathbb{E}(X))^2] = \lambda$