Theorems

Sum of Geometric Series and Sequences

$$S_n = \sum_{i=1}^{n} a_i r^{i-1} = a_1 \left(\frac{1-r^n}{1-r} \right)$$

$$S = \sum_{i=0}^{\infty} a_i r^i = \frac{a_1}{1-r}$$

Newton Binomial Theorem

$$(x+y)^n = \sum_{k=0}^{n} C(n,k) x^k y^{n-k}$$

$$\frac{x^k y^{n-k}}{(x+y)^n} = \left(\frac{x}{x+y} \right)^k \left(\frac{y}{x+y} \right)^{n-k}$$

Maclaurin Series of e^x

$$e^x = \sum_{n \geq 0} \frac{x^n}{n!} = \sum_{n \geq 1} \frac{x^{n-1}}{(n-1)!}$$

Combination Rules

$$\binom{n}{k} = \binom{n}{n-k}$$

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

$$k\binom{n}{k} = (n-k+1)\binom{n}{k-1}$$

$$(n-k)\binom{n}{k} = n\binom{n-1}{k}$$

$$n\binom{n}{k} = n\binom{n-1}{k-1}$$

$$\binom{n+m}{k} = \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-1}, \text{ Vandermonde equality}$$

Balls and Bags

$$k_1 + ... + k_i = n, k \in \mathbb{N}_0, n \in \mathbb{N}$$

$$\binom{n+i-1}{i-1}$$

Enumeration (Combinatorics)

Combinatorics is an art of counting sets.

Cardinal

The cardinal of a set A is a number of elements contained in A.

Types of Sets

- Countable set: can list all elements in a sequence indexed by integers.
  - Finite set: the number of elements in the set is finite. Example: $A = \{1, 2, 3, 6, 7\}, card(A) = 5$.
  - Infinite set: the number of elements in the set is infinite. Example: set of $\mathbb{N} = \{1, 2, 3, ...\}$
- Uncountable set: set that cannot be indexed by the sequence of integers. Example: $\mathbb{R}, \{0,1\}^{\mathbb{N}} = 0010101..., 0101101111..., \mathcal{P}(\mathbb{N})$.

Basic Rules

Bijection

In set theory, a bijection is a function between two sets that establishes a one-to-one correspondence between their elements. More formally, a function $f : A \to B$ is considered a bijection if the following conditions are met:

1. Injective (One-to-One): For every pair of distinct elements x and y in set A, $f(x)$ and $f(y)$ are also distinct in set B. In other words, no two elements in set A map to the same element in set B.

2. Surjective (Onto): For every element y in set B, there exists an element x in set A such that $f(x) = y$. In other words, the function f covers all elements in set B.

3. Bijective: If a function f is both injective and surjective, it is called a bijection. This means that there is a one-to-one correspondence between the elements of sets A and B, and every element in set B has a unique pre-image in set A.

If A and B are in bijection, then

$$card(A) = card(B)$$

Sets

- Product: Let A, B be two finite sets. We define the product $A \cdot B$ as $A \cdot B = \{(a,b) | a \in A, b \in B\}$ and

$$card(A \cdot B) = card(A) \cdot card(B)$$

- Disjoint Let A, B be two subsets of a set Ω. If A, B are disjoint, then

$$card(A \cup B) = card(A) + card(B)$$

- Joint: If A, B are joint, then:

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

Since $card(A) + card(B)$ counts $card(A \cap B)$ two times, we subtract one $card(A \cap B)$.

Indicator Function

Let Ω be a set and $A \subset \Omega$. We denote by 1_A the indicator function of the set A s.t.

$$1_A : \Omega \to \{0, 1\}$$

$$\omega \to \begin{cases} 1, \omega \in A \\ 0, \omega \notin A \end{cases}$$

Power Set

Let Ω be a set. We denote by $\mathcal{P}(\Omega)$ the power set of Ω. It is defined by $\mathcal{P}(\Omega) = \{A | A \subset \Omega\}$ (nested sets).

Cardinal of Power Set

$$card(\mathcal{P}(\Omega)) = 2^{card(\Omega)}$$

Proof:
The set $\mathcal{P}(\Omega)$ can be put in bijection with the set $\{0,1\} \times \{0,1\} \times ... \times \{0,1\} = \{0,1\}^{\Omega} = \{0,1\}^{card(\Omega)}$

$\mathcal{P}(\Omega) \to \{\{0,1\} \times \{0,1\} \times ... \times \{0,1\}\}$

We map each element of Ω to either 0 or 1. Thus:
$card(\mathcal{P}(\Omega)) = card(\{0,1\}^{card(\Omega)}) = card(\{0,1\})^{card(\Omega)}$
$2^{card(\Omega)}$

Permutation

A set Ω is not ordered, there is no repetition:
$\{a,b,c\} = \{b,c,a\} = \{a,c,b\} = \{a,b,c,b,c,a\} = ...$

A permutation is a way to order the elements of Ω. Alternatively, it is a bijetion from Ω to $\{1,2,3,...,n\}$ where $n = card(\Omega)$

If Ω has n elements, then there are $n!$ number of permutations of Ω.

$(w_1, w_2, w_3, ..., w_n)$, n possible ways to choose w_1, $n-1$ possible ways to choose w_2, $\ldots$, 1 possible way to choose $w_n = n!$

Question: how many anagrams of the word MISSISSIPPI ? $\frac{11!}{4!4!2!}$

Arrangement

An arrangement of k elements from n elements of a set Ω is an ordered sequence of k distinct elements of Ω. In arrangement, the order of selection matters. There are

$$A_n^k = n(n-1)(n-2)...(n-k+1) = \frac{n!}{(n-k)!}$$

ways to arrange k elements among n elements.

Combination

A combination of k elements from n elements of Ω is a subset of Ω with k elements. In combination, the order does not matter ($\{1,2\} = \{2,1\}$).

$$C(n,k) = \frac{n!}{k!(n-k)!}$$

Proof:
Given a combination of k elements, we can form exactly $k!$ distinct arrangements by permuting the k elements chosen. This gives the equality:

$A_n^k = \frac{n!}{(n-k)!}$

$k!\binom{n}{k} = A_k^n = \frac{n!}{k!(n-k)!}$

Example:
$\Omega = \{1,2,3\}$
$k = 2, C(n,k) = \{\{1,2\},\{1,3\},\{2,3\}\}$ - the number ~~of~~ combinations of k elements among a set of n elements ($C(n,k)$ is called "n" choose "k").

Probability Space

A probability space is a triplet (Ω, F, P) where: - Ω is the universe, set of possible configurations of the experiment, simply a set; - F is a σ-algebra on Ω. F represents the information we can acquire during the experiment. An element in F is called an **event**. An event is the subset of Ω universe. Therefore, F is a set of events. It is possible to apply certain operations between events : union, intersection, difference, complementary, etc. In other words, the set F is stable by a number of operations; - P is a probability measure. It is used to quantify the probability of a given event occurring. For a given event A, it associates a number between 0 and 1, denoted $P(A)$, which reflects the probability of the event A occurring.

Measurable Space

The pair (Ω, F) is a measurable space (events in F are measurable).

sigma-algebra

A σ-algebra F on a space Ω is a subset of $P(\Omega)$ (the power set of Ω) such that:

1. $\Omega \in F$ (the universe is an event)
2. If $A \in F$, then $\overline{A} \in F$ (stability by passing to the complementary)
3. If $(A_n)_{n\geq 0} \in F$, then $\bigcup_{n\geq 0} A_n \in F$ (stability by countable union)

σ-algebra - everything you can construct with union, intersection, complementary, ... of events

Trivial / Discrete sigma-algebras

The σ-algebra $\{\emptyset, \Omega\}$ is called the trivial σ-algebra. It is the smallest σ-algebra on Ω that we can consider. The σ-algebra $P(\Omega)$ is called the discrete σ-algebra. It is the largest σ-algebra on Ω that can be considered.

Smallest sigma-algebra

Let C be a subset of $P(\Omega)$. Let $\sigma(C)$ be the smallest σ-algebra containing C. It is the intersection of all σ-algebras containing C. For example, if Ω is countable and C is the set of singletons, it's easy to see that $\sigma(C) = P(\Omega)$.

Let $\Omega = [0, 1]$. If I is an interval with ends a and b (not necessarily open or closed), then $\mu(I) = b - a$. This definition is "consistent" with the notion of a probability space: $\mu(\Omega) = 1$.

I_1 and I_2 are disjoint intervals such that $I_1 \cup I_2$ is an interval. E.g.: $I_1 = [a, b]$ and $I_2 = [b, c]$, then: $\mu(I_1 \cup I_2) = \mu(I_1) + \mu(I_2)$

Borelian sigma-algebra

Let $B([0, 1])$ be the σ-algebra generated by the intervals included in $[0, 1]$. It is called the Borelian σ-algebra on $[0, 1]$. Similarly, $B(\mathbb{R})$ is the σ-algebra generated by the intervals in $\mathbb{R}$.

$B([0, 1]) \neq \mathcal{P}([0, 1]), \mathcal{P}([0, 1])$ is much bigger.

$C = \{[a, b] | a, b \in [0, 1]\}$
$B([0, 1]) = \sigma(C)$

Probability Measure

A probability measure P on a measurable space (Ω, F) is an application:

$P : F \to [0, 1]$ $A \to P(A)$

Such that:

1. $P(\Omega) = 1$ (the universe is an event with probability 1)
2. If $(A_n)_{n\geq 0}$ is a countable family of pairwise disjoint events, then: $P(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} P(A_n)$

Almost-Surely / Negligible

If A is an event in a probability space (Ω, F, P) such that $P(A) = 1$, we say that A is realized **almost-surely** (sometimes denoted as a.s.). Conversely, if A is an event in a probability space (Ω, F, P) such that $P(A) = 0$, we'll say that A is a **negligible** event.

Conditional Probability and Independence

When we want to model two distinct quantities by a probabilistic model, it often happens that these two quantities are **correlated**.

Conditional Probability

Let (Ω, F, P) be a probability space, and B an event of non-zero measure. Then for any event A, we call the quantity:

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

It indicates the probability of an event A occurring knowing that event B has occurred.

$$(\Omega, F, P) \to (B, \mathcal{F}_B, P(\cdot|B))$$

$$\mathcal{F}_B = \{A \cap B \mid A \in F\}$$

Complete System of Events

A family $\{B_1, \ldots, B_n\}$ is said to be a complete system of events if:

- $\forall i \in \{1, \ldots, n\}, P(B_i) \neq 0$ $(B_i \neq)$
- $\forall i, j \in \{1, \ldots, n\}, P(B_i \cap B_j) = 0$ $(B_i \cap B_j =)$
- $P\left(\bigcup_{i=1}^n B_i\right) = 1$ or equivalently $\sum_{i=1}^n P(B_i) = 1$ $(=> \cup B_i = \Omega)$

Law of Total Probability (LTP)

Let $\{B_1, \ldots, B_n\}$ be a complete system of events, and A be any event. Then:

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A|B_i)$$

Example:
We have 2 classes: A and B. We are interested in probability of a students passing.
$P(Pass|A) = 0.8; P(Pass|B) = 0.6; P(A) = 0.6$
Thus, by LTP we have:
$P(Pass) = P(A)P(Pass|A) + P(\overline{A})P(Pass|\overline{A}) = P(A)P(Pass|A) + (1 - P(A))P(Pass|B) = 0.6 * 0.8 + (1 - 0.6)0.6 = 0.72$

Bayes' Formula

Let A and B be two events of non-zero probability. Then we have the identity:

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

If the family $\{A_1, \ldots, A_n\}$ is a complete system of events, then for $1 \leq i \leq n$ we have:

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

Example:
$P(Disease) = P(D)$ - probabiltiy that randomly selected person is sick.
$P(P|D)$ - probablity of positive testing given sick person.
$P(P|ND)$ - probability of positive testing given healthy person.
We have (D, ND) - compelete system of events. Thus:
$P(D|P) = \frac{P(D)P(P|D)}{P(D)P(P|D)+P(ND)P(P|ND)} = \frac{P(D)P(P|D)}{P(D)P(P|D)+(1-P(D))P(P|ND)}$

Independence

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

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

$\mathcal{F}_1 \perp \mathcal{F}_2$ if $\forall A \in \mathcal{F}_1, \forall B \in \mathcal{F}_2 : P(A \cap B) = P(A)P(B)$

Mutual Independence: Events | Family of Events

Events $A_1, \ldots, A_n$ are said to be mutually independent if for any subset I of $\{1, \ldots, n\}$ we have:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Similarly, the σ-algebras $F_1, \ldots, F_n$ are mutually independent if for any events $A_1, \ldots, A_n$ such that: $A_1 \in F_1, \ldots, A_n \in F_n$, the events $A_1, \ldots, A_n$ are independent.

Same with a family $(A_n)_{n \in \mathbb{N}}$ of events.

Let $F_1, \ldots, F_n, G_1, \ldots, G_m$ be mutually independent σ-algebras. Then:

$$\sigma\left(\bigcap_{i=1}^{n} F_i\right) \perp \sigma\left(\bigcap_{j=1}^{m} G_j\right)$$

Example:
Let A_1, A_2, A_3, A_4, A_5 be mutually independent events. Then:

$(A_1 \cup A_3) \perp (A_2 \cap (A_4 \cup A_5))(\sigma - \text{algebra operations})$

Independence vs Disjoint

- Independence: Two events A and B are said to be independent if: $P(A \cap B) = P(A) \cdot P(B)$. In simple terms, if two events A and B are independent, the occurrence of one event does not provide any information about the occurrence or non-occurrence of the other event. They are unrelated in terms of probability.
- Disjoint (Mutually Exclusive): Two events A and B are said to be disjoint or mutually exclusive if they cannot occur at the same time. In other words, if one event happens, the other cannot happen simultaneously. Mathematically, A and B are disjoint if $A \cap B = \emptyset$, where $\emptyset$ represents the empty set. Disjoint events are not independent because if one event occurs, it implies that the other event cannot occur.

Limit Theorems

Limit Superior | Limit Inferior

Let $(A_n)_{n \geq 0}$ be a sequence of sets. The limit superior of the sequence is defined as:

$$\limsup_{n \geq 0} A_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k$$

The limit inferior of the sequence is defined as:

$$\liminf_{n \geq 0} A_n = \bigcup_{n \geq 0} \bigcap_{k \geq n} A_k$$

The interpretations of these two quantities are as follows:

- An element $\omega \in \Omega$ belongs to the set $\limsup_{n \geq 0} A_n$ if and only if ω belongs to an infinite number of events in the sequence $(A_n)_{n \geq 0}$.
- An element $\omega \in \Omega$ belongs to the set $\liminf_{n \geq 0} A_n$ if and only if ω belongs to all events of the sequence $(A_n)_{n \geq 0}$ starting from a certain rank (from some point onward).

Random Variable

A random variable models the different values that the outcome of a random experiment can take.

- The roll of the dice. Consider the random variable $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$ the application to which a given configuration associates the value of the dice.

For coin game, we have X mapping:

$$X : \Omega \to \{\text{"}Head\text{"}, \text{"}Tail\text{"}\}, \omega \to x(\omega)$$

We have to ensure that $\{\omega \in \Omega : x(\omega) = \text{"}Tail\text{"}\}$ (*Notation:* $\{\omega \in \Omega : x(\omega) = \text{"}Tail\text{"}\} = \{X = \text{"}Tail\text{"}\}$) is an event, i.e. an element of the σ-algebra $\mathcal{F}$. Indeed, the proposition "the probability that the coin lands on tails is 1/2" is written mathematically:

$$\mathbb{P}(X = \text{"}Tails\text{"}) = 1/2$$

Countable Space

If X takes value in a countable (finite) space we require that for any x in E the set

$$\{X = x\} = \{\omega \in \Omega \,|\, X(\omega) = x\}$$

is an event (i.e., belongs to the σ-algebra $\mathcal{F}$). Since any subset U of E can be written as a countable union of singletons, then:

$$\{X \in U\} = \bigcup_{x \in U} \{X = x\}$$

$$=> \forall U \subset E, \{X \in U\} \subset \mathcal{F}$$

Uncountable Space (X is Real-Valued)

If we ask only that the sets $\{X = x\}$ are events, we can't assert that the set:

$$\{X \in [0, 1]\} = \bigcup_{x \in [0,1]} \{X = x\}$$

is an event, since a σ-algebra is only stable by countable union, and the set $[0, 1]$ is uncountable. In general, we require that for any $x \in \mathbb{R}$, the following set is an event:

$$\{X \leq x\} = \{X \in (-\infty, x]\}$$

So the intersection, union and complement of events are still events:

$$\{X < x\} = \bigcup_{n \geq 0} \left\{X \leq x - \frac{1}{n}\right\}$$

$$\{X > x\} = \{X \leq x\}$$

$$\{X \geq x\} = \{X < x\}$$

$$\{x \leq X \leq y\} = \{X \in [x,y]\} = \{X \geq x\} \cap \{X \leq y\}$$

$$\{X = x\} = \{x \leq X \leq x\}$$

Discrete Random Variable

Let X be an application of a probability space $(\Omega, \mathcal{F}, P)$ with values in a countable space E, provided with the discrete σ-algebra. This means

$$X : (\Omega, \mathcal{F}, P) \to (E, P(E))$$

We say that X is a discrete random variable with values in E, if for any $e \in E$, the set $\{X = e\} = \{\omega \in \Omega | X(w) = e\}$ is an event of the σ-algebra $\mathcal{F}$.

Real Random Variable

Let X be an application of a probability space $(\Omega, \mathcal{F}, P)$ with values in $\mathbb{R}$, provided with the Borelian σ-algebra $\mathcal{B}(\mathbb{R})$ ($B(\mathbb{R})$ is the σ-algebra generated by the intervals in $\mathbb{R}$), i.e.

$$X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

We say that X is a real random variable if for any $x \in \mathbb{R}$, the set $\{X \leq x\} = \{\omega \in \Omega | X(w) \leq x\}$ is an event of the σ-algebra $\mathcal{F}$.

Distribution of Random Variables

The Distribution of a Random Variable

Distribution of Discrete Random Variable

Let X be a discrete random variable with values in a countable space E. The distribution of X is entirely characterized by the quantities for $e \in E$ (X is a collection of quantities $\mathbb{P}(X = e)$):

$$P(X = e)$$

Distribution of Real Random Variable

Let X be a real random variable. The distribution of X is entirely characterized by the quantities for $x \in \mathbb{R}$:

$$P(X \leq x)$$

Cumulative Distribution Function (CDF)

Let X be a real random variable. The function F is called the cumulative distribution function of the variable X:

$$F_X : \mathbb{R} \to [0,1], x \mapsto P(X \leq x)$$

For $a < b$ and if F does not have atoms:

$P(X \in [a,b]) = P(X = a)_{=0 \text{ (no atoms)}} + P(X \in (a,b]) = P(X \in (a,b]) = P((X \leq b) \cap (X > a)) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$

CDF Properties

The cumulative distribution function F_X of a random variable verifies the following properties:

- F_X is increasing and:
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to +\infty} F_X(x) = 1$
- F_X is continuous on the right: If $(x_n)_{n \geq 0}$ is a decreasing sequence that converges to a real number x, then:
  - $\lim_{n \to +\infty} P(X \leq x_n) = P(X \leq x)$
- F_X admits a left limit at any point: If $(y_n)_{n \geq 0}$ is an increasing sequence that converges to a real number y, then:
  - $\lim_{n \to +\infty} P(X \leq y_n) = P(X < y)$

A function F verifying the last to points is said to be *cadlag* (continious on the right, limit on the left).

Proof Let $x, y \in \mathbb{R}$ be real points s.t. $x \leq y$. Then:

$$\{X \leq x\} \subset \{X \leq y\}$$

$$P(X \leq x) \leq P(X \leq y)$$

Furthermore, let $(x_n)_{n \in \mathbb{N}}$ be a sequence increasing towards infinity. The events $(\{X \leq x_n\})_{n \in \mathbb{N}}$ form an increasing sequence of events whose union is $\mathbb{R}$ any number. Hence:

$$\lim_{n \to +\infty} F_X(x_n) = \lim_{n \to +\infty} P(X \leq x_n) =$$

$$= P\left(\bigcup_{n=0}^{\infty}\{X \le x_n\}\right) = P(X \in \mathbb{R}) = 1$$

The case for $-\infty$ is similar.

Let $(x_n)_{n\in\mathbb{N}}$ be a decreasing sequence towards a real x. Then:

$$\lim_{n\to+\infty} F_X(x_n) = \lim_{n\to+\infty} P(X \le x_n) =$$

$$= P\left(\bigcup_{n=0}^{\infty}\{X \le x_n\}\right) = P(X \le x) = F_X(x)$$

Let $(y_n)_{n\in\mathbb{N}}$ be a sequence increasing towards a real y.

$$\lim_{n\to+\infty} F_X(y_n) = \lim_{n\to+\infty} P(X \le y_n) =$$

$$= P\left(\bigcup_{n=0}^{\infty}\{X \le y_n\}\right) = P(X < y)$$

Quantile Function

Let F be a function verifying the three properties of CDF. We define the quantile function $F^{-1} : [0,1] \to \mathbb{R}$ by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} \mid F(x) \ge u\}$$

(in general: $F(x) = y \iff x = F^{-1}(y)$)

Let U be a uniform distribution on $[0,1]$. The quantile function F^{-1} verifies the following properties: - If F is continuous and increasing, then F^{-1} is the reciprocal bijection of F. - The random variable $X = F^{-1}(U)$ has the cumulative distribution function F.

Atom

Let $x \in \mathbb{R}$. A real random variable X (usually discrete) is said to have an atom at the point x if:

$$P(X = x) > 0$$

CDF and Atom | Jump of Discontinuity

Let X be a real random variable. The cumulative distribution function of X exhibits a jump of discontinuity at the point x if and only if X has an atom at the point x. In this case, the jump is of size $P(X = x)$.

Probability Density Function (PDF)

Let X be a real random variable with a cumulative distribution function F_X. The random variable X is said to have a density if there exists an integrable function f_X such that

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

The function F_X is a primitive of the function f_X. We deduce that F_X is continuous, so X is atom-free.

CDF and PDF

Let X be a real random variable. If its cumulative distribution function F_X is continuous and piecewise derivable, then X is a random variable with probability density f and:

$$f_X = (F_X)' = \frac{d}{dx}F_X$$

Conversely, if X is a random variable that has a density, its characteristic function is continuous, so X has no atom. On the other hand, there are atom-free random variables that do not admit a density (Cantor's staircase is a counterexample).

Let a and b be real numbers with $a < b$, and let X be a random variable with density f_X. Then:

$$P(a \le X \le b) = \int_{a}^{b} f_X(t)\,dt$$

If now $I = [x, x+\varepsilon]$ is a small interval, then we have:

$$P(x \le X \le x+\varepsilon) = \int_{x}^{x+\varepsilon} f_X(t)\,dt \approx \varepsilon f_X(x)$$

Let X be a random variable with density f_X. The probability that X belongs to a small interval around a point $x \in \mathbb{R}$ is proportional to the size of this interval. The proportionality coefficient is exactly $f_X(x)$.

PDF: Positive and Mass

Let X be a random variable with density f_X. Then:

- f_X is positive.
- f_X has mass 1:

$$\int_{-\infty}^{\infty} f_X(t)\,dt = 1$$

If f is a function verifying these two properties, then there exists a random variable X of density f.

Demonstration (Proof)

For the first point, we have for ε a small positive real:

$$0 \leq P(x \leq X \leq X + \varepsilon) \approx \varepsilon f(x)$$

For the second point, we have:

$$1 = P(-\infty < X < +\infty) = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(t)\,dt$$

Expectation, Variance, Median

Expectation

Expectation Properties

Let X and Y be two integrable random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and λ a real number. Then,

- **Linearity** : For $\lambda \in \mathbb{R}$, we have

$$E[\lambda X + Y] = \lambda E[X] + E[Y]$$

- **Monotonicity** : If, for almost all $\omega \in \Omega$, we have $X(\omega) \leq Y(\omega)$, then

$$E[X] \leq E[Y]$$

- $E[1_A] = P[A]$
- $E[C] = C$.

Expectation Discrete

$$\sum_{n=0}^{\infty} |x_n| P(X = x_n) < +\infty$$

$$E[X] = \sum_{n=0}^{\infty} x_n P(X = x_n)$$

Expectation Density

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < +\infty$$

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

Transfer Theorem

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, and g be a function from $\mathbb{R}$ to $\mathbb{R}$. Then, the quantity $E[g(X)]$ depends only on the function g and the distribution of X.

$$E[g(X)] = \sum_{n=0}^{\infty} g(x_n) P(X = x_n)$$

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

Markov's Inequality

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}, \quad \alpha > 0$$

$$P(X \geq \alpha) \leq \frac{E[g(X)]}{g(\alpha)}, \quad g \text{ is positive and strictly increasing}$$

Median

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2}$$

Variance

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

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

$$\sigma(X) = \sqrt{\mathrm{Var}(X)} - \text{ standard deviation}$$

$$Var(X)(\text{discrete}) = \left(\sum_{n=1}^{+\infty} x_n^2 P(X = x_n)\right) - \left(\sum_{n=1}^{+\infty} x_n P(X = x_n)\right)^2$$

$$Var(X)(\text{real}) = \left(\int_R t^2 f_X(t) dt\right) - \left(\int_R t f_X(t) dt\right)^2$$

Let X be a random variable with finite variance, and λ a real number. We have:

- $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$
- $\text{Var}(X + \lambda) = \text{Var}(X)$
- X is almost surely (a.s.) constant if and only if $\text{Var}(X) = 0$
- $\text{Var}(X) = E[X^2] - E[X]^2$

Chebyshev's Inequality

$$P(|X - E[X]| \geq \beta) \leq \frac{\text{Var}(X)}{\beta^2}, \quad \beta \in R \backslash 0$$

Probability Distributions

Discrete

Bermoulli $\mathcal{B}(\theta)$

$$P(X = 0) = 1 - \theta \text{ and } P(X = 1) = \theta$$

$$E[X] = \theta$$

$$\text{Var}(X) = \theta(1 - \theta)$$

Poisson $\mathcal{P}(\theta)$

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

$$E[X] = \theta$$

$$\text{Var}(X) = \theta$$

Binomial $\mathbb{B}(n, \theta)$

$$P(X = k) = \binom{n}{k}\theta^k(1 - \theta)^{n-k}$$

$$\text{if } S_n = \sum_{i=1}^{n} X_i, X_i \sim \mathcal{B}(\theta) => S_n \sim \mathbb{B}(n, \theta)$$

$$E[X] = n\theta$$

$$\text{Var}(X) = n\theta(1 - \theta)$$

Uniform $\mathbb{U}(a, b)$

$$n = b - a + 1$$

$$k \in \{1, ..., n\}$$

$$P(X = k) = \frac{1}{n}$$

$$E[X] = \frac{n + 1}{2}$$

$$\text{Var}(X) = \frac{n^2 - 1}{12}$$

Geometric $\mathcal{G}(\theta)$

$$k \in N$$

$$P(X = k) = \theta(1 - \theta)^{k-1}$$

$$P(X \leq k) = 1 - (1 - \theta)^k$$

$$P(X > k) = (1 - \theta)^k$$

$$E[X] = \frac{1}{\theta}$$

$$\text{Var}(X) = \frac{1 - \theta}{\theta^2}$$

Density

Normal $\mathcal{N}(\theta, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x - \theta)^2)$$

$$\theta = E[X]$$

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

Exponential (Memoryless) $\varepsilon(\lambda)$

$$P(X \leq x) = 1 - e^{-\lambda x} \cdot 1_{[0, +\infty]}$$

$$P(X \geq x) = e^{-\lambda x} \cdot 1_{[0, +\infty]}$$

$$P(X \geq x + s | X \geq x) = P(X \geq s)$$

$$f_X(t) = \lambda e^{-\lambda t} 1_{[0, +\infty]}(t)$$

$$E[X] = \frac{1}{\lambda}$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Uniform $\mathcal{U}([a, b])$

$$a \leq c \leq d \leq b$$

$$P(c < X < d) = \frac{d - c}{b - a}$$

$$f_X(x) = \frac{1}{b - a} 1_{[a, b]}$$

$$E[X] = \frac{b + a}{2}$$

$$\text{Var}(X) = \frac{(b - a)^2}{12}$$

Pair of Random Variables

Discrete

Joint Law

$$P(X = x_k, Y = y_j)$$

If you know the distribution of the pair (X, Y), then you can deduce the Marginal Distributions of X and Y.

Marginal Law

$$P(X = x_k) = \sum_j P(X = x_k, Y = y_j)$$

$$P(Y = y_j) = \sum_k P(X = x_k, Y = y_j)$$

Knowing marginals does not allow us to compute the joint, unless we add some hypothesis, like independence.

Continious Case

Joint Law

$$P(X \leq x, Y \leq y)$$

Joint Density

$$P(X \in I, Y \in J) = \iint_{I \times J} f(X,Y)(x,y)\,dx\,dy$$

If you have joint density, you can compute marginal densities.

Marginal Density

$$f_X(x) \to \int_R f_{(X,Y)}(x,y)\,dy$$

$$f_Y(y) \to \int_R f_{(X,Y)}(x,y)\,dx$$

Independent Random Variables

$\forall A \in G, \forall B \in H, P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$

Discrete Case

$$P(X = x_k, Y = y_j) = P(X = x_k)P(Y = y_j)$$

Continious Case

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

Joint Density

$$f_{(X,Y)}(x,y) \to f_X(x)f_Y(y)$$

Expected Values

$$E[XY] = E[X]E[Y]$$

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)], f : R \to R, g : R \to R$$

$$E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n]$$

Variances

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X + Y) = E[(X - E[X]) + (Y - E[X])]^2 =$$

$$= Var(X) + Var(Y) - 2E[(X - E[X]) + (Y - E[X])] =$$

$$= Var(X) + Var(Y) - 2Cov(X, Y) = Var(X) + Var(Y)$$

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)$$

Covariance

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

PMF ($Z = X + Y, X \perp Y$)

Discrete

$$P(Z = n) = \sum_{k=0}^{n} P(X = k)P(Y = n - k)$$

Continious

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

Limit Theorems

Modes of Convergence of Random Variables

Almost Sure Convergence

We say that $(X_n)_{n \geq 0}$ converges to X **almost surely** if for almost every $\omega \in \Omega$

$$\lim_{n \to +\infty} X_n(\omega) = X(\omega)$$

$$X_n \to^{a.s.}_{n \to +\infty} X$$

A simple case of almost sure convergence occurs when dealing with an **increasing** and **bounded sequence** of random variables. In that case, it converges almost surely to a limiting random variable.

Convergence in Law L^p

We say that $(X_n)_{n\geq 0}$ converges to X **in Law** L^p if

$$\lim_{n\to+\infty} E[|X_n - X|^p] = 0$$

$$X_n \to^{L^p}_{n\to+\infty} X$$

When $p = 2$ it is called quadratic convergence.

In other words, as we compute distances for each ω for $(X_n)_{n\geq 0}$ and X, each ω should be close for $E[|X_n - X|^p]$ to go to 0.

Convergence in Probability

We say that $(X_n)_{n\geq 0}$ converges to X **in probability** if, for every $\varepsilon > 0$

$$\lim_{n\to+\infty} P[|X_n - X| > \varepsilon] = 0$$

$$X_n \to^{P}_{n\to+\infty} X$$

In other words, a sequence of random variables converges in probability to its limit when the probability that the sequence takes values far from its limit tends to 0.

Convergence in Distribution

We say $(X_n)_{n\geq 0}$ (with distribtuion functions $(F_n)_{n\geq 0}$) **converges in distribution** to X (with distribution function F) if, at every continuity point x of function F we have

$$\lim_{n\to+\infty} F_n(x) = F(x)$$

$$X_n \to^{L}_{n\to+\infty} X$$

In other words, if $(X_n)_{n\geq 0}$ is Gaussian, or Bernoulli, or etc., then X is distributed the same way.

Instead of X we can put a distribution, for example

$$X_n \to^{L}_{n\to+\infty} \mathcal{N}(0,1)$$

Example $(X_n)_{n\geq 0}$ can converge in **distribution** to X, but not converge in other modes of convergence. Take sequence $(X, -X, X, -X, ...)$. We have Gaussian in every X_i, so sequence converges to Gaussian. Convergence in distribution generally does not imply other modes of convergence. Convergence in distribution is **weak mode of convergence**.

Convergence to Constant

$$\text{if } (X_n)_{n\geq 0} \to^{L}_{n\to+\infty} C, C \in R$$
$$=> (X_n)_{n\geq 0} \to^{P}_{n\to+\infty} C$$

Convergence Implications

- Almost Sure Convergence => Convergence in Probability => Convergence in Distribution
- Convergence in L^p => Convergence in Probability => Convergence in Distribution

LLN | Monte Carlo | CLT

Independent and Identically Distributed Sequence of Random Variables (i.i.d.)

If $(X_n)_{n\geq 1}$ are i.i.d., then they all have the same distribution.

Empirical Mean

We define the sequence of empirical means $\overline{(X_n)_{n\geq 1}}$ of i.i.d. random variables as

$$\forall n \geq 0 \quad \overline{X_n} = \frac{X_1 + ... + X_n}{n}$$

$$\sqrt{Var(\overline{X_n})} = \frac{\sqrt{Var(X)}}{\sqrt{n}}$$

$$Var(\overline{X_n}) = \frac{Var(X)}{n}$$

Law of Large Numbers

If $(X_n)_{n\geq 1}$ are i.i.d., distributed the same way as X, and expectation of X is finite, then

$$\overline{X_n} \to^{\text{a.s.}}_{n\to+\infty} E[X]$$

For example, the law of large numbers asserts that, on average, after a large number of dice rolls, I would have rolled a 5 approximately one in six times.

Monte Carlo

An application of the law of large numbers is the Monte Carlo method for estimating the value of an integral. This method is particularly useful in dimensions $d \geq 1$.

Let $f : [0,1]^d \to R$ be an integrable function, and $(U_n)_{n\geq 1}$ be an i.i.d. sequence of uniform random variables on $[0,1]^d$. Then

$$\frac{f(U_1) + ... + f(U_n)}{n} \to_{n\to+\infty} \int_{[0,1]^d} f(x)dx$$

Proof by LLN

Define $X_n = f(U_n)$. The sequence $(X_n)_{n\geq 1}$ is a sequence of real r.v. i.i.d. with

$$\overline{f(U_n)} \to^{\text{a.s.}}_{n\to\infty} E[f(U)] = \int_{[0,1]^d} f(x)dx$$

Central Limit Theorem (CLT)

Assuming that the variance of X is finite, then

$$\frac{\overline{X_n} - E[\overline{X_n}]}{\sqrt{Var(\overline{X_n})}} \to^{L}_{n\to+\infty} Z = \mathcal{N}(0,1)$$

$$Var(X) = \sigma^2$$

$$E[\overline{X_n}] = E[X]$$

$$Var(\overline{X_n}) = \frac{\sigma^2}{n}$$

$$\sqrt{n}(\overline{X_n} - E[X]) \to^{L}_{n\to+\infty} \mathcal{N}(0,\sigma^2)$$

$$\sqrt{\frac{n}{\sigma^2}}(\overline{X_n} - E[X]) \to^{L}_{n\to+\infty} Z = \mathcal{N}(0,1)$$

$$\frac{(X_1 + ... + X_n) - nE[X]}{\sqrt{n}\sigma} \to^{L}_{n\to+\infty} \mathcal{N}(0,1)$$

$$\overline{X_n} \approxeq \mathcal{N}(E[X], \frac{\sigma^2}{n})$$

The CLT asserts that the error between the theoretical and empirical means is of the order of $\sqrt{Var(\overline{X_n})} = \frac{\sigma}{\sqrt{n}}$. This allows the calculation of asymptotic confidence interval.

Asymptotic Confidence Interval at level α

Let Z be a standard normal random variable, and $0 < \alpha < 1$. We define the number $q_{\alpha/2}$ such that

$$P(-q_{\alpha/2} \leq Z \leq q_{\alpha/2}) = 1 - \alpha$$

$$\lim_{n\to+\infty} P\left(\overline{X_n} - \frac{\sigma}{\sqrt{n}} q_{\alpha/2} \leq E[X] \leq \overline{X_n} + \frac{\sigma}{\sqrt{n}} q_{\alpha/2}\right) = 1 - \alpha$$

$$\lim_{n\to+\infty} P\left(\left|\frac{\overline{X_n} - E[\overline{X_n}]}{\sqrt{Var(\overline{X_n})}}\right| \leq q_{\alpha/2}\right) \approxeq 1 - \alpha$$

$$\lim_{n\to+\infty} P\left(\left|\frac{\overline{X_n} - E[X]}{\frac{1}{\sqrt{n}}\sqrt{Var(X)}}\right| \leq q_{\alpha/2}\right) \approxeq 1 - \alpha$$

Confidence interval does not exclude the risk of ...

The intersection is not empty, thus we cannot exclude the risk of ...

With probability > 95 the ... will not happen.