MTL108
Random Variables, Expectation and Moments

Rahul Singh

Discrete probability space

Definition 1 (Discrete probability space)**.** Let Ω be a finite or countable set, $\mathcal{C} = 2^{\Omega}$, power set of Ω, and $\mathbb{P} : \Omega \to [0,1]$ defined by

$$\mathbb{P}(\{\omega\}) = p_w \text{ for } \omega \in \Omega \text{ and } \mathbb{P}(A) = \sum_{\omega \in A} p_w \text{ for } A \in \mathcal{C},$$

where

$$p_w \geq 0 \text{ and } \sum_{w \in \Omega} p_w = 1.$$

Then, $(\Omega, \mathcal{C}, \mathbb{P})$ is said to be a discrete probability space and $\mathbb{P}$ is known as a discrete probability measure on Ω.

Example 1 (Fair coins)**.** Consider tossing two coins; $\Omega = \{HH, HT, TH, TT\}$, $\mathcal{C} = 2^{\Omega}$. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a discrete probability space, where

$$\mathbb{P}(\{\omega\}) = p_w = \frac{1}{4} \text{ if } \omega \in \Omega.$$

Further, for any $A \in \mathcal{C}$

$$\mathbb{P}(A) = \sum_{\omega \in A} p_w,$$

for example, if $A = \{HH, TT\}$ then

$$\mathbb{P}(A) = \mathbb{P}(\{HH, TT\}) = \sum_{\omega \in \{HH, TT\}} p_w = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Example 2 (Maybe biased coins)**.** Consider tossing two coins; $\Omega = \{HH, HT, TH, TT\}$, $\mathcal{C} = 2^{\Omega}$. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a discrete probability space, where

$$\mathbb{P}(\{HH\}) = p_1, \;\; \mathbb{P}(\{HT\}) = p_2, \;\; \mathbb{P}(\{TH\}) = p_3, \;\; \mathbb{P}(\{TT\}) = p_4,$$

where $p_1, p_2, p_3, p_4 \geq 0$ and $p_1 + p_2 + p_3 + p_4 = 1$. Further, for any $A \in \mathcal{C}$

$$\mathbb{P}(A) = \sum_{\omega \in A} p_w,$$

for example, if $A = \{HH, TT\}$ then

$$\mathbb{P}(A) = \mathbb{P}(\{HH, TT\}) = \sum_{\omega \in \{HH,TT\}} p_w = p_1 + p_4.$$

Theorem 1. *A discrete probability measure is a probability measure.*

Proof. We assume the setup in the above definition. To complete the proof, we need to verify three axioms. Verifications of the first two axioms are straightforward.

Axiom 1 (Non-negativity) $\mathbb{P}(A) = \sum_{\omega \in A} p_w \geq 0$ for all $A \in \mathcal{C}$ as $p_w \geq 0$; non-negativity holds.

Axiom 2 (Normalization) $\mathbb{P}(\Omega) = \sum_{w \in \Omega} p_w = 1$; normalization holds.

Axiom 3 (Countable additivity) Consider a collection of pairwise disjoint events $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{C}$ (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$),

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} p_w = \sum_{\omega \in A_1} p_w + \sum_{\omega \in A_2} p_w + \ldots \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots = \sum_{i=1}^{\infty} \mathbb{P}(A_i).\end{aligned}$$

So,

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

Consequently, the third axiom also holds.

Therefore, all three axioms for probability measure are satisfied by $\mathbb{P}$, so a discrete probability measure is a probability measure. □

Remark 1. For a finite or countable set, defining probability on singleton sets is sufficient for defining a probability measure.

Random Variable

A random variable assigns a numerical value to each outcome in a sample space, allowing us to quantify and analyze uncertainty mathematically. Random variables bridge the gap between abstract probabilistic events and concrete numerical computations, enabling the use of algebraic and analytical tools in probability problems.

Definition 2 (Random Variable)**.** A random variable X on measurable space $(\Omega, \mathcal{C})$ is a function $X : \Omega \to \mathbb{R}$ such that for any $a \in \mathbb{R}$,

$$X^{-1}((-\infty, a]) := \{\omega \in \Omega \,:\, X(\omega) \in (-\infty, a]\} \in \mathcal{C}.$$

Remark 2. A random variable is usually denoted by upper-case letters.

Example 3 (Two coins)**.** Consider tossing two coins; $\Omega = \{HH, HT, TH, TT\}$, $\mathcal{C} = 2^{\Omega}$. Define a random variable $X : \Omega \to \mathbb{R}$ such that

$$X(\omega) = \text{ number of heads in } \omega.$$

Then,

$$X(HH) = 2, \;\; X(HT) = X(TH) = 1, \;\; X(TT) = 0.$$

Precisely, random variable $X : \Omega \to \{0, 1, 2\}$. Further, events can also be written in terms of the random variable, e.g.,

1. $\{\omega \in \Omega : X(\omega) = 0\} = \{TT\}$
2. $\{\omega \in \Omega : X(\omega) = 1\} = \{TH, HT\}$
3. $\{\omega \in \Omega : X(\omega) = 2\} = \{HH\}$
4. $\{\omega \in \Omega : X(\omega) \leq 1\} = \{\omega \in \Omega : X(\omega) \in \{0, 1\}\} = \{TT, TH, HT\}$
5. $\{\omega \in \Omega : X(\omega) < 1\} = \{\omega \in \Omega : X(\omega) = 0\} = \{TT\}$
6. $\{\omega \in \Omega : X(\omega) \leq 0.5\} = \{\omega \in \Omega : X(\omega) = 0\} = \{TT\}$
7. $\{\omega \in \Omega : X(\omega) \leq 1.5\} = \{\omega \in \Omega : X(\omega) \in \{0, 1\}\} = \{TT, TH, HT\}$.

Remark 3 (Simplified notations)**.** Usually, for a random variable we write X instead of $X(\omega)$ and in probability statements we drop curly braces, i.e., we write $\mathbb{P}(X \leq 1)$ instead of $\mathbb{P}(\{X \leq 1\})$ or $\mathbb{P}(\{X(\omega) \leq 1\})$.

Random variables are classified into two main types: discrete and continuous. Discrete random variables take on a countable number of values (finite or infinite), while continuous random variables take on uncountably many values, typically over an interval.

Definition 3 (Probability distribution/law of a random variable)**.** Let X be a random variable defined on probability space $(\Omega, \mathcal{C}, \mathbb{P})$. The probability distribution/law of the random variable X induced by $(\Omega, \mathcal{C}, \mathbb{P})$ is a function $\mathbb{P}_X : \mathbb{R} \to [0, 1]$, such that $\mathbb{P}_X$ is a probability measure on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ satisfying

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) \quad \text{for any} \quad A \in \mathbb{B}(\mathbb{R}).$$

Here, $(\mathbb{R}, \mathbb{B}(\mathbb{R}), \mathbb{P}_X)$ is a probability space induced by $(\Omega, \mathcal{C}, \mathbb{P})$.

Definition 4 (Cumulative distribution function (CDF))**.** The cumulative distribution function (CDF) for a random variable X is a function $F_X : \mathbb{R} \to [0, 1]$, defined by

$$F_X(x) = \mathbb{P}_X(X \leq x).$$

Theorem 2 (Without proof). *There is one-to-one correspondence between $\mathbb{P}_X$ and F_X.*

The cumulative distribution function (CDF) of a random variable X possesses several key properties regardless of whether X is discrete or continuous.

Theorem 3 (Characterizing Properties of CDF, without proof). *1.* **Limit at $-\infty$ and ∞**

$$\lim_{x\to-\infty} F_X(x) = 0 \textit{ and } \lim_{x\to\infty} F_X(x) = 1.$$

2. **Non-decreasing** *A CDF $F_X(x)$ is a non-decreasing function, that is,*

$$F_X(a) \le F_X(b) \textit{ if } a \le b.$$

3. **Right continuity** *A CDF $F_X(x)$ is right-continuous, i.e., for any $x \in \mathbb{R}$,*

$$\lim_{h\to0^+} F_X(x+h) = F_X(x).$$

Remark 4. As x approaches negative infinity, x becomes an extremely small number. The event $X \le x$ means the random variable X takes a value less than or equal to this very small x. In most practical scenarios, random variables have a lower bound or the probability of extremely small values is negligible. Heuristically, as x gets smaller and smaller, fewer (or no) outcomes satisfy $X \le x$, so the probability approaches 0. This reflects that the total probability starts from 0 and builds up as x increases.

Next, as x approaches positive infinity, x becomes arbitrarily large. The event $X \le x$ encompasses almost all possible values of X, since even the largest conceivable outcomes are less than or equal to such a large x. Thus, the probability approaches 1, capturing the fact that the total probability mass or density over the entire range of X sums to 1. This ensures the CDF accounts for all possible outcomes as x covers the entire real line.

Remark 5. The nondecreasing property of the CDF reflects the natural accumulation of probability as the threshold x grows. The value $F_X(x)$ represents the probability that X is less than or equal to x, which is the total probability mass or density accumulated up to that point. As x increases, the event $X \le x$ includes all outcomes where X is less than or equal to the previous x, plus any additional outcomes where X falls between the old x and the new x. Since probabilities are non-negative, adding these additional outcomes (if any) can only increase or leave unchanged the total probability. It cannot decrease, as no probability mass is removed when expanding the interval.

In discrete cases, the CDF increases in steps at points where X has positive probability, while in continuous cases, it increases smoothly (or remains constant where the density is zero), but never decreases.

Remark 6. Right-continuity means that at any point x, $\lim_{h\to0^+} F_X(x+h) = F_X(x)$, while there might be a jump discontinuity from the left. Heuristically, this arises because the CDF includes the probability up to and including x (i.e., $\mathbb{P}(X \le x)$). For continuous distributions, the CDF is smooth and continuous everywhere. For discrete distributions, jumps occur at points where X has positive probability, and right-continuity ensures the jump is "attached" to the right side, reflecting that the probability at the point is included when approaching from the right. This convention aligns with how we define intervals in probability calculations, ensuring consistency in applications like quantile functions.

These properties collectively ensure the CDF is a non-decreasing function that starts at 0, ends at 1, and handles discontinuities in a standardized way, making it a reliable tool for probability computations.

Example 4 (Two Coin Flips)**.** Consider a random variable X representing the number of heads in two independent flips of a fair coin. The possible values are 0, 1, or 2, each with probabilities $\mathbb{P}(X = 0) = 1/4$, $\mathbb{P}(X = 1) = 1/2$, $\mathbb{P}(X = 2) = 1/4$.
The CDF $F_X(x)$ is:

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 1/4 & 0 \leq x < 1, \\ 3/4 & 1 \leq x < 2, \\ 1 & x \geq 2. \end{cases}$$

- As $x \to -\infty$, $F_X(x) = 0$, since no outcomes are less than 0.
- As $x \to \infty$, $F_X(x) = 1$, covering all possible heads.
- Right-continuity: At $x = 1$, approaching from the right (e.g., 1.0001) gives $3/4$, matching $F_X(1) = 3/4$. From the left (e.g., 0.999), it's $1/4$, showing a jump, but the function is continuous from the right.

Example 5 (Rolling a Fair Die)**.** Consider a random variable X representing the number of spots on the upward face of a fair six-sided die. The possible values are $\{1, 2, 3, 4, 5, 6\}$, each with probability $\mathbb{P}(X = k) = 1/6$.
The CDF $F_X(x)$ is:

$$F_X(x) = \begin{cases} 0 & x < 1, \\ 1/6 & 1 \leq x < 2, \\ 2/6 & 2 \leq x < 3, \\ 3/6 & 3 \leq x < 4, \\ 4/6 & 4 \leq x < 5, \\ 5/6 & 5 \leq x < 6, \\ 1 & x \geq 6. \end{cases}$$

- At $x = 0.5$, $F_X(0.5) = 0$ (no outcomes ≤ 0.5).
- At $x = 1.5$, $F_X(1.5) = 1/6$ (includes only $X = 1$).
- At $x = 2.5$, $F_X(2.5) = 2/6$ (includes $X = 1, 2$).

As x increases from 0.5 to 1.5 to 2.5, $F_X(x)$ increases from 0 to $1/6$ to $2/6$, reflecting the addition of new outcomes (e.g., $X = 2$ at $x = 2.5$). The function never decreases; it either stays constant between integer jumps or increases at each step where a new value of X is included. This step-wise increase illustrates the nondecreasing nature, driven by the cumulative inclusion of probability mass.

Discrete Random Variable

Definition 5 (Discrete Random Variable)**.** A random variable X is discrete if its range (the set of possible values) is countable, i.e., it can take values in a set like $\{x_1, x_2, \dots\}$ where the x_i are distinct real numbers.

See Example 3 for an example. The probability distribution of a discrete random variable is described by its probability mass function (PMF).

Definition 6 (Probability Mass Function (PMF))**.** The PMF of a discrete random variable X, denoted $p_X(x)$ or $\mathbb{P}(X = x)$, is defined as

$$p_X(x) = \mathbb{P}(X = x),$$

for each possible value x in the range of X. The PMF satisfies:

1. $p_X(x) \geq 0$ for all x,
2. $p_X(x) = 0$ if x is not in the range of X,
3. $\sum_{\text{all } x} p_X(x) = 1$.

Lemma 1 (CDF of discrete random variable)**.** *The cumulative distribution function (CDF) for a discrete random variable can be expressed*

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} p_X(y).$$

Lemma 2. *CDF for a discrete random variable is a jump function. Jumps are at mass points.*

Example 6 (Bernoulli Distribution)**.** A Bernoulli random variable models a single trial with two outcomes: success (1) or failure (0), with success probability p (where $0 < p < 1$).
- Range: $\{0, 1\}$
- PMF:

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

This is often denoted as $X \sim \text{Bernoulli}(p)$.

Example 7 (Binomial Distribution)**.** A binomial random variable counts the number of successes in n independent Bernoulli trials, each with success probability p.
- Range: $\{0, 1, 2, \ldots, n\}$
- PMF:

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

This is denoted as $X \sim \text{Binomial}(n, p)$.

Uncountable Probability Space

When Ω is uncountable, defining probabilities using probabilities for singleton sets is not possible. For example, let $\Omega = (0, 1)$, then each interval subset of Ω is an uncountable set. A probability measure on Ω is defined using length of interval inside $[0, 1]$, i.e., for (a, b) where $0 \leq a \leq b \leq 1$

$$\mathbb{P}((a, b)) = b - a.$$

Continuous Random Variable

Continuous random variables are essential in probability theory for modeling phenomena that can take any value within a continuum, such as time, height, or temperature. Unlike discrete random variables, which have countable outcomes, continuous random variables have uncountably many possible values, typically over an interval. This requires different tools for describing their distributions, such as probability density functions (PDFs) instead of probability mass functions.

Definition 7 (Continuous Random Variable)**.** A random variable X is continuous if its cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x)$ is continuous everywhere. Typically, continuous random variables take values in an interval or the entire real line, and the probability of X taking any specific value is zero, i.e., $\mathbb{P}(X = c) = 0$ for any constant c.

Definition 8 (Probability Density Function)**.** The PDF of a continuous random variable X, denoted $f_X(x)$, is a non-negative function such that the probability of X falling in an interval $[a, b]$ is the integral of the PDF over that interval:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx.$$

The PDF satisfies:
1. $f_X(x) \geq 0$ for all x,
2. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

Definition 9 (Cumulative Distribution Function)**.** The CDF of X, denoted $F_X(x)$, is

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Theorem 4 (Without proof)**.** *The PDF is the derivative of the CDF:* $f_X(x) = \frac{d}{dx} F_X(x)$ *where the derivative exists.*

Example 8 (Uniform Distribution)**.** The uniform random variable models equal likelihood over an interval $[a, b]$, where $a < b$.
- Support: $[a, b]$
- PDF:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

- CDF:

$$F_X(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{x-a}{b-a} & \text{if } a \leq x < b, \\ 1 & \text{if } x \geq b. \end{cases}$$

Example 9 (Exponential Distribution)**.** The exponential random variable models the time between events in a Poisson process, with rate parameter $\lambda > 0$.
- Support: $[0, \infty)$
- PDF:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- CDF:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

Expectation (or expected value)

The expectation (or expected value) of a random variable is a fundamental concept in probability theory, representing the average or mean value of the variable over many trials. It serves as a measure of central tendency and is a cornerstone for further statistical analysis. The expectation of a function g of a random variable X is denoted by $\mathbb{E}[g(X)]$.

Expectation of Discrete Random Variables

Definition 10 (Expectation of a Discrete Random Variable)**.** Let X be a discrete random variable with probability mass function (PMF) $p_X(x)$, where the possible values of X are $\{x_1, x_2, \dots\}$. The expectation of X, denoted $\mathbb{E}[X]$, is defined as

$$\mathbb{E}[X] = \sum_x x p_X(x),$$

where the sum is taken over all possible values x in the support of X, provided the sum converges absolutely (i.e., $\sum |x| p_X(x) < \infty$).

For a function $g(X)$ of the discrete random variable X,

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x).$$

Example 10. Consider a fair six-sided die, where X is the outcome, and $p_X(x) = 1/6$ for $x = 1, 2, \dots, 6$.

$$\mathbb{E}[X] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5.$$

Expectation of Continuous Random Variables

Definition 11 (Expectation of a Continuous Random Variable)**.** Let X be a continuous random variable with probability density function (PDF) $f_X(x)$. The expectation of X, denoted $\mathbb{E}[X]$, is defined as

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

provided the integral converges absolutely (i.e., $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$).

For a function $g(X)$ of the continuous random variable X,

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

Example 11. Consider X uniformly distributed over $[0, 1]$, with $f_X(x) = 1$ for $0 \le x \le 1$, and 0 otherwise.

$$\mathbb{E}[X] = \int_0^1 x \cdot 1 \, dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

Linearity Property of Expectation

Theorem 5 (Linearity of Expectation)**.** *For any two functions $g(\cdot)$ and $h(\cdot)$ of random variables X (discrete or continuous) and constants a and b,*

$$\mathbb{E}[a\, g(X) + b\, h(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)].$$

Proof. Note that $a\, g(X) + b\, h(X)$ is also a function of random variable X, so the expectation is defined accordingly. We prove for both discrete and continuous cases.

Discrete Case: Let X has PMF $p_x = \mathbb{P}(X = x)$. Then, using properties of addition we have

$$\begin{aligned}
\mathbb{E}[a\, g(X) + b\, h(X)] &= \sum_x [a\, g(x) + b\, h(x)]\, p_x \\
&= \sum_x a\, g(x)\, p_x + \sum_x b\, h(x)\, p_x, \text{ by splitting sum} \\
&= a \sum_x g(x)\, p_x + b \sum_x h(x)\, p_x, \text{ by factoring out constants.}
\end{aligned}$$

Now observe that

$$\mathbb{E}[g(X)] = \sum_x g(x)\, p_x \quad \text{and} \quad \mathbb{E}[h(X)] = \sum_x h(x)\, p_x.$$

Substituting in above expression we get

$$\mathbb{E}[a\, g(X) + b\, h(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)].$$

Continuous Case: Let X has PDF $f(x)$. Then, using properties of integral we have

$$\begin{aligned}
\mathbb{E}[a\, g(X) + b\, h(X)] &= \int_x [a\, g(x) + b\, h(x)]\, f(x)\, dx \\
&= \int_x a\, g(x)\, f(x)\, dx + \int_x b\, h(x)\, f(x)\, dx, \text{ by splitting integral} \\
&= a \int_x g(x)\, f(x)\, dx + b \int_x h(x)\, f(x)\, dx, \text{ by factoring out constants.}
\end{aligned}$$

Now observe that

$$\mathbb{E}[g(X)] = \int_x g(x)\, f(x)\, dx \quad \text{and} \quad \mathbb{E}[h(X)] = \int_x h(x)\, f(x)\, dx.$$

Substituting in above expression we get

$$\mathbb{E}[a\, g(X) + b\, h(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)].$$

□

Moments of Random Variables

Moments of random variables are statistical measures that provide insights into the shape, spread, and central tendency of a probability distribution. They are fundamental in understanding the behavior of random variables, both discrete and continuous, and are widely used in fields such as engineering and data science.

Definition 12 (Moment)**.** The k-th moment of a random variable X about the origin is the expected value of X^k, denoted $\mathbb{E}[X^k]$, where k is a non-negative integer. For a discrete random variable with probability mass function $p_X(x)$, it is

$$\mathbb{E}[X^k] = \sum_x x^k p_X(x),$$

and for a continuous random variable with probability density function $f_X(x)$, it is

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

provided the integrals or sums converge.

Definition 13 (Central Moment)**.** The k-th central moment of a random variable X is the expected value of $(X-\mathbb{E}[X])^k$, denoted $\mathbb{E}[(X-\mathbb{E}[X])^k]$. The first central moment is zero, and higher moments describe the shape of the distribution around the mean.

Remark 7 (Raw Moments vs. Central Moments)**. Raw Moments** are the expected values of powers of X, defined as $\mathbb{E}[X^k]$ for $k = 1, 2, \ldots$. The first raw moment ($\mathbb{E}[X]$) is the mean, the second raw moment ($\mathbb{E}[X^2]$) relates to the squared values, and so on. **Central Moments**, on the other hand, are the expected values of powers of the deviation of X from its mean, $\mathbb{E}[(X-\mathbb{E}[X])^k]$. The first central moment is zero (since $\mathbb{E}[X-\mathbb{E}[X]] = 0$), the second central moment is the variance ($\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2]$), and higher central moments (e.g., skewness for $k = 3$, kurtosis for $k = 4$) describe the distribution's shape around the mean. Central moments are particularly valuable for understanding the spread and asymmetry of the data relative to its center.

Mean (First Moment)

Definition 14 (Mean)**.** The mean, or expected value, of a random variable X, denoted $\mathbb{E}[X]$, is the first moment about the origin. It represents the average value of X over many trials. For a discrete random variable,

$$\mathbb{E}[X] = \sum_x x p_X(x),$$

and for a continuous random variable,

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

The mean is the center of gravity of the distribution.

Variance (Second Central Moment)

Definition 15 (Variance)**.** The variance of a random variable X, denoted $\text{Var}(X)$, is the second central moment, measuring the spread of X around its mean $\mu = \mathbb{E}[X]$. It is defined as

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

For a discrete random variable,

$$\text{Var}(X) = \sum_x (x - \mu)^2 p_X(x),$$

and for a continuous random variable,

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)\, dx.$$

Lemma 3. *For both discrete and continuous random variables,*

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

Proof. See Solution to Problem Set-3. □

Standard Deviation

Definition 16 (Standard Deviation)**.** The standard deviation of a random variable X, denoted σ, is the square root of the variance:

$$\sigma = \sqrt{\text{Var}(X)}.$$

It provides a measure of spread in the same units as X, making it more interpretable than variance.

Examples

Discrete Example: Bernoulli Random Variable

Example 12 (Bernoulli Distribution)**.** Consider a Bernoulli random variable X with $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$.

- Mean:

$$\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p.$$

- Second Moment:

$$\mathbb{E}[X^2] = 0^2 \cdot (1 - p) + 1^2 \cdot p = p.$$

- Variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p).$$

- Standard Deviation:

$$\sigma = \sqrt{p(1 - p)}.$$

For $p = 0.3$, $\mathbb{E}[X] = 0.3$, $\text{Var}(X) = 0.3 \cdot 0.7 = 0.21$, and $\sigma = \sqrt{0.21} \approx 0.458$.

Continuous Example: Uniform Distribution

Example 13 (Uniform Distribution)**.** Consider a random variable X uniformly distributed over $[0, 1]$, with PDF $f_X(x) = 1$ for $0 \leq x \leq 1$, and 0 otherwise.
- Mean:

$$\mathbb{E}[X] = \int_0^1 x \cdot 1 \, dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

- Second Moment:

$$\mathbb{E}[X^2] = \int_0^1 x^2 \cdot 1 \, dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}.$$

- Variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{4}{12} - \frac{3}{12} = \frac{1}{12}.$$

- Standard Deviation:

$$\sigma = \sqrt{\frac{1}{12}} \approx 0.289.$$

This reflects the even spread of X across $[0, 1]$.

Remarks

- The mean provides the balance point of the distribution.
- Variance quantifies dispersion, with larger values indicating greater variability.
- Standard deviation is useful for comparing spread across different scales.
- Higher moments (e.g., skewness, kurtosis) can be derived similarly, offering insights into asymmetry and tail behavior.

These measures are critical for statistical inference, hypothesis testing, and modeling real-world phenomena like test scores or machine failures.

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer