Probability

Derek Li

Contents

1 Review

1.1 Sequence

Theorem 1.1. $\sup(-x_n) = -\inf(x_n)$ and $\inf(-x_n) = \sup(x_n)$.

Proof. We know $\forall x_n \in \{x_n\}, \exists x$ s.t. $x \leqslant x_n \Rightarrow -x \geqslant -x_n$, i.e., $-x$ is the upper bound for $\{-x_n\}$ and x is the lower bound for $\{x_n\}$.

Besides, $\exists y$ s.t. $y = \sup(-x_n)$, i.e., $-x_n \leqslant y \leqslant -x \Rightarrow x \leqslant -y \leqslant x_n$. Hence, $-y = -\sup(-x_n)$ is the greatest lower bound for $\{x_n\}$ and wherefore

$$-\sup(-x_n) = \inf(x_n) \Rightarrow \sup(-x_n) = -\inf(x_n).$$

Similarly, we can show that $\inf(-x_n) = \sup(x_n)$. □

Theorem 1.2. $\inf\limits_{k\geqslant m} x_k \leqslant \sup\limits_{k\geqslant n} x_k, \forall m, n$.

Proof. We have

$$\inf_{k\geqslant n} x_k \leqslant x_n \leqslant \sup_{k\geqslant n} x_k.$$

Assume $m \leqslant n$, we have

$$\inf_{k\geqslant m} x_k \leqslant \inf_{k\geqslant n} x_k \leqslant x_n \leqslant \sup_{k\geqslant n} x_k.$$

Assume $m \geqslant n$, we have

$$\inf_{k\geqslant m} x_k \leqslant x_n \leqslant \sup_{k\geqslant m} x_k \leqslant \sup_{k\geqslant n} x_k.$$

Wherefore, $\inf\limits_{k\geqslant m} x_k \leqslant \sup\limits_{k\geqslant n} x_k, \forall m, n$. □

Definition 1.1 (Upper Limit)**.** We define upper limit $\overline{\lim} x_n = \limsup\limits_{n\to\infty} x_n$ as

$$\lim_{n\to\infty} \sup_{i\geqslant n} x_i = \inf_{n=1}^{\infty} \sup_{i=n}^{\infty} x_i.$$

Definition 1.2 (Lower Limit)**.** We define lower limit $\underline{\lim} x_n = \liminf\limits_{n\to\infty} x_n$ as

$$\lim_{n\to\infty} \inf_{j\geqslant n} x_j = \sup_{n=1}^{\infty} \inf_{j=n}^{\infty} x_j.$$

Theorem 1.3. $\underline{\lim} x_n \leqslant \overline{\lim} x_n$.

Proof. Since $\inf\limits_{k\geqslant n} x_k \leqslant \sup\limits_{k\geqslant n} x_k$,

$$\underline{\lim} x_n = \lim_{n\to\infty} \inf_{k\geqslant n} x_k \leqslant \lim_{n\to\infty} \sup_{k\geqslant n} x_k = \overline{\lim} x_n.$$

□

Theorem 1.4. $\sup\limits_{k\geqslant n} x_k - \inf\limits_{k\geqslant n} x_k = \sup\limits_{i,j\geqslant n} |x_i - x_j|$.

Proof. We have

$$\sup_{i \geqslant n} x_i - x_j = \sup_{i \geqslant n}(x_i - x_j),$$

for any fixed j.

Wherefore

$$\begin{aligned}\sup_{i \geqslant n} x_i - \inf_{j \geqslant n} x_j &= \sup_{i \geqslant n} x_i + \sup_{j \geqslant n}(-x_j) = \sup_{j \geqslant n}\sup_{i \geqslant n}(x_i - x_j) \\ &= \sup_{j \geqslant n}\sup_{i \geqslant n}|x_i - x_j| = \sup_{i,j \geqslant n}|x_i - x_j|.\end{aligned}$$

$\square$

Definition 1.3 (Cauchy)**.** x_n is Cauchy iff

$$\sup_{i,j \geqslant n}|x_i - x_j| \to 0,$$

as $n \to \infty$.

Theorem 1.5. If a sequence converges, it must be Cauchy.

Proof. Suppose $\lim_{n\to\infty} x_n = x$, then

$$\forall \varepsilon > 0, \exists N \text{ s.t. } n \geqslant N \Rightarrow |x_n - x| < \frac{\varepsilon}{2}.$$

Therefore, $\forall i, j \geqslant N$, we have

$$|x_i - x_j| = |x_i - x + (x_j - x)| \leqslant |x_i - x| + |x_j - x| < \varepsilon,$$

i.e., the sequence is Cauchy. $\square$

Theorem 1.6. $x = \overline{\lim} x_n = \underline{\lim} x_n \Leftrightarrow x_n \to x$.

Proof. ($\Rightarrow$) We have $\inf_{k \geqslant n} x_k \leqslant x_n \leqslant \sup_{k \geqslant n} x_k$.

Since $x = \overline{\lim} x_n = \underline{\lim} x_n$, then

$$\inf_{k \geqslant n} x_k \leqslant \lim_{n\to\infty}\inf_{k \geqslant n} x_k = x = \lim_{n\to\infty}\sup_{k \geqslant n} x_k \leqslant \sup_{k \geqslant n} x_k.$$

As a consequence,

$$|x_n - x| \leqslant \sup_{k \geqslant n} x_k - \inf_{k \geqslant n} x_k \to 0, \text{as } n \to \infty,$$

i.e., $x_n \to x$.

($\Leftarrow$) Since $x_n \to x$, then the sequence is Cauchy, i.e.,

$$\sup_{i,j \geqslant n}|x_i - x_j| = \sup_{k \geqslant n} x_k - \inf_{k \geqslant n} x_k \to 0 \text{ as } n \to \infty.$$

Therefore, $\overline{\lim} x_n = \underline{\lim} x_n = x$. $\square$

1.2 Set

Definition 1.4 (Power Set)**.** For a given set Ω, the power set is the set of all of its subsets

$$\mathcal{P}(\Omega) = \{A | A \subset \Omega\}.$$

The power set is closed w.r.t. all the usual set-theoretic operations.

Definition 1.5 (Symmetric Difference)**.** For any two sets A and B,

$$A \Delta B = (A - B) + (B - A) = A \cup B - AB.$$

Definition 1.6 (Arbitrary Unions)**.** Let $\omega \in \Omega, A_n \subset \Omega, n \in \mathbb{N}$.

$$\omega \in \bigcup_{n=1}^{\infty} A_n \text{ iff } \exists n \text{ s.t. } \omega \in A_n$$

Definition 1.7 (Arbitrary Intersections)**.** Let $\omega \in \Omega, A_n \subset \Omega, n \in \mathbb{N}$.

$$\omega \in \bigcap_{n=1}^{\infty} A_n \text{ iff } \forall n, \omega \in A_n.$$

Hence, we have

$$P(\omega \in A_n, \exists n) = P\left(\omega \in \bigcup_{n=1}^{\infty} A_n\right) \text{ and } P(\omega \in A_n, \forall n) = P\left(\omega \in \bigcap_{n=1}^{\infty} A_n\right).$$

Definition 1.8 (Infinitely Often)**.** Let $\omega \in \Omega, A_n \subset \Omega, n, N \in \mathbb{N}$.

$$\omega \in \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n \text{ iff } \forall N, \exists n \geqslant N \text{ s.t. } \omega \in \bigcup_{n=N}^{\infty} A_n.$$

1.3 Number System and Euclidean Space

With the notation of set, we can consider whole number as: $0 = \varnothing, 1 = \{\varnothing\}, 2 = \{0, 1\}, \cdots$. Therefore

$$\begin{aligned} n + 1 &= n \cup \{n\} \\ &= \{0, 1, \cdots, n-1\} \cup \{n\} \\ &= \{0, 1, \cdots, n\}. \end{aligned}$$

We can also define number systems with set:

$$\begin{aligned} &\mathbb{N} = \{1, 2, \cdots\}, \mathbb{W} = \mathbb{N} \cup \{0\}, \mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\}, \mathbb{Q} = \left\{\frac{n}{m}\middle| n \in \mathbb{Z}, m \in \mathbb{N}\right\}, \\ &\mathbb{R} = \left\{x = \lim_{n \to \infty} r_n \middle| r_n \in \mathbb{Q}, n \in \mathbb{N}\right\}, \mathbb{C} = \{z = x + iy | x, y \in \mathbb{R}\}. \end{aligned}$$

In multi-variable calculus, we define

$$\mathbb{R}^n = \{\mathbf{x} | x_i \in \mathbb{R}. i = 1, \cdots, n\},$$

where $\mathbf{x} = (x_i, i = 1, \cdots, n)$ and

$$\mathbb{R}^\infty = \{\mathbf{x} = (x_i, i = 1, 2, \cdots) | x_i \in \mathbb{R}, i \in \mathbb{N}\}.$$

1.4 Function

Before we define a function, we look at the product $A \times B$ of any two sets A and B, which is defined as the set of all ordered pairs that may be formed of the elements of the first set A, with the second set B :

$$A \times B = \{(a, b) | a \in A, b \in B\}.$$

Definition 1.9 (Ordered Pairs)**.** An ordered pair is $(a, b) = \{\{a\}, \{a, b\}\}$.

Definition 1.10 (Function)**.** A function f with domain A and range B, denoted by $f : A \to B$, is any $f \subset A \times B$ s.t. $\forall a \in A, \exists! b \in B$ with $(a, b) \in f$.

From the definition, b is uniquely determined by a and we may write $b = f(a)$.

The collection of all functions from a particular domain A to a certain B is denoted by

$$B^A = \{f \subset A \times B | f : A \to B\}.$$

Definition 1.11 (Inverse Image)**.** For any function say $X : \Omega \to \mathcal{X}$, the inverse image of any $A \subset \mathcal{X}$ is defined as

$$X^{-1}(A) := \{\omega \in \Omega | X(\omega) \in A\}.$$

1.4.1 Indicator Function and Indicator Map

Definition 1.12 (Indicator Function)**.** For any $A \subset \Omega$, we define $I_A \in 2^\Omega$ by

$$I_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}.$$

Indicator function defines a bijective correspondence between subsets of Ω and their indicator functions, that is referred to as the indicator map

$$\begin{aligned} I : \mathcal{P}(\Omega) &\overset{\cong}{\to} 2^\Omega \\ A &\mapsto I_A. \end{aligned}$$

Theorem 1.7. The indicator map is bijective.

Proof. We want to show the indicator map is both injective and surjective.

(Injection) Let $I_A = I_B$, then $I_A(\omega) = I_B(\omega), \forall \omega$.

We have

$$\omega \in A \Leftrightarrow I_A(\omega) = 1 = I_B(\omega) \Leftrightarrow \omega \in B,$$

i.e., $A = B$.

(Surjection) Want to show $\forall f \in 2^\Omega, \exists A \in \mathcal{P}(\Omega)$ s.t. $I(A) = I_A = f$.

Take any $f \in 2^\Omega$ and let $A = \{\omega | f(\omega) = 1\}$. We have

$$\omega \in A \Leftrightarrow \begin{cases} f(\omega) = 1 \\ I_A(\omega) = 1 \end{cases} \Rightarrow f(\omega) = I_A(\omega), \forall \omega.$$

Hence, $f = I_A$. ∎

From the proof, we also have

$$A = f^{-1}(1) = I_A^{-1}(1).$$

Definition 1.13 (Convergence of Set). $A_n \to A$ iff $I(A_n) \to I(A)$.

Note. By the theorem, we have

$$\begin{aligned}
A_n \to A &\Leftrightarrow \overline{\lim} I(A_n) = \underline{\lim} I(A_n) = I(A) \\
&\Leftrightarrow \inf_{n=1} \sup_{k\geqslant n} I(A_k) = \sup_{n=1} \inf_{k\geqslant n} I(A_k) = I(A) \\
&\Leftrightarrow I\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = I\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) = I(A) \\
&\Leftrightarrow \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = A.
\end{aligned}$$

For any monotone sequence of sets, there will always exist a limit set:

$$\begin{aligned}
A_n \text{ is increasing} &\Rightarrow A_n \to A = \bigcup_{n=1}^{\infty} A_n, \\
A_n \text{ is decreasing} &\Rightarrow A_n \to A = \bigcap_{n=1}^{\infty} A_n.
\end{aligned}$$

Note that

$$I_{\bigcap_{n=1}^{\infty} A_n}(\omega) = \inf_{n=1}^{\infty} I_{A_n}(\omega) \text{ and } I_{\bigcup_{n=1}^{\infty} A_n}(\omega) = \sup_{n=1}^{\infty} I_{A_n}(\omega).$$

Also,

$$\sum_{n=1}^{\infty} I_{A_n}(\omega) \in \mathbb{W} \cup \{\infty\}.$$

Example 1.1. If $A_n \to A, B_n \to B, C_n \to C$. Show that $A_n(B_n - C_n) \to A(B - C)$.

1.5 Linear Algebra

Definition 1.14 (Dot Product). We define

$$\mathbf{x}^T \cdot \mathbf{y} = (x_1, \cdots, x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{n} x_i y_i.$$

Definition 1.15 (Vector Norm). $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$.

2 Probability Space

Definition 2.1 (σ-algebra)**.** Let Ω be a set, then $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is called a σ-algebra s.t.

- $\mathcal{F} \neq \varnothing$;
- if $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$, i.e., $\mathcal{F}$ is closed under complements;
- if $A_i \in \mathcal{F}, i \in \mathbb{N}$, then $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$, i.e, $\mathcal{F}$ is closed under countable unions.

Note. From the second and third condition above we know that $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{F}$ since $\bigcap\limits_{n=1}^{\infty} A_i = \left(\bigcup\limits_{n=1}^{\infty} A_i^C\right)^C$. We can also show that $\Omega, \varnothing \in \mathcal{F}$.

Definition 2.2 (Probability Measure)**.** The probability measure $P : \mathcal{F} \to [0, 1]$ is a function s.t.

- σ-additivity: $P\left(\sum\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} P(A_n)$ provided $A_iA_j = \varnothing, \forall i \neq j$.
- non-negativity: $P(A) \geqslant 0, \forall A \in \mathcal{F}$;
- normalization: $P(\Omega) = 1$.

Definition 2.3 (Probability Space)**.** A probability space is a mathematical triplet $(\Omega, \mathcal{F}, P)$, where the sample space Ω is the set of all possible outcomes, the σ-algebra $\mathcal{F}$ is a collection of all the events, and the probability measure P is a function returning an event's probability.

Theorem 2.1 (Sequential Continuity)**.** $A_n \to A \Rightarrow P(A_n) \to P(A)$.

3 Expectation, Variance and Covariance

Definition 3.1 (Expect Value)**.** We define expect value as

$$\mathbb{E}[X] = \lim_{n\to\infty} \frac{x_1 + \cdots + x_n}{n}.$$

Example 3.1. $\langle X, Y\rangle = \mathbb{E}[XY] = \lim\limits_{n\to\infty} \frac{\sum\limits_{i=1}^{n} x_i y_i}{n} = \lim\limits_{n\to\infty} \frac{\mathbf{x}^T\cdot\mathbf{y}}{n}$.

Example 3.2. $|X| = \sqrt{\mathbb{E}[X^2]} = \lim\limits_{n\to\infty} \frac{\sqrt{\sum\limits_{i=1}^{n} x_i^2}}{\sqrt{n}} = \lim\limits_{n\to\infty} \frac{\|\mathbf{x}\|}{\sqrt{n}}$.

Example 3.3. Let $\measuredangle(\mathbf{x}, \mathbf{y})$ be the angle between $\mathbf{x}$ and $\mathbf{y}$. Then

$$\cos\measuredangle(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}/n}{(\|\mathbf{x}\|/\sqrt{n})(\|\mathbf{y}\|/\sqrt{n})} \to \frac{\mathbb{E}[XY]}{\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}}.$$

Example 3.4. We have

$$\begin{aligned} r &= \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i - \overline{y})^2}} \\ &\to \frac{\mathbb{E}\left[\dot{X}\dot{Y}\right]}{\sqrt{\mathbb{E}\left[\dot{X}^2\right]}\sqrt{\mathbb{E}\left[\dot{Y}^2\right]}} = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{|X - \mathbb{E}[X]||Y - \mathbb{E}[Y]|} = \cos\measuredangle\left(\dot{X}, \dot{Y}\right). \end{aligned}$$

Definition 3.2. $\cos\measuredangle(X, Y) \overset{\text{def'n}}{=} \frac{\mathbb{E}[XY]}{|X||Y|} \overset{\text{LLN}}{=} \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \cos\measuredangle(\mathbf{x}, \mathbf{y})$.

Property 3.1. $\mathbb{E}[X]$ is closest constant to X, i.e.,

$$|X - \mathbb{E}[X]| = \inf_{t\in\mathbb{R}} |X - t|.$$

Proof. Hint: $f(t) = \sqrt{\mathbb{E}[(X - t)^2]}, g(t) = \mathbb{E}[(X - t)^2]$. □

Definition 3.3 (Covariance)**.** Define

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[\dot{X}\dot{Y}\right] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Property 3.2. Here are some properties of covariance.

(1) Bilinear: $\mathrm{Cov}\left[\sum\limits_{i=1}^{m} a_i X_i, \sum\limits_{j=1}^{n} b_j Y_j\right] = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} a_i b_j \mathrm{Cov}(X_i, Y_j)$;

(2) Symmetric: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$;

(3) Non-negative: $\mathrm{Cov}(X, X) = \mathbb{E}\left[\dot{X}\dot{X}\right] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}[X] \geqslant 0$ with equality iff $X = \mathbb{E}[X]$, with probability 1.

Definition 3.4. Define

$$\rho(X,Y) = \cos\measuredangle\left(\dot{X},\dot{Y}\right) = \frac{\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)},$$

where $\sigma(X) = \sqrt{\mathrm{Var}[X]} = \left|\dot{X}\right| = |X - \mathbb{E}[X]|$.

Theorem 3.1 (Markov Inequality)**.** Given $Z \geqslant 0, t \geqslant 0$, for any $g : [0,\infty) \to [0,\infty)$ be increasing, we have

$$P(Z \geqslant t) \leqslant \frac{\mathbb{E}[g(Z)]}{g(t)}.$$

Proof. We have $g(Z) \geqslant g(t)I(Z \geqslant t) \Rightarrow \mathbb{E}[g(Z)] \geqslant g(t)\mathbb{E}[I(Z \geqslant t)] = g(t)P(Z \geqslant t)$, and thus $P(Z \geqslant t) \leqslant \frac{\mathbb{E}[g(Z)]}{g(t)}$. □

Corollary 1 (Chebyshev Inequality)**.** We have

$$P(|X-\mathbb{E}[X]| \geqslant k) \leqslant \frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{k^2}$$

or

$$P(|X-\mathbb{E}[X]| \geqslant k\sigma) \leqslant \frac{1}{k^2}.$$

Property 3.3. $\mathrm{Var}[X] = 0 \Leftrightarrow X = \mathbb{E}[X]$, with probability 1.

Proof. ($\Rightarrow$) Suppose $\mathrm{Var}[X] = 0$, then

$$P\left(|X-\mathbb{E}[X]| \geqslant \frac{1}{n}\right) \leqslant n^2 \times 0 = 0,$$

i.e., $P\left(|X-\mathbb{E}[X] \geqslant \frac{1}{n}\right) = 0$ for all $n \Rightarrow P\left(|X-\mathbb{E}[X]| < \frac{1}{n}\right) = 1$ for all n.

Let $A_n = \left(|X-\mathbb{E}[X]| < \frac{1}{n}\right)$. As $n \to \infty$, A_n is decreasing and $A_n \to A = (|X-\mathbb{E}[X]| = 0)$.

Thus, $P(A_n) \to P(A) = P(|X-\mathbb{E}[X]| = 0) = 1 \Rightarrow P(X = \mathbb{E}[X]) = 1$.

($\Leftarrow$) Suppose $P(X = \mathbb{E}[X]) = 1$, then $P(|X-\mathbb{E}[X]^2| = 0) := P(Y = 0) = 1$.

Therefore, $\mathrm{Var}[X] = \mathbb{E}[Y] = 0 \times 1 = 0$. □

Definition 3.5 (Equality in Distribution)**.** $X \overset{d}{=} Y$ on sample space $\mathcal{X}$ iff

$$\mathbb{E}[g(X)] = \mathbb{E}[g(Y)], \forall g : \mathcal{X} \to \mathbb{R}.$$

Example 3.5. If $X \overset{d}{=} Y$, then let $g = I_A, A \subset \mathcal{X}$, then

$$P(x \in A) = \mathbb{E}[I_A(X)] = \mathbb{E}[I_A(Y)] = P(Y \in A).$$

Theorem 3.2. If $X \overset{d}{=} Y$, then $\phi(X) \overset{d}{=} \phi(Y), \forall\phi : \mathcal{X} \to \mathcal{Y}$.

Proof. Since $X \overset{d}{=} Y$, then $\mathbb{E}[g(X)] = \mathbb{E}[g(Y)], \forall g : \mathcal{X} \to \mathbb{R}$.

Let $g = h\phi, \forall h : \mathcal{Y} \to \mathbb{R}$, then we have

$$\mathbb{E}[h\phi(X)] = \mathbb{E}[h\phi(Y)], \forall h : \mathcal{Y} \to \mathbb{R},$$

i.e., $\phi(X) \overset{d}{=} \phi(Y), \forall\phi : \mathcal{X} \to \mathcal{Y}$. □

Note that $\mathbb{E} : \mathcal{R} \to \mathbb{R}^* \cup \{*\}$ defined on $\mathcal{R} = \{X : \text{Real valued random variable}\}$.

4 Probability Distribution

4.1 General Finite Discrete Distribution

Definition 4.1 (Finite Scheme)**.** A finite scheme is a name for any finite discrete distribution where

$$X \sim \begin{pmatrix} a_1, \cdots, a_N \\ p_1, \cdots, p_N \end{pmatrix},$$

i.e., $X = \sum\limits_{j=1}^{N} a_j I_{\{a_j\}}(X)$.

By the definition, we have

$$\mathbb{E}[X] = \sum_{j=1}^{N} a_j \mathbb{E}\left[I_{\{a_j\}}(X)\right] = \sum_{j=1}^{N} a_j P(X = a_j) = \sum_{j=1}^{N} a_j p_j.$$

Definition 4.2 (Cumulative Distribution Function)**.** We define

$$F(x) = P(X \leqslant x) = P(X \in (-\infty, x]) = P_X((-\infty, x]).$$

Example 4.1. $F(x + n^{-1}) = P_X((-\infty, x + n^{-1}])$. As $n \to \infty$,

$$F(x + n^{-1}) \to P_X((-\infty, x]) = F(x).$$

Example 4.2. $F(x - n^{-1}) = P_X((-\infty, x - n^{-1}])$. As $n \to \infty$,

$$F(x - n^{-1}) \to P_X((-\infty, x)) = P(X < x).$$

Definition 4.3. $F(x+) := \lim\limits_{n\to\infty} F(x+n^{-1}) = F(x)$, i.e., any distribution function is right continuous at every point x.

Definition 4.4. $F(x-) := \lim\limits_{n\to\infty} F(x - n^{-1}) = P(X < x)$.

Definition 4.5 (Probability Mass Function)**.** A probability mass function of X is $p : \mathbb{R} \to [0, 1]$ given by

$$p(x) = P(X = x) = P_X(\{x\}) = P(X \in \{x\}).$$

Property 4.1. $p(x) = F(x) - F(x-)$.

Proof. We have $p(x) = P(X \leqslant x) - P(X < x) = F(x) - F(x-)$. ☐

Property 4.2. $|\{x \in \mathbb{R} | p(x) > 0\}| \leqslant |\mathbb{N}|$.

Proof. Note that $\{x \in \mathbb{R}|p(x) > 0\} = \bigcup\limits_{n=1}^{\infty} \{x \in \mathbb{R}|p(x) > \frac{1}{n}\}$. Actually,

$$\forall n \in \mathbb{N}, |\{x \in \mathbb{R}|p(x) > \frac{1}{n}\}| < n.$$

Otherwise, $\exists A_n = \{a_1, \cdots, a_n\} \subset \{x \in \mathbb{R}|p(x) > \frac{1}{n}\}$ s.t. $P(A_n) > \frac{n}{n} = 1$. ☐

Property 4.3. F is continuous at $x \Leftrightarrow p(x) = 0$.

Proof. F is continuous at $x \Leftrightarrow F(x-) = F(x+) \Leftrightarrow F(x-) = F(x) \Leftrightarrow F(x) - F(x-) = p(x) = 0$. ☐

4.2 Lebesgue-Stieltjes Integral

Consider any $\mathbb{R}$-valued $X \geqslant 0$ and let

$$X_n = \sum_{j=1}^{n} \frac{j-1}{\sqrt{n}} I_{\left(\frac{j-1}{\sqrt{n}}, \frac{j}{\sqrt{n}}\right]}(X).$$

Then we have $0 \leqslant X_n \leqslant X$, and thus

$$\begin{aligned}
0 \leqslant X - X_n &= \sum_{j=1}^{n} \left(X - \frac{j-1}{\sqrt{n}}\right) I_{\left(\frac{j-1}{\sqrt{n}}, \frac{j}{\sqrt{n}}\right]}(X) + X I_{(\sqrt{n},\infty)}(X) \\
&\leqslant \sum_{j=1}^{n} \frac{1}{\sqrt{n}} I_{\left(\frac{j-1}{\sqrt{n}}, \frac{j}{\sqrt{n}}\right]}(X) + X I_{(\sqrt{n},\infty)}(X) \\
&= \frac{1}{\sqrt{n}} I_{(0,\sqrt{n}]}(X) + X I_{(\sqrt{n},\infty)}(X) \\
&\leqslant \frac{1}{\sqrt{n}} + X I_{(\sqrt{n},\infty)}(X) \to 0
\end{aligned}$$

as $n \to \infty$, i.e., $X_n \to X$ as $n \to \infty$.

Let $h : [0,\infty) \to [0,\infty)$ be continuous and we have $h(X_n) \to h(X) \Rightarrow \mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$, i.e.,

$$\begin{aligned}
\mathbb{E}[h(X)] &= \lim_{n\to\infty} \sum_{j=1}^{n} h\left(\frac{j-1}{\sqrt{n}}\right) P\left(\frac{j-1}{\sqrt{n}} < X \leqslant \frac{j}{\sqrt{n}}\right) \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} h\left(\frac{j-1}{\sqrt{n}}\right) \left[F\left(\frac{j}{\sqrt{n}}\right) - F\left(\frac{j-1}{\sqrt{n}}\right)\right] \\
&:= \int_{o}^{\infty} h(x) \mathrm{d}F(x),
\end{aligned}$$

which is called the Lebesgue-Stieltjes integral.

4.3 Uniform Distribution

Definition 4.6 (Finite Discrete Uniform Distribution)**.** $U \sim \mathrm{unif}(\Omega)$ with $|\Omega| < |\mathbb{N}|$ iff

$$P(U = \omega) = \frac{1}{\Omega} \Leftrightarrow P(U \in A) = \frac{|A|}{|\Omega|}.$$

Example 4.3. $U \sim \mathrm{unif}\{1, \cdots, n\}$ iff $P(U = i) = \frac{1}{n}, i = 1, \cdots, n$.

Example 4.4. Let $U \sim \mathrm{unif}\{1, \cdots, n\}$, then $-U \sim \mathrm{unif}\{-n, \cdots, -1\}$ and $n+1-U \sim \mathrm{unif}\{1, \cdots, n\}$. Hence we say $n + 1 - U \overset{\mathrm{d}}{=} U$ and thus

$$n + 1 - \mathbb{E}[U] = \mathbb{E}[U] \Rightarrow \mathbb{E}[U] = \frac{n+1}{2} = \frac{1 + \cdots + n}{n}.$$

Example 4.5. Let $U \sim \mathrm{unif}\{1, \cdots, n\}$, then $U^k \sim \mathrm{unif}\{1^k, \cdots, n^k\}$, and thus we have

$$\mathbb{E}\left[U^k\right] = \frac{1^k + 2^k + \cdots + n^k}{n} \text{ and } \mathbb{E}\left[(U-1)^k\right] = \frac{0^k + 1^k + \cdots + (n-1)^k}{n},$$

and thus

$$\mathbb{E}\left[U^k\right] - \mathbb{E}\left[(U-1)^k\right] = n^{k-1}.$$

Recall that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Therefore,

$$\mathbb{E}\left[U^3\right] - \mathbb{E}\left[(U-1)^3\right] = n^2 = \mathbb{E}\left[U^3\right] - \left(\mathbb{E}\left[U^3\right] - 3\mathbb{E}\left[U^2\right] + 3\mathbb{E}\left[U\right] - 1\right),$$

i.e.,

$$3\mathbb{E}\left[U^2\right] = n^2 + 3\mathbb{E}\left[U\right] - 1 = n^2 + \frac{3(n+1)}{2} - 1 = \frac{(n+1)(2n+1)}{2}.$$

Thus,

$$\mathbb{E}\left[U^2\right] = \frac{(n+1)(2n+1)}{6} = \frac{1^2+2^2+\cdots+n^2}{n}.$$

Example 4.6. Let $U \sim \text{unif}\{1, \cdots, n\}$, then

$$\text{Var}[U] = \mathbb{E}\left[U^2\right] - (\mathbb{E}[U])^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.$$

Definition 4.7 (Standard Uniform)**.** $Z \sim \text{unif}(p)$, where $p = \{0, 1, \cdots, p-1\}$, iff

$$P(Z = i) = \frac{1}{p}, \forall i \in p.$$

$U \sim \text{unif}[0,1]$ iff

$$P(U \leqslant u) = u, \forall 0 \leqslant u \leqslant 1.$$

Example 4.7. $U \sim \text{unif}[0,1] \Leftrightarrow [nU] \sim \text{unif}\{0, \cdots, n-1\}, \forall n.$

Proof. $(\Rightarrow)$ $P([nU] = k) = P(k \leqslant nU < k+1) = P(\frac{k}{n} \leqslant U < \frac{k+1}{n}) = \frac{1}{n}$, provided $k = 0, \cdots, n-1$.

$(\Leftarrow)$ Consider $P(U < r), \forall r \in \mathbb{Q}[0,1)$. We have

$$\begin{aligned} P(U < r) &= P(nU < k), \text{for } r = \frac{k}{n} \text{ and } k < n \\ &= \sum_{i=0}^{n-1} P(nU < k, [nU] = i) \\ &= \sum_{i=1}^{n-1} P(0 \leqslant nU < k, i \leqslant nU < i+1) \\ &= \sum_{i=0}^{k-1} P(i \leqslant nU < i+1) = \sum_{i=0}^{k-1} P([nU] = i) = \frac{k}{n} = r. \end{aligned}$$

Besides, $\forall 0 \leqslant u < 1, \exists \{r_n\} \in \mathbb{Q}[0,1)$ s.t. r_n is decreasing to u and thus $(0, r_n) \to (0, u]$ and

$$P(U \leqslant u) = \lim_{n\to\infty} P(U < r_n) = \lim_{n\to\infty} r_n = u.$$

□

4.3.1 Fundamental Theorem of Applied Probability

For any $p \in \mathbb{N}$ with $p \geqslant 2$ we define the p-adic series

$$U = \sum_{i=1}^{\infty} Z_i p^{-i}.$$

Example 4.8. $Z : z_{11}, z_{12}, \cdots, z_{1n}, \cdots; z_{21}, z_{22}, \cdots, z_{2n}$ and
$U = .z_{11}z_{12}z_{13}\cdots, .z_{21}z_{22}z_{23}\cdots$.

If $Z \sim \text{unif}(10)$, then $.z_1z_2\cdots z_n\cdots = \sum\limits_{i=1}^{\infty} z_i 10^{-i}$.

Lemma 4.1. Let $\dot{p}^{\infty} = \{\mathbf{z} = (z_i, i \in \mathbb{N}) | z_i \in p, i \in \mathbb{N}, z_i < p-1 \text{ io}(i)\}$. Then $u = \sum\limits_{i=1}^{\infty} z_i p^{-i}$ defines a bijective function $\Phi : \dot{p}^{\infty} \overset{\cong}{\to} [0,1)$.

Note. The range cannot include 1, because it is not allowed to end in $p-1$ repeated and

$$\sum_{i=1}^{\infty} p^{-i}(p-1) = \frac{p-1}{p}\sum_{i=0}^{\infty} p^{-i} = \frac{p-1}{p} = 1.$$

Proof. We know $0 \leqslant u < \frac{p-1}{p}\sum_{i=0}^{\infty} p^{-i} = 1$.

Besides,

$$\begin{aligned}
& u = \sum_{i=1}^{\infty} z_i p^{-i} \\
& \Leftrightarrow 0 \leqslant u - \sum_{i=1}^{n} z_i p^{-i} = \sum_{i=n+1}^{\infty} z_i p^{-i} < \sum_{i=n+1}^{\infty} p^{-i}(p-1) = p^{-(n+1)}\frac{p-1}{1-1/p} = p^{-n}. \\
& \Leftrightarrow z_n p^{-n} \leqslant u - \sum_{i=1}^{n-1} z_i p^{-i} < p^{-n} + z_n p^{-n} = (z_n+1)p^{-n}. \\
& \Leftrightarrow z_n \leqslant p^n\left(u - \sum_{i=1}^{n-1} z_i p^{-i}\right) < z_n + 1.
\end{aligned}$$

Recall that $[x] = m$ iff $m \leqslant x < m+1$ uniquely determines m as the greatest integer less than or equal to x. Therefore,

$$z_n = \left[p^n\left(u - \sum_{i=1}^{n-1} z_i p^{-i}\right)\right], n \geqslant 2,$$

and $z_1 = [pu]$. □

Lemma 4.2. $\sum\limits_{i=0}^{n} a_i p^i = 0$, where $|a_i| < p, \forall i \Leftrightarrow a_i = 0, \forall i$.

Proof. $(\Rightarrow)$ Assume $\sum\limits_{i=0}^{n} a_i p^i = 0$.

(1) When $n = 1 : |a_1|p = |a_0| < p \Rightarrow |a_1| < 1 \Rightarrow a_1 = 0 = a_0$.

(2) Suppose it holds for all n, then

$$\sum_{i=1}^{n+1} a_i p^i = \sum_{i=1}^{n} a_i p^i + a_{n+1}p^{n+1} = a_{n+1}p^{n+1} = 0 \text{ and } a_0 = \cdots = a_n = 0.$$

Therefore,

$$|a_{n+1}p^{n+1}| = |a_n p^n| < p^{n+1} \Rightarrow |a_{n+1}| < 1 \Rightarrow a_{n+1} = 0.$$

Wherefore, by induction, it holds for all i.

($\Leftarrow$) Suppose $a_i = 0, \forall i$, then $\sum_{i=1}^{n} a_i p^i = 0$, where $|a_i| < p, \forall i$. □

Lemma 4.3. For $u = \sum_{i=1}^{\infty} z_i p^{-i}, \mathbf{z} \in \dot{p}^{\infty}$, we have

$$z_1 = b_1, \cdots, z_n = b_n \Leftrightarrow u \in \left[\sum_{i=1}^{n} b_i p^{-i}, \sum_{i=1}^{n} b_i p^{-i} + p^{-n}\right).$$

Proof. ($\Leftarrow$) We have

$$\sum_{i=1}^{n} b_i p^{-i} \leqslant u < \sum_{i=1}^{n} b_i p^{-i} + p^{-n} \Rightarrow \sum_{i=1}^{n} b_i p^{-i} \leqslant \sum_{i=1}^{n} z_i p^{-i} + \sum_{i=n+1}^{\infty} z_i p^{-i} < \sum_{i=1}^{n} b_i p^{-i} + p^{-n}$$

$$\Rightarrow 0 \leqslant \sum_{i=1}^{n} (z_i - b_i)p^{-i} + \sum_{i=n+1}^{\infty} z_i p^{-i} < p^{-n}.$$

Besides,

$$0 \leqslant \sum_{i=n+1}^{\infty} z_i p^{-i} < (p-1) \sum_{i=n+1}^{\infty} p^{-i} = p^{-n},$$

then

$$-p^{-n} < -\sum_{i=n+1}^{\infty} z_i p^{-i} \leqslant 0.$$

Therefore,

$$-p^{-n} < \sum_{i=1}^{n} (z_i - b_i)p^{-i} < p^{-n} \Rightarrow \left|\sum_{i=1}^{n} (z_i - b_i)p^{-i}\right| < p^{-n} \Rightarrow \left|\sum_{i=1}^{n} (z_i - b_i)p^{n-i}\right| < 1,$$

where $|z_i - b_i| < p$. Since $\sum_{i=1}^{n} (z_i - b_i)p^{n-i} \in \mathbb{Z}$, then $\sum_{i=1}^{n} (z_i - b_i)p^{n-i} = 0$. By lemma, $z_i = b_i$.

($\Rightarrow$) Suppose $z_i = b_i, \forall i$, then

$$0 \leqslant \sum_{i=1}^{n} (z_i - b_i)p^{-i} + \sum_{i=n+1}^{\infty} z_i p^{-i} < 0 + (p-1) \sum_{i=n+1}^{\infty} p^{-i} = p^{-n},$$

i.e.,

$$\sum_{i=1}^{n} b_i p^{-i} \leqslant \sum_{i=1}^{\infty} z_i p^{-i} < \sum_{i=1}^{n} b_i p^{-i} + p^{-n}.$$

□

Theorem 4.1 (Fundamental Theorem of Applied Probability)**.** For $U = \sum_{i=1}^{\infty} Z_i p^{-i}, p \geqslant 2$, we have

$$U \sim \text{unif}[0,1] \Leftrightarrow Z_i \overset{\text{i.i.d.}}{\sim} \text{unif}(p).$$

4.4 Bernoulli Distribution

Definition 4.8 (Bernoulli Distribution)**.** $Z \sim \mathrm{Bern}(p), p \in [0,1]$ iff

$$Z \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}.$$

Example 4.9. $Z^{-1} \sim \begin{pmatrix} \infty & 1 \\ q & p \end{pmatrix}$.

Example 4.10. $Z^s = Z, \forall s > 0$.

Property 4.4. $\mathbb{E}[Z] = p, \mathbb{E}[Z^2] = \mathbb{Z} = p$.

Property 4.5. $\mathrm{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = p - p^2 = pq, \sigma(Z) = \sqrt{pq}$.

4.5 Binomial Distribution

Definition 4.9 (Binomial Distribution)**.** $X \sim \mathrm{Bin}(n,p), n \in \mathbb{N}, p \in [0,1]$ iff

$$X \overset{d}{=} Z_1 + \cdots + Z_n,$$

where $Z_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(p)$.

Property 4.6. $\mathbb{E}[X] = \sum\limits_{i=1}^{n} \mathbb{E}[Z_i] = np$.

Property 4.7. $\mathrm{Var}[X] = \sum\limits_{i=1}^{n} \mathrm{Var}[Z_i] = npq, \sigma(X) = \sqrt{n}\sqrt{pq}$.

Property 4.8. Let $X \sim \mathrm{Bin}(m,p), Y \sim \mathrm{Bin}(n,p), X \perp Y$, then $X + Y \sim \mathrm{Bin}(m+n,p)$.

Proof. Take $Z_1, \cdots, Z_{m+n} \overset{\text{i.i.d.}}{\sim} \mathrm{Bern}(p)$. Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \overset{d}{=} \begin{pmatrix} Z_1 + \cdots + Z_m \\ Z_{m+1} + \cdots Z_{m+n} \end{pmatrix},$$

then we have $X + Y \overset{d}{=} \sum\limits_{i=1}^{m+n} Z_i$. □