Theory Notes

Note: Finished Lecture 5

Introduction

- **Probability** is a branch of mathematics concerned with the study of *random phenomenon* (e.g., experiments, models of populations).
- We are primarily interested in probability as it relates to **statistical inference**, the science of drawing inferences about populations based on only a part of the population (i.e., a sample).

Some Definitions

1. **population**: the entire set of objects that we are interested in studying
   e.g., all ISU students
2. **sample**: the subset of the population available for observation
   e.g., STAT 542 students

Note: population and sample are crucial terms in understanding statistics (i.e., STAT 543), but will not occur very often in our discussions of probability theory (i.e., STAT 542).

3. **experiment**: process of obtaining an observed result of a random phenomenon
4. **sample space** S: the set of all possible outcomes of the experiment

- elements $s \in S$ of a sample space are called **sample points** (s)
- a sample space may be
- **discrete**
  (finite or countably infinite, i.e., listable as a finite/infinite sequence)

$$S = \{s_1, s_2, \ldots, s_n\}$$

or

$$S = \{s_1, s_2, s_3, \ldots\}$$

- or **continuous**
  (uncountably infinite, i.e., a continuum of sample points like
  $S = [0, \infty)$)

5. **event** (e.g., $A, B, \ldots$): subset of the sample space S

- **set**: A is a collection of elements
(in our case, A is a collection of outcomes)
- **membership**: $x \in A$ or $x \notin A$
(x is in A or x is not in A)
- **complement**:
$$A^c = \{x : x \notin A\}$$
(x such that x is not in A)
- **union**:
$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$
(x is in A or B or both)
- **intersection**:
$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$
- **subset**: $A \subset B$ means that A is contained in B
(formally, $x \in A \Rightarrow x \in B$)
- **equality**: $A = B$ if $A \subset B$ and $B \subset A$
- **empty set**: $\varnothing$

Algebraic Laws

- **commutativity**:
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$
- **associativity**:
$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$$
- **distributive law**:
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
- **DeMorgan's laws**:
$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

Aside on disjoint and partitions

- events A and B are **disjoint** (mutually exclusive) if

$$A \cap B = \varnothing$$

- For a sequence $A_1, A_2, \ldots$ of events, we say $A_1, A_2, \ldots$ are **pairwise disjoint** if

$$A_i \cap A_j = \varnothing \quad \text{for all } i \neq j$$

- $A_1, A_2, \ldots$ is a **partition** of S if the A_i's are pairwise disjoint and exhaustive, that is,

$$\bigcup_{i=1}^{\infty} A_i = S \quad \text{and} \quad A_i \cap A_j = \varnothing \quad \text{for all } i \neq j$$

Probability Functions

- A **probability function** is a function P defined on a Borel field $\mathcal{B}$ of the sample space S that satisfies:
  1. $P(A) \geq 0$ for all $A \in \mathcal{B}$
  2. $P(S) = 1$
  3. If $A_1, A_2, \ldots \in \mathcal{B}$ are *pairwise disjoint*, then

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

- Any function satisfying the above is a legitimate probability function.

Theorem 1.2.8.
If P is a probability function and A is any set in $\mathcal{B}$, then:

(a)

$$P(\varnothing) = 0$$

(b)

$$P(A) \leq 1$$

(c)

$$P(A^c) = 1 - P(A)$$

Proof of (c) (parts (a) and (b) follow from (c) and the axioms):

Since

$$S = A \cup A^c,$$

and A and A^c are disjoint, by the axioms of probability,

$$P(S) = P(A \cup A^c) = P(A) + P(A^c).$$

Because $P(S) = 1$, we have

$$1 = P(A) + P(A^c),$$

which implies

$$P(A^c) = 1 - P(A).$$

Theorem 1.2.9.
If P is a probability function and A, B are sets in $\mathcal{B}$, then:

(a)

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

(b)

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

(c) If $A \subset B$, then

$$P(A) \leq P(B).$$

Theorem 1.2.11.
If P is a probability function, then

(a) For any partition $C_1, C_2, \ldots \in \mathcal{B}$ (i.e., disjoint C_i's and $\bigcup_{i=1}^{\infty} C_i = S$),

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

(b) For any sets $A_1, A_2, \ldots \in \mathcal{B}$,

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

Principle of Inclusion–Exclusion.
For any sets $A_1, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right).$$

Equivalently,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i \cap A_j) + \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

This generalizes

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

and is proven by induction.

Bonferroni's Inequalities.
For any sets $A_1, \ldots, A_n$ and any $m \in \{1, \ldots, n\}$,

- if m is odd,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{k=1}^{m} (-1)^{k-1} \left(\sum_{1\le i_1<\cdots<i_k\le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right),$$

- if m is even,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \ge \sum_{k=1}^{m} (-1)^{k-1} \left(\sum_{1\le i_1<\cdots<i_k\le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right).$$

In particular,

$$\sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_i \cap A_j) \le P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i).$$

Combinatorics

Permutations / ordered arrangements II.
When selecting r objects from n objects (without replacement), the number of ordered arrangements possible is

$$n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

Combinations / unordered selections.
The number of ways to choose r objects from n objects (without replacement), where the ordering doesn't matter, is

$$\binom{n}{r} \equiv \frac{n!}{r!\,(n-r)!}.$$

Summary table: number of ways to select r objects from a group of n

	objects chosen **without replacement**	objects chosen **with replacement**
ordered	$\dfrac{n!}{(n-r)!}$	n^r
unordered	$\dbinom{n}{r}$	$\dbinom{n+r-1}{r}$

Conditional Probability

- **Definition:** If A, B are events in S with $P(B) > 0$, then
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$
- In conditioning, B can be thought of as the **updated sample space**, i.e., not all of S is relevant since we know B has occurred.

$P(\cdot \mid B)$ is a probability function that satisfies the usual axioms and properties.

Axioms:

- $P(A \mid B) \geq 0$ for all events A
- $P(B \mid B) = 1$
  (B is the updated sample space)
- If $A_1, A_2, \ldots$ are pairwise disjoint events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

Some properties:

$$P(A^c \mid B) = 1 - P(A \mid B)$$

$$P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B) - P(A_1 \cap A_2 \mid B)$$

It also follows from our definition of conditional probability that

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

More generally, for events $A_1, A_2, \ldots, A_n$,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \,\cdots\, P(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

It is possible to reverse the conditioning of A and B to obtain **Bayes' rule**:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}.$$

More generally, if $A_1, A_2, \ldots$ is a partition of the sample space S, then we obtain a general version of Bayes' rule:

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)\, P(A_j)}.$$

Independence

If $P(A \mid B) = P(A)$, then the occurrence of B does not affect the probability of A. It then follows that

$$P(A \cap B) = P(A)P(B) \quad \text{and} \quad P(B \mid A) = P(B).$$

We define two events A and B as **independent** if

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

More than two events.
$A_1, \ldots, A_n$ are **independent** if and only if, for any subcollection $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ of distinct indices (with any $2 \leq k \leq n$), it holds that

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

- If $A_1, \ldots, A_n$ are independent, then

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \text{for any } i \neq j.$$

- However,

$$P(A_i \cap A_j) = P(A_i)P(A_j) \text{ for } i \neq j$$

does **not** imply that $A_1, \ldots, A_n$ are independent.

If $A_1, \ldots, A_n$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n).$$

However,

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

holding does **not** imply that $A_1, \ldots, A_n$ are independent.

The assumption of independence of events allows the computation of joint occurrences of events through simple calculations.

Random Variables

Definition: A **random variable** (r.v.) X is a function defined on a sample space S that associates a real number with each outcome in S.

That is, for each $s \in S$, we have

$$X(s) \in \mathbb{R}.$$

In function notation,

$$X : S \to \mathbb{R}.$$

We usually suppress the dependence of X on $s \in S$ and write

$$X = X(s).$$

We have $P(A)$ defined on events $A \subset S$, which can be used to assign probabilities for events concerning a random variable X on $\mathbb{R}$ ($X : S \to \mathbb{R}$).

Define $P_X(\cdot)$ for events $B \subset \mathbb{R}$ as follows:

$$P_X(B) = P_X(X \in B) = P\big(\{s \in S : X(s) \in B\}\big).$$

$P_X(\cdot)$ satisfies the axioms and is therefore a legitimate probability function.

CDF

Definition.
The **cumulative distribution function** (cdf) of a random variable X, denoted by $F(\cdot)$, is defined by

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Sometimes written with subscript as $F_X(x)$.

A function $F(x)$, $x \in \mathbb{R}$, is a cdf for some random variable if and only if the following hold:

1. $F(x)$ is a nondecreasing function of x.
2. $$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$
3. $F(x)$ is right continuous, i.e.,
$$\lim_{x \downarrow x_0} F(x) = F(x_0) \quad \text{for any } x_0 \in \mathbb{R}.$$

Discrete Random Variables

Definition.
If a cdf F is a step function (with jumps at a countable collection of points $x_i \in \mathbb{R}$), then we say the distribution described by F is **discrete** (with support or range $x_i \in \mathbb{R}$).

If a random variable X has a cdf $F = F_X$ which is a step function, then we say X is a **discrete random variable**.

Besides the cdf, there are other (equivalent) ways to state the probability distribution for a discrete distribution / discrete r.v. X.

1. **Probability mass function (pmf).**
   The pmf of a discrete random variable X is given by
$$f(x) = P(X = x) \geq 0, \quad \text{for any } x \in \mathbb{R}.$$
2. **Equivalent characterization via the cdf.**
   The pmf of a discrete r.v. X can also be written as
$$f(x) = P(X \leq x) - P(X < x) = F(x) - \lim_{y \to x^-} F(y).$$

Continuous Random Variables and Probability Density Functions

- If a cdf F is such that there exists a nonnegative function f satisfying

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \quad \text{for any } x \in \mathbb{R},$$

  then the distribution described by F is said to be (absolutely) **continuous** with **probability density function (pdf)** f.

A random variable X with an (absolutely) continuous cdf F, or a pdf f, is said to be a **continuous random variable**.

- If F is (absolutely) continuous, then its derivative at $x \in \mathbb{R}$ is its pdf $f(x)$:

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

- If X is a continuous random variable, then

$$P(X = x) = 0 \quad \text{for any } x \in \mathbb{R}.$$

For $a < b$,

$$P(a < X < b) = P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = F(b) - F(a) = \int_a^b f(t)\,dt.$$

Properties of Probability Density or Mass Functions

A function $f(x)$ is a pdf (or pmf) for some random variable if and only if

1. $f(x) \ge 0$ for any $x \in \mathbb{R}$
2. $$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad (\text{or } \sum_x f(x) = 1)$$

Any nonnegative function having a finite integral (or sum) can be turned into a pdf (or pmf) f by dividing by its integral (or sum).

We will write $X \sim f_X(x)$ (or $X \sim F_X(x)$) to denote that X has a distribution given by f (or F).

Computing Probabilities Using a pmf or pdf

To find general probabilities using a pmf or pdf, note that for $A \subset \mathbb{R}$,

Discrete case (using pmf):

$$P(X \in A) = \sum_{x \in A} f_X(x) = \sum_{x \in A,\ f_X(x) > 0} f_X(x)$$

Continuous case (using pdf):

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

Relating the CDF to the PMF / PDF

Discrete random variable case

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

$$P(a \leq X \leq b) = F(b) - F(a) + f(a)$$

$$P(a \leq X < b) = F(b) - F(a) + f(a) - f(b)$$

$$P(a < X < b) = F(b) - F(a) - f(b)$$

Continuous random variable case

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

$$P(a \leq X \leq b) = F(b) - F(a)$$

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

$$P(a < X < b) = F(b) - F(a)$$

Equivalently,

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

Functions of a Random Variable

Introduction

- Consider a random variable $X \sim F_X(\cdot)$ and a function
$$g : \mathbb{R} \to \mathbb{R}.$$
(Here, X is a random variable and g may be *any* function.)
- Then
$$Y = g(X)$$
is also a random variable, having its own cdf $F_Y(\cdot)$.

Since Y is a function of X, we can describe the probabilistic behavior of Y in terms of that of X.

- Formally, there is also an inverse mapping g^{-1} defined by
$$g^{-1}(A) = \{x \in \mathbb{R} : g(x) \in A\}, \qquad \text{for any } A \subset \mathbb{R}.$$

Distribution of a Function of a Random Variable

- The distribution of $Y = g(X)$ is completely determined by the distribution of X and the function g.

For any set $A \subset \mathbb{R}$,
$$P_Y(Y \in A) = P_X(g(X) \in A) = P_X(X \in g^{-1}(A)).$$

That is, the distribution of Y depends on the cdf (or pdf/pmf) F_X of X together with the function g.

Support (Range) Under Transformations

- If X has pdf/pmf $f_X(x)$, then the **range (support)** of X is
$$\mathcal{X} = \{x \in \mathbb{R} : f_X(x) > 0\}.$$
- If $Y = g(X)$ has pdf/pmf $f_Y(y)$, then the **range (support)** of Y is
$$\mathcal{Y} = \{y \in \mathbb{R} : f_Y(y) > 0\} = \{g(x) : x \in \mathcal{X}\}.$$

Discrete Case

Result.
If X is a discrete random variable with pmf $f_X(x)$ (i.e., X has range
$$\mathcal{X} = \{x \in \mathbb{R} : f_X(x) > 0\},$$
which is either finite or countably infinite), then
$$Y = g(X)$$
is also a discrete random variable with pmf
$$f_Y(y) = P(Y = y) = \begin{cases} \displaystyle\sum_{x \in g^{-1}(\{y\})} f_X(x) = \sum_{x \in \mathcal{X}: g(x) = y} f_X(x), & y \in \mathcal{Y}, \\ 0, & y \notin \mathcal{Y}, \end{cases}$$
where the range (support) of Y is
$$\mathcal{Y} = \{g(x) : x \in \mathcal{X}\} = \{y \in \mathbb{R} : f_Y(y) > 0\}.$$

Continuous Case

For a continuous random variable X, the random variable
$$Y = g(X)$$
will *typically* (but not always) be continuous.

To determine the distribution of Y, one can use either of the following two approaches.

CDF Method

Compute the cdf $F_Y(\cdot)$ of Y:

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(g(X) \le y) \\ &= P(\{x \in \mathbb{R} : g(x) \le y\}) \\ &= \int_{\{x \in \mathbb{R}: g(x) \le y\}} f_X(x)\, dx. \end{aligned}$$

This is a general approach, but its success depends on being able to evaluate the integral.

PDF (Transformation) Method

Alternatively, one may compute the pdf $f_Y(\cdot)$ directly using a transformation technique.

This method is **only valid** when the function g is **monotone** or **piecewise monotone**.

Key Result

Theorem 2.1.5 (Monotone Transformation)

If X has pdf $f_X(x)$ and

$$Y = g(X),$$

where the function $g(\cdot)$ has either a **strictly positive** or a **strictly negative** derivative on

$$\mathcal{X} = \{x \in \mathbb{R} : f_X(x) > 0\},$$

then the pdf of Y has support

$$\mathcal{Y} = \{g(x) : x \in \mathcal{X}\},$$

and is given by

$$f_Y(y) = f_X\big(g^{-1}(y)\big) \left| \frac{dg^{-1}(y)}{dy} \right| > 0, \quad \text{for } y \in \mathcal{Y},$$

with

$$f_Y(y) = 0, \quad \text{for } y \notin \mathcal{Y}.$$

Note that unless g is **strictly monotone** (or at least there is a way to break up

$$\mathcal{X} = \{x \in \mathbb{R} : f_X(x) > 0\}$$

into several intervals on each of which g is strictly increasing or strictly decreasing),
X being a continuous random variable does **not** necessarily imply that

$$Y = g(X)$$

will be a continuous random variable.

Probability Integral Transform (PIT)

This is a famous (and for some purposes very useful) transformation connected with continuous cdfs.

If F is a continuous cdf, then

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \quad t \in \mathbb{R}.$$

If X has a continuous cdf $F(\cdot)$, then the random variable

$$Y = F(X)$$

is uniformly distributed on $(0, 1)$.

That is, Y has pdf

$$f_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise}, \end{cases}$$

and cdf

$$F_Y(y) = \begin{cases} 0, & y \leq 0, \\ y, & 0 \leq y \leq 1, \\ 1, & y \geq 1. \end{cases}$$

Expected Value of a Function of a Random Variable

Definition.
The expected value (or mean) of a random variable $g(X)$, denoted by $\mathrm{E}g(X)$, $\mathrm{E}[g(X)]$, or $\mathrm{E}(g(X))$, is defined as follows.

Discrete case:

$$\mathrm{E}g(X) = \sum_x g(x)\, f_X(x).$$

Continuous case:

$$\mathrm{E}g(X) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\,dx.$$

Existence of the Expectation

The expectation $\mathrm{E}g(X)$ is defined **provided that**

Discrete case:

$$\sum_x |g(x)|\, f_X(x) < \infty,$$

Continuous case:

$$\int_{-\infty}^{\infty} |g(x)|\, f_X(x)\,dx < \infty.$$

(That is, we require $\mathrm{E}[g(X)]$ to be a real, finite number.)

Nonexistence of the Expectation

We say that the expected value (or mean) $\mathrm{E}g(X)$ **does not exist** if

Discrete case:

$$\sum_x |g(x)|\, f_X(x) = \infty,$$

Continuous case:

$$\int_{-\infty}^{\infty} |g(x)|\, f_X(x)\,dx = \infty.$$

Theorem 2.2.5 (Properties of Expectation)

Theorem.
Suppose X is a random variable such that

$$\mathrm{E}|g_1(X)| < \infty \quad \text{and} \quad \mathrm{E}|g_2(X)| < \infty,$$

and let $a, b, c \in \mathbb{R}$ be fixed constants. Then:

1. $$\mathrm{E}[ag_1(X) + b] = a\,\mathrm{E}g_1(X) + b.$$

2. $$\mathrm{E}[ag_1(X) + bg_2(X) + c] = a\,\mathrm{E}g_1(X) + b\,\mathrm{E}g_2(X) + c.$$

3. If $g_1(x) \geq a$ for all x, then
$$\mathrm{E}g_1(X) \geq a.$$

4. If $g_1(x) \leq b$ for all x, then
$$\mathrm{E}g_1(X) \leq b.$$

5. If $g_1(x) \geq g_2(x)$ for all x, then
$$\mathrm{E}g_1(X) \geq \mathrm{E}g_2(X).$$

Invariance of Expectation Under Transformation

Expectations are invariant under transformation.

If

$$Y = g(X),$$

then

$$\mathrm{E}Y = \sum_y y\, f_Y(y) = \sum_y y\, P(Y = y) = \sum_x g(x)\, f_X(x) = \mathrm{E}g(X)$$

in the discrete case.

(In the continuous case, replace sums with integrals.)

That is,

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx = \mathrm{E}g(X).$$

Variance

An important instance of the $\mathrm{E}g(X)$ notation arises when

$$g(X) = (X - \mathrm{E}X)^2.$$

Definition.
The **variance** of a random variable X, denoted $\mathrm{Var}(X)$ or σ_X^2, is

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

the expected squared distance between X and its mean $\mathrm{E}X$.

Two Important Variance Facts

1. For any real numbers a, b,
$$\mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X).$$

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

Other Moments and Distributional Summaries

Moments are an important summary of a distribution.

1. $$\mu = \mu_X = \mathrm{E}X$$
is often called the **mean**.

2. $$\mu'_n = \mathrm{E}X^n$$
is the nth (raw) moment, provided $\mathrm{E}X^n$ exists, i.e.,

Discrete case:
$$\sum_x |x^n| f_X(x) < \infty,$$

Continuous case:
$$\int_{-\infty}^{\infty} |x^n| f_X(x)\, dx < \infty.$$

3. $$\mu_n = \mathrm{E}\big[(X - \mu)^n\big]$$
is the nth **central moment**, provided $\mathrm{E}X^n$ exists.

(a) $$\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}\big[(X - \mu)^2\big] = \mu_2$$
is the **variance**.

(b) $$\sigma_X = \sqrt{\mathrm{Var}(X)}$$
is the **standard deviation**.

(c) $$\mu_3$$
is **skewness** (i.e., measures distributional balance around μ).

(d) $$\mu_4$$
is **kurtosis** (i.e., a measure of how long the distributional tails are).

Regarding Moments

1. If $\mathrm{E}X^r$ exists for some $r > 0$, then $\mathrm{E}X^s$ exists for all

$$0 \leq s \leq r.$$

2. If $\mathrm{E}X^r$ does not exist for some $r > 0$, then $\mathrm{E}X^s$ will not exist for any

$$s > r.$$

3.

$$\mathrm{E}X^2 \text{ exists if and only if } \mathrm{Var}(X) \text{ exists.}$$

4. For $r > 0$, the existence of $\mathrm{E}X^r$ is a matter of the distribution of X not having **heavy tails**, i.e., X does not assume large values with large probability.