Probability

Derek Li

Contents

1 Review **2**
1.1 Sequence . . . 2
1.2 Set . . . 4
1.3 Number System and Euclidean Space . . . 5
1.4 Function . . . 5
1.4.1 Indicator Function and Indicator Map . . . 5
1.5 Linear Algebra . . . 7

2 Probability Space **8**

3 Expectation **9**
3.1 Expectation . . . 9
3.2 Variance, Covariance and Correlation . . . 9
3.3 Equality in Distribution . . . 11
3.4 Conditional Expectation . . . 11

4 Probability Distribution **12**
4.1 General Finite Discrete Distribution . . . 12
4.2 Lebesgue-Stieltjes Integral . . . 13
4.3 Uniform Distribution . . . 13
4.3.1 Fundamental Theorem of Applied Probability . . . 15
4.4 Bernoulli Distribution . . . 17
4.5 Binomial Distribution . . . 17
4.6 Negative-Binomial and Geometric Distribution . . . 19
4.7 Poisson Distribution . . . 21
4.8 Exponential Distribution . . . 22
4.9 Gamma Distribution . . . 23
4.10 Beta Distribution . . . 24
4.11 Normal Distribution . . . 25
4.12 χ^2 Distribution . . . 26

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

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

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

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

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

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

□

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. □

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

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 $\omega \in \Omega, A_n \subset \Omega, n \in \mathbb{N}$.

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

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

$$\omega \in \bigcap_{n \geqslant 1} A_n \text{ iff } \forall n, \omega \in A_n.$$

Hence, we have

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

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

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

Property 1.1. For any monotone sequence of sets, there will always exist a limit set, i.e.,

$$A_n \uparrow A = \bigcup_{n \geqslant 1} A_n, \text{ or } A_n \downarrow A = \bigcap_{n \geqslant 1} A_n.$$

Proof. Suppose A_n is increasing, then $\bigcap_{k \geqslant n} A_k = A_n$ and thus

$$\underline{\lim} A_n = \bigcup_{n \geqslant 1} \bigcap_{k \geqslant n} A_k = \bigcup_{n \geqslant 1} A_n.$$

Besides, $\bigcup_{k \geqslant n} A_k = \bigcup_{n} A_n$ and thus

$$\overline{\lim} A_n = \bigcap_{n \geqslant 1} \bigcup_{k \geqslant n} A_k = \bigcap_{n \geqslant 1} \bigcup_{n} A_n = \bigcup_{n \geqslant 1} A_n.$$

Wherefore, $\underline{\lim} A_n = \overline{\lim} A_n = \bigcup_{n \geqslant 1} A_n = A \Rightarrow A_n \uparrow A$.

Similarly, we can show that if $A_n \downarrow \bigcap_{n \geqslant 1} A_n$. □

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:

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

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

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

Definition 1.12. $f = g$ iff $f(x) = g(x), \forall x$.

1.4.1 Indicator Function and Indicator Map

Definition 1.13 (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

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

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 1.14 (Convergence of Set)**.** $A_n \to A$ iff $I(A_n) \to I(A)$.

By the theorem, we have

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

Note that

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

Also,

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

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

Proof. Since $A_n \to A, B_n \to B, C_n \to C$, then $I(A_n) \to I(A), I(B_n) \to I(B), I(C_n) \to I(C)$. Besides,

$$I_{A_n(B_n-C_n)} = I_{A_n(B_n-B_nC_n)} = I_{A_nB_n} - I_{A_nB_nC_n} = I_{A_n}I_{B_n} - I_{A_n}I_{B_n}I_{C_n} \to I_AI_B - I_AI_BI_C = I_{A(B-C)}.$$

Therefore, $A_n(B_n - C_n) \to A(B - C)$. □

Example 1.2. If $A_n \to A, B_n \to B$, then $A_n \Delta B_n \to A \Delta B$.

Proof. Since $A_n \to A, B_n \to B$, then $I(A_n) \to I(A), I(B_n) \to I(B)$. Besides,

$$I_{A_n\Delta B_n} = I_{A_n \cup B_n - A_nB_n} = I_{A_n+B_n-2A_nB_n} = I_{A_n} + I_{B_n} - 2I_{A_n}I_{B_n} \to I_A + I_B - 2I_AI_B = I_{A\Delta B}.$$

Therefore, $A_n \Delta B_n \to A \Delta B$. □

1.5 Linear Algebra

Definition 1.15 (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 1.16 (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.

(1) $\mathcal{F} \neq \varnothing$;

(2) if $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$, i.e., $\mathcal{F}$ is closed under complements;

(3) if $A_i \in \mathcal{F}, i \in \mathbb{N}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, i.e, $\mathcal{F}$ is closed under countable unions.

Note. From the second and third condition above we know that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ since $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^C\right)^C$. 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.

(1) σ-additivity: $P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ provided $A_i A_j = \varnothing, \forall i \neq j$;

(2) non-negativity: $P(A) \geqslant 0, \forall A \in \mathcal{F}$;

(3) 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.

Property 2.1. If $A_n \uparrow A$ or $A_n \downarrow A$, then $P(A_n) \to P(A)$.

Proof. If $A_n \uparrow A$, we have

$$A = \bigcup_{n \geqslant 1} A_n = \bigcup_{n \geqslant 1} (A_n - A_{n-1}),$$

where $A_0 = \varnothing$. Let $B_n = A_n - A_{n-1}$ and we have B_n's are disjoint for all n. Hence,

$$\begin{aligned} P(A) = P\left(\bigcup_{n \geqslant 1} (A_n - A_{n-1})\right) &= \sum_{n=1}^{\infty} P(A_n - A_{n-1}) = \lim_{j \to \infty} \sum_{n=1}^{j} P(A_n - A_{n-1}) \\ &= \lim_{j \to \infty} P\left(\bigcup_{n=1}^{j} (A_n - A_{n-1})\right) = \lim_{j \to \infty} P(A_j) = \lim_{n \to \infty} P(A_n), \end{aligned}$$

i.e., $P(A_n) \to P(A)$.

If $A_n \downarrow A$, then $A_n^C \uparrow A^C$, and $P(A_n^C) \to P(A^C) \Leftrightarrow P(A_n) \to P(A)$. □

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

Proof. Let $B_n = \bigcap_{k \geqslant n} A_k$ and $C_n = \bigcup_{k \geqslant n} A_k$. Since $A_n \to A$, we have $B_n \uparrow A, C_n \downarrow A$ and thus $P(B_n) \to P(A), P(C_n) \to P(A)$. Besides, since $B_n \subseteq A, A_n \subseteq C_n$, then $P(B_n) \leqslant P(A), P(A_n) \leqslant P(C_n)$ and thus

$$P(A) - P(A_n) \leqslant P(C_n) - P(B_n) \to P(A) - P(A) = 0.$$

□

3 Expectation

3.1 Expectation

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

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

Note that $\mathbb{E} : \mathcal{R} \to \mathbb{R}^* \cup \{*\}$ defined on $\mathcal{R} = \{X : \text{Real valued random variable}\}$, and $\mathcal{R}$ itself is a vector space.

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

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

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

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

as $n \to \infty$.

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

3.2 Variance, Covariance and Correlation

Definition 3.2 (Orthogonal Projection)**.** We say

$$\text{o.p.}(X|Y) = \widehat{Y}$$

satisfies two properties: $\widehat{Y} = tY$ for some $t \in \mathbb{R}$ and $X - \widehat{Y} \perp Y$.

Corollary 3.1. Since $\text{o.p.}(X|Y) = \widehat{Y}$, then

$$t = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}.$$

Proof. Since $\text{o.p.}(X|Y) = \widehat{Y}$, then $X - \widehat{Y} \perp Y$ and thus

$$\mathbb{E}\left[\left(X - \widehat{Y}\right) Y\right] = 0 \Rightarrow t = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}.$$

□

Corollary 3.2. Since $\text{o.p.}(X|Y) = \widehat{Y}$, then

$$\cos\measuredangle(X, Y) = \frac{\mathbb{E}[XY]}{\|X\|\|Y\|} \overset{\text{LLN}}{=} \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \cos\measuredangle(\mathbf{x}, \mathbf{y}).$$

Proof. We have

$$\cos \measuredangle(X, Y) = \frac{\|\widehat{Y}\|}{\|X\|} = \frac{t\|Y\|}{\|X\|} = \frac{\mathbb{E}[XY]\|Y\|}{\|Y\|^2\|X\|} = \frac{\mathbb{E}[XY]}{\|X\|\|Y\|} \overset{\text{LLN}}{=} \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \cos \measuredangle(\mathbf{x}, \mathbf{y}).$$

□

Definition 3.3 (Covariance)**.** Define

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

Property 3.2. Covariance satisfies:

(1) bilinear: $\text{Cov}\left[\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \text{Cov}(X_i, Y_j)$;

(2) symmetric: $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;

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

Definition 3.4 (Correlation)**.** Correlation is the cosine of the angle between the centered variables, i.e.,

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

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

Example 3.4. We have

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{(\mathbf{x} - \overline{x}\mathbf{1})^T(\mathbf{y} - \overline{y}\mathbf{1})}{\|\mathbf{x} - \overline{x}\mathbf{1}\|\|\mathbf{y} - \overline{y}\mathbf{1}\|} = \frac{\dot{\mathbf{x}}^T\dot{\mathbf{y}}}{\|\dot{\mathbf{x}}\|\|\dot{\mathbf{y}}\|} \to \rho(X, Y),$$

as $n \to \infty$.

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

Proof. ($\Rightarrow$) Suppose $\text{Var}[X] = 0$, 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, $\text{Var}[X] = \mathbb{E}[Y] = 0 \times 1 = 0$. □

3.3 Equality in Distribution

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

3.4 Conditional Expectation

Definition 3.6 (Conditional Expectation)**.** X is a $\mathbb{R}$-valued random variable, and

$$\mathbb{E}[X|W] = \text{o.p.}(X|L_2(W)),$$

where $L_2(W) = \{g(W)|g : \mathcal{W} \to \mathbb{R}, E[g(W)^2] < \infty\}, \mathcal{W}$ is the sample space of W.

Example 3.6. Take $Z \sim \mathcal{N}(0, 1), X = \frac{1}{Z}, X^2 = \frac{1}{Z^2}$. Thus $X, X^2 \notin L_2(Z)$.

Theorem 3.3. If $X, Y \in L_2$, then $\mathbb{E}[|XY|] < \infty$.

Proof. We have $|XY| \leqslant \frac{X^2+Y^2}{2}$ and thus

$$\mathbb{E}[|XY|] \leqslant \frac{\mathbb{E}[X^2] + \mathbb{E}[Y^2]}{2} < \infty.$$

□

Corollary 3.4. If $X, Y \in L_2$, then $X + Y \in L_2$, i.e., L_2 is a subvector space of $\mathcal{R}$.

Proof. We have $(X + Y)^2 = X^2 + 2XY + Y^2 \leqslant X^2 + 2|XY| + Y^2$ and thus

$$\mathbb{E}[(X + Y)^2] \leqslant \mathbb{E}[X^2] + 2\mathbb{E}[|XY|] + \mathbb{E}[Y^2] < \infty.$$

□

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., $X = \sum_{j=1}^{N} a_j I_{\{a_j\}}(X)$.

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

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

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 $\{x \in \mathbb{R} | p(x) > 0\} = \bigcup_{n=1}^{\infty} \{x \in \mathbb{R} | p(x) > \frac{1}{n}\}$. 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.,

$$\begin{aligned} \mathbb{E}[h(X)] &= \lim_{n\to\infty} \sum_{j=1}^{n} h\left(\frac{j-1}{\sqrt{n}}\right) P\left(\frac{j-1}{\sqrt{n}} < X \leqslant \frac{j}{\sqrt{n}}\right) \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} h\left(\frac{j-1}{\sqrt{n}}\right) \left[F\left(\frac{j}{\sqrt{n}}\right) - F\left(\frac{j-1}{\sqrt{n}}\right) \right] \\ &:= \int_0^\infty h(x) \mathrm{d}F(x), \end{aligned}$$

which is called the Lebesgue-Stieltjes integral.

4.3 Uniform Distribution

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

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

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

Example 4.4. Let $U \sim \text{unif}\{1, \cdots, n\}$, then $-U \sim \text{unif}\{-n, \cdots, -1\}$ and $n+1-U \sim \text{unif}\{1, \cdots, n\}$. 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 $U \sim \text{unif}\{1, \cdots, n\}$, then $U^k \sim \text{unif}\{1^k, \cdots, n^k\}$, 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,

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

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. $(\Rightarrow)$ $P([nU] = k) = P(k \leqslant nU < k+1) = P(\frac{k}{n} \leqslant U < \frac{k+1}{n}) = \frac{1}{n}$. We also need $\frac{k}{n} \geqslant 0$ and $\frac{k+1}{n} \leqslant 1$ and thus $k = 0, \cdots, n-1$, i.e., $[nU] \sim \text{unif}\{0, \cdots, n-1\}, \forall n$.

$(\Leftarrow)$ Consider $P(U < r), \forall r \in \mathbb{Q}[0,1)$. Suppose $r = \frac{k}{n}, k < n$, we have

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

Besides, $\forall 0 \leqslant u < 1, \exists \{r_n\} \in \mathbb{Q}[0,1)$ s.t. $r_n \downarrow u$ 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

Definition 4.8 (p-Adic Series)**.** 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 $.z_1 z_2 \cdots z_n \cdots = \sum_{i=1}^{\infty} z_i 10^{-i}$.

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 $u = \sum_{i=1}^{\infty} z_i p^{-i}$ defines a bijective function $\Phi : \dot{p}^{\infty} \xrightarrow{\cong} [0, 1)$.

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 $z_1 \leqslant pu < z_1 + 1 \Rightarrow z_1 = [pu]$. □

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

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

(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 4.9 (Bernoulli Distribution)**.** $Z \sim \text{bern}(p), p \in [0,1]$ 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. $\text{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = p - p^2 = pq, \sigma(Z) = \sqrt{pq}$.

4.5 Binomial Distribution

Definition 4.10 (Binomial Distribution)**.** $X \sim \text{bin}(n,p), n \in \mathbb{N}, p \in [0,1]$ iff

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

where $Z_i \overset{\text{i.i.d.}}{\sim} Z \sim \text{bern}(p), i \in \mathbb{N}$.

Note that p is the probability per trial.

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

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

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

Proof. Take $Z_1, \cdots, Z_{m+n} \overset{\text{i.i.d.}}{\sim} \text{bern}(p)$. 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 $X + Y \overset{d}{=} \sum_{i=1}^{m+n} Z_i$. □

Property 4.9. $P(X = k) = \binom{n}{k} p^k q^{n-k}$.

Proof. We have $P(X = k) = P(Z_1 + \cdots + Z_n = k)$ and $P(Z = z) = p^z q^{1-z}, z \in \{0,1\}$. Also,

$$\begin{aligned} P((Z_1, \cdots, Z_n) = (z_1, \cdots, z_n)) &= P(Z_1 = z_1, \cdots, Z_n = z_n) \overset{\text{independent}}{=} P(Z_1 = z_1) \cdots P(Z_n = z_n) \\ &= p^{z_1} q^{1-z_1} \cdots p^{z_n} q^{1-z_n} = p^{\sum z_i} q^{n - \sum z_i}. \end{aligned}$$

Hence,

$$P(X = k) = \sum P((Z_1, \cdots, Z_n) = (z_1, \cdots, z_n)), (z_1, \cdots, z_n) \in C_k^n,$$

where $C_k^n = \left\{ (z_1, \cdots, z_n) | z_i \in \{0,1\}, i = 1, \cdots, n \text{ s.t. } \sum_{i=1}^{n} z_i = k \right\}$.

Thus,

$$P(X = k) = \sum_{(z_1, \cdots, z_n) \in C_k^n} p^{\sum z_i} q^{n - \sum z_i} = \sum_{(z_1, \cdots, z_n) \in C_k^n} p^k q^{n-k} = |C_k^n| p^k q^{n-k} := \binom{n}{k} p^k q^{n-k}.$$

□

Property 4.10. $\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!}$.

Proof. We have

$$\sum_{k=0}^{n} P(X=k) = 1 = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

Let $p = \frac{t}{1+t}, q = \frac{1}{1+t}$, then

$$\begin{aligned} (1+t)^n &= \sum_{k=0}^{n} \binom{n}{k} t^k \\ \Rightarrow n(1+t)^{n-1} &= \sum_{k=0}^{n} k \binom{n}{k} t^{k-1} \\ \Rightarrow n \sum_{k=0}^{n-1} \binom{n-1}{k} t^k &= \sum_{k=0}^{n} k \binom{n}{k} t^{k-1} \end{aligned} .$$

Let $j = k+1, k = j-1$, then

$$\sum_{j=1}^{n} n \binom{n-1}{j-1} t^{j-1} = \sum_{k=1}^{n} n \binom{n-1}{k-1} t^{k-1} = \sum_{k=1}^{n} k \binom{n}{k} t^{k-1}.$$

Wherefore

$$k \binom{n}{k} = n \binom{n-1}{k-1} \Rightarrow \binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1} = \frac{n(n-1)\cdots(n-(k-1))}{k(k-1)\cdots(k-(k-1))} \binom{n-k}{0}.$$

Since

$$\begin{aligned} P(X=0) = \binom{n}{0} q^n &= P(Z_1 + \cdots + Z_n = 0) = P(Z_1 = \cdots = Z_n = 0) \\ &\overset{\perp}{=} P(Z_1 = 0) \cdots P(Z_n = 0) \overset{\text{indentical}}{=} P(Z=0)^n = q^n, \end{aligned}$$

then $\binom{n}{0} = 1$ and thus

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-(k-1))}{k(k-1)\cdots(k-(k-1))} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!}.$$

□

Property 4.11. $\mathbb{E}[X^{(r)}] = \begin{cases} n^{(r)} p^r, & r = 0, \cdots, n \\ 0, & r = n+1, \cdots \end{cases}$.

Proof. Let $g(X) = X^{(r)} = X(X-1)\cdots(X-(r-1))$, we have $P(X=k) = \binom{n}{k} p^k q^{n-k}$. Hence

$$\begin{aligned} \mathbb{E}[X^{(r)}] = \mathbb{E}[g(X)] &= \sum_{k=0}^{n} g(k) p(k) = \sum_{k=0}^{n} k^{(r)} \frac{n^{(k)}}{k!} p^k q^{n-k} \\ &= \sum_{k=r}^{n} \frac{n^{(k)}}{(k-r)!} p^k q^{n-k} = n^{(r)} p^r \sum_{k=r}^{n} \frac{(n-r)!}{(k-r)!(n-k)!} p^{k-r} q^{n-k} \\ &= n^{(r)} p^r \sum_{j=0}^{N} \frac{N!}{j!(N-j)!} p^j q^{N-j} = n^{(r)} p^r, \end{aligned}$$

when $r = 0, \cdots, n$. □

Example 4.11. $X^{(2)} = X(X-1) = X^2 - X$ and $\mathbb{E}[X^{(2)}] = n^{(2)}p^2 = \mathbb{E}[X^2] - \mathbb{E}[X]$. Therefore, $\mathbb{E}[X^2] = n^{(2)}p^2 + np$.

Example 4.12. Let $Z_i \overset{\text{independent}}{\sim} \text{bin}\left(i, \frac{2}{3}\right), X = (Z_1 + 3Z_2 - 2Z_3)^2$.

We have

$$\begin{aligned}\mathbb{E}[X] &= \text{Var}[Z_1 + 3Z_2 - 2Z_3] + (\mathbb{E}[Z_1 + 3Z_2 - 2Z_3])^2 \\ &= 1 \times \frac{2}{3} \times \frac{1}{3} + 9 \times 2 \times \frac{2}{3} \times \frac{1}{3} + 4 \times 3 \times \frac{2}{3} \times \frac{1}{3} + \left(\frac{2}{3} + 3 \times 2 \times \frac{2}{3} - 2 \times 3 \times \frac{2}{3}\right)^2 = \frac{22}{3}.\end{aligned}$$

4.6 Negative-Binomial and Geometric Distribution

Definition 4.11 (Negative-Binomial Distribution)**.** Suppose $Z_i \overset{\text{i.i.d.}}{\sim} Z \sim \text{bern}(p), S_n = \sum_{i=1}^{n} Z_i, n \in \mathbb{N}$. We say $T_k \sim \text{negbin}(k, p)$ iff

$$(T_k = n) = (S_{n-1} = k-1, Z_n = 1) = (S_{n-1} < k \leqslant S_n),$$

where $n = k, k+1, \cdots$.

Definition 4.12 (Geometric Distribution)**.** $W \sim \text{geo}(p)$ iff

$$W \overset{d}{=} T_1,$$

i.e., $\text{geo}(p) \equiv \text{negbin}(1, p)$.

Property 4.12. $P(W = n) = P(S_{n-1} = 0, Z_n = 1) = P(S_{n-1} = 0)P(Z_n = 1) = pq^{n-1}$, where $n = 1, \cdots$.

Property 4.13. If $T \overset{d}{=} T_k, W \overset{d}{=} T_1, T \perp W$, then $T + W \overset{d}{=} T_{k+1}$.

Proof. We have

$$\begin{aligned}P(T+W=n) &= \sum_{i=1}^{n-k} P(T = n-i, W = i) = \sum_{i=1}^{n-k} P(T = n-i)P(W = i) \\ &= \sum_{i=1}^{n-k} P(T_k = n-i)P(T_1 = i) = \sum_{i=1}^{n-k} P(S_{n-i-1} = k-1, Z_{n-i} = 1)P(S_{i-1} = 0, Z_i = 1) \\ &= \sum_{i=1}^{n-k} P(Z_1 + \cdots + Z_{n-i-1} = k-1, Z_{n-i} = 1)P(Z_{n-i+1} + \cdots + Z_{n-1} = 0, Z_n = 1) \\ &= \sum_{i=1}^{n-k} P(Z_1 + \cdots + Z_{n-i-1} = k-1, Z_{n-i} = 1, Z_{n-i+1} = \cdots = Z_{n-1} = 0)P(Z_n = 1).\end{aligned}$$

Let $n - i = j$, then $i = n - j$ and thus

$$\begin{aligned}P(T+W=n) &= \sum_{j=k}^{n-1} P(Z_1 + \cdots + Z_{j-1} = k-1, Z_j = 1, Z_{j+1} = \cdots = Z_{n-1} = 0)P(Z_n = 1) \\ &= P(Z_1 + \cdots + Z_{n-1} = k)P(Z_n = 1) = P(Z_1 + \cdots + Z_{n-1} = k, Z_n = 1) \\ &= P(T_{k+1} = n).\end{aligned}$$

Therefore, $T + W \overset{d}{=} T_{k+1}$. □

Note that

$$\begin{aligned} P(Z_1+\cdots+Z_n=k) &= \sum_{j=k}^{n} P(Z_1+\cdots+Z_n=k, k^{\text{th}} \text{ success occurs on } j^{\text{th}} \text{ trail}) \\ &= \sum_{j=k}^{n} P(Z_1+\cdots+Z_n=k, Z_1+\cdots+Z_{j-1}=k-1, Z_j=1) \\ &= \sum_{j=k}^{n} P(Z_1+\cdots+Z_{j-1}=k-1, Z_j=1, Z_{j+1}=\cdots=Z_n=0). \end{aligned}$$

Corollary 4.1. If $T \overset{d}{=} T_1, W \overset{d}{=} T_1, T \perp W$, then $T+W \overset{d}{=} T_2$.

Corollary 4.2. If $W_1, \cdots, W_k \overset{\text{i.i.d.}}{\sim} W \sim \text{geo}(p)$, then

$$W_1+\cdots+W_k \overset{d}{=} T_k.$$

Corollary 4.3. If $T_1 \overset{d}{=} T_{k_1}, T_2 \overset{d}{=} T_{k_2}, T_1 \perp T_2$, then

$$T_1+T_2 \overset{d}{=} T_{k_1+k_2}.$$

Property 4.14. $\mathbb{E}[W] = \sum\limits_{n=1}^{\infty} np(n) = \sum\limits_{n=1}^{\infty} npq^{n-1} = p\frac{\mathrm{d}}{\mathrm{d}q}\sum\limits_{n=0}^{\infty} q^n = p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{1}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{1}{p}$.

Property 4.15. $\text{Var}[W] = \frac{q}{p^2}$.

Proof. We have $\mathbb{E}[W(W-1)] = \mathbb{E}[W^2] - \mathbb{E}[W]$ and

$$\begin{aligned} \mathbb{E}[W(W-1)] &= \sum_{n=1}^{\infty} n(n-1)P(W=n) = pq\sum_{n=1}^{\infty} n(n-1)q^{n-2} \\ &= pq\sum_{n=1}^{\infty} \frac{\mathrm{d}^2}{\mathrm{d}q^2} q^n = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}. \end{aligned}$$

Hence, $\mathbb{E}[W^2] = \frac{2q}{p^2} + \frac{1}{p} = \frac{p+2q}{q} = \frac{1+q}{p^2}$, and thus

$$\text{Var}[W] = \mathbb{E}[W^2] - (\mathbb{E}[W])^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

□

Property 4.16. $\mathbb{E}[T_k] = \frac{k}{p}$.

We can say S_n is the random number of successes in fixed number n of trails and T_k is the random number of trails in fixed number k of successes.

Since $\mathbb{E}[S_n] = np$, then $p = \frac{\mathbb{E}[S_n]}{n}$, i.e., the average number of successes per trail. Since $\mathbb{E}[T_k] = \frac{k}{p}$, then $\frac{1}{p} = \frac{\mathbb{E}[T_k]}{k}$, i.e., the average number of trials per success.

Property 4.17. If $T_k > n \Leftrightarrow S_n < k, k, n \in \mathbb{N}$, then $T_k \sim \text{negbin}(k,p) \Leftrightarrow S_n \sim \text{bin}(n,p)$.

Proof. ($\Leftarrow$) Suppose $S_n \sim \text{bin}(n,p)$. We have

$$\begin{aligned} P(T_k = n) &= P(T_k \leqslant n) - P(T_k < n) = P(T_k \leqslant n) - P(T_k \leqslant n-1) \\ &= P(T_k > n-1) - P(T_k > n) = P(S_{n-1} < k) - P(S_n < k). \end{aligned}$$

Since $(S_n < k) \subset (S_{n-1} < k)$, then

$$P(T_k = n) = P((S_{n-1} < k)(S_n < k)^C) = P(S_{n-1} < k \leqslant S_n) = P(S_{n-1} = k-1, Z_n = 1).$$

Hence, $T_k \sim \text{negbin}(k,p)$.

($\Rightarrow$) Suppose $T_k \sim \text{negbin}(k,p)$ and thus $(T_k = n) = (X_{n-1} = k-1, Z_n = 1)$, where $X_n \sim \text{bin}(n,p)$. We have

$$\begin{aligned} P(S_n = k) &= P(S_n \leqslant k) - P(S_n < k) = P(S_n < k+1) - P(S_n < k) \\ &= P(T_{k+1} > n) - P(T_k > n) = P(T_k \leqslant n) - P(T_{k+1} \leqslant n) \\ &= P(T_k \leqslant n-1) - P(T_{k+1} \leqslant n-1) + P(T_k = n) - P(T_{k+1} = n) \\ &= P(S_{n-1} = k) + P(T_k = n) - P(T_{k+1} = n). \end{aligned}$$

We can show $P(S_n = k) = P(X_n = k)$ by induction:

(i) $n = 1$: since $k = \mathbb{N}$, then $k = 1$. We have

$$P(S_1 = 1) = P(T_1 \leqslant 1) - P(T_2 \leqslant 1) = p = P(X_1 = 1).$$

(ii) $n > 1$: suppose $P(S_{n-1} = k) = P(X_{n-1} = k)$, we have

$$\begin{aligned} P(S_n = k) &= P(S_{n-1} = k) + P(T_k = n) - P(T_{k+1} = n) \\ &= P(X_{n-1} = k) + P(T_k = n) - P(T_{k+1} = n) \\ &= P(X_{n-1} = k, Z_n = 1) + P(X_{n-1} = k, Z_n = 0) \\ &\quad + P(X_{n-1} = k-1, Z_n = 1) - P(X_{n-1} = k, Z_n = 1) \\ &= P(X_n = k, Z_n = 0) + P(X_n = k, Z_n = 1) = P(X_n = k). \end{aligned}$$

Therefore, $S_n \overset{d}{=} X_n$, i.e., $S_n \sim \text{bin}(n,p)$. □

4.7 Poisson Distribution

Before we define Poisson distribution, we first let $\lambda = np$ and thus $p = \frac{\lambda}{n}$. Therefore

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{\frac{n(n-1)\cdots(n-k+1)}{n^k}}{\left(1-\frac{\lambda}{n}\right)^k} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^n.$$

We have

$$\frac{\frac{n(n-1)\cdots(n-k+1)}{n^k}}{\left(1-\frac{\lambda}{n}\right)^k} = \frac{1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k} \to 1, \text{ as } n \to \infty.$$

Besides,

$$\begin{aligned} n \ln\left(1-\frac{\lambda}{n}\right) &= \frac{-\lambda\left[\ln\left(1-\frac{\lambda}{n}\right) - \ln 1\right]}{-\frac{\lambda}{n}} := -\lambda \frac{f(1+h) - f(1)}{h} \\ &\to -\lambda f'(1) = -\lambda, \text{ as } n \to \infty. \end{aligned}$$

Hence, $\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$, as $n \to \infty$.

As a consequence,

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \text{ as } n \to \infty.$$

Definition 4.13 (Poisson Distribution)**.** $N \sim \text{Poisson}(\lambda)$ iff

$$P(N=k) = \lim_{n\to\infty} P(X_n = k),$$

where $X_n \sim \text{bin}\left(n, \frac{\lambda}{n}\right)$.

Property 4.18. $\mathbb{E}[N] = \lambda$.

Proof. We have

$$\mathbb{E}[N] = \sum_{k=0}^{\infty} kp(k) = \sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.$$

□

Suppose N_t is the random number of successes in time $t > 0$. We assume $N_t \sim \text{Poisson}(\mathbb{E}[N_t])$ and $\mathbb{E}[N_t] \propto t$, i.e., $\mathbb{E}[N_t] = \lambda t$ for some λ.

Definition 4.14 (Poisson Process)**.** $(N_t, t > 0)$ is a Poisson process iff $N_t \sim \text{Poisson}(\lambda t)$, for $t > 0$ and for any strictly increasing sequences $0 < t_n$ as $n \to \infty, N_{t_1}, N_{t_2} - N_{t_1}, \cdots, N_{t_n} - N_{t_{n-1}}, \cdots$ are mutually statistically independent.

We say T_n is the random amount of time for n successes, we have

$$T_n > t \Leftrightarrow N_t < n.$$

Therefore,

$$F_n(t) = P(T_n \leqslant t) = P(N_t \geqslant n) = 1 - P(N_t < n) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

and

$$f_n(t) = F_n'(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

4.8 Exponential Distribution

Definition 4.15 (Exponential Distribution)**.** $X \sim \exp(1)$ iff

$$f(x) = e^{-x},$$

and $Y \sim \exp(\theta)$ iff

$$Y = \theta X.$$

4.9 Gamma Distribution

Before we define Gamma distribution, we first let $Z_n = \lambda T_n$. Suppose $G_n(z) = P(Z_n \leqslant z)$, then

$$G_n(z) = P(\lambda T_n \leqslant z) = P\left(T_n \leqslant \frac{z}{\lambda}\right) = F_n\left(\frac{z}{\lambda}\right).$$

Hence,

$$g_n(z) = G_n'(z) = \frac{1}{\lambda} f_n\left(\frac{z}{\lambda}\right) = \frac{z^{n-1}e^{-z}}{(n-1)!} = \frac{z^{n-1}e^{-z}}{\Gamma(n)}.$$

Definition 4.16 (Gamma Distribution)**.** $Z \sim G(p), p > 0$ iff

$$g(z) = \frac{z^{p-1}e^{-z}}{\Gamma(p)}, z > 0,$$

where

$$\Gamma(p) = \int_0^\infty z^{p-1}e^{-z}\mathrm{d}z,$$

and $X \sim G(p, \theta)$ iff

$$X = \theta Z.$$

Property 4.19. If $T_n > t \Leftrightarrow N_t < n, t \geqslant 0, n \in \mathbb{N}$ then $T_n \sim \text{negbin}(n, \lambda^{-1}) \Leftrightarrow N_t \sim \text{Poisson}(\lambda t)$.

Property 4.20. $\Gamma(p+1) = p\Gamma(p)$ and thus $\Gamma(n) = (n-1)!$.

Example 4.13. $\Gamma(1) = \Gamma(2) = 1, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Property 4.21. $\mathbb{E}[Z^s] = \frac{\Gamma(p+s)}{\Gamma(p)}$.

Corollary 4.4. $\mathbb{E}[Z] = \text{Var}[Z] = p$.

Proof. We have

$$\mathbb{E}[Z] = \frac{\Gamma(p+1)}{\Gamma(p)} = p.$$

Since

$$\mathbb{E}[Z^2] = \frac{\Gamma(p+2)}{\Gamma(p)} = \frac{\Gamma((p+1)+1)}{\Gamma(p)} = \frac{(p+1)\Gamma(p+1)}{\Gamma(p)} = (p+1)p,$$

then

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

□

Corollary 4.5. $\mathbb{E}[Z^{-1}] = \frac{\Gamma(p-1)}{\Gamma(p)} = \frac{1}{p-1}$.

Corollary 4.6. If $Z \sim G(p), W \sim G(q), Z \perp W$, then

$$\mathbb{E}\left[\frac{Z}{W}\right] = \mathbb{E}[ZW^{-1}] = \mathbb{E}[Z]\mathbb{E}[W^{-1}] = \frac{p}{q-1}.$$

Corollary 4.7. If $X \sim G(p, \theta)$, then

$$\mathbb{E}[X^s] = \theta^s \mathbb{E}[Z^s] = \frac{\theta^s \Gamma(p+s)}{\Gamma(p)}.$$

Property 4.22. Suppose $T = Z + W, U = \frac{Z}{T}$, then $Z \sim G(p), W \sim G(q), Z \perp W$

$$\Leftrightarrow T \sim G(p+q), f_U(u) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1}, 0 < u < 1, \text{ and } T \perp U.$$

Proof. Since $T = Z + W, U = \frac{Z}{T}$, then $Z = TU, W = T - Z = T(1-U)$. Since $Z \perp W$, then

$$g(z,w) = \frac{z^{p-1}e^{-z}w^{q-1}e^{-w}}{\Gamma(p)\Gamma(q)}.$$

Thus

$$h(t,u) = g(z,w)\left|\frac{\partial(z,w)}{\partial(t,u)}\right|_+ = \frac{u^{p-1}(1-u)^{q-1}t^{p+q-1}e^{-t}}{\Gamma(p)\Gamma(q)} = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}u^{p-1}(1-u)^{q-1} \cdot \frac{t^{p+q-1}e^{-t}}{\Gamma(p+q)}.$$

Therefore,

$$g(t) = \frac{t^{p+q-1}e^{-t}}{\Gamma(p+q)} \Rightarrow T \sim G(p+q).$$

Besides,

$$f_U(u) = \int_{-\infty}^{\infty} h(t,u)\mathrm{d}t = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}u^{p-1}(1-u)^{q-1}$$

and thus $T \perp U$. □

4.10 Beta Distribution

Definition 4.17 (Beta Distribution)**.** $U \sim \text{Beta}(p;q)$ iff

$$U = \frac{Z}{Z+W},$$

where $Z \sim G(p), W \sim G(q), Z \perp W$.

Property 4.23. $\mathbb{E}[U] = \frac{p}{p+q}$.

Proof. Let $T = Z + W$, then $Z = UT$. By the property in Gamma distribution, we have and $T \sim G(p+q)$ and $T \perp U$. Therefore

$$\mathbb{E}[Z] = \mathbb{E}[U]\mathbb{E}[T]$$

and thus

$$\mathbb{E}[U] = \frac{\mathbb{E}[Z]}{\mathbb{E}[T]} = \frac{p}{p+q}.$$

□

Property 4.24. We have

$$\mathbb{E}[U^2] = \frac{\mathbb{E}[Z^2]}{\mathbb{E}[T^2]} = \frac{\Gamma(p+2)/\Gamma(p)}{\Gamma(p+q+2)/\Gamma(p+q)}.$$

4.11 Normal Distribution

Definition 4.18 (Standard Normal Distribution)**.** $Z \sim \mathcal{N}(0,1)$ iff

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Definition 4.19 (Normal Distribution)**.** $X \sim \mathcal{N}(\mu, \sigma^2)$ iff

$$X = \mu + \sigma Z.$$

Property 4.25. $Z \sim \mathcal{N}(0,1) \Leftrightarrow -Z \overset{d}{=} Z, \frac{Z^2}{2} \sim G\left(\frac{1}{2}\right)$.

Proof. ($\Rightarrow$) Suppose $Z \sim \mathcal{N}(0,1)$. By definition, we immediately have

$$-Z \overset{d}{=} Z.$$

Let $W = \frac{Z^2}{2}$, we have

$$\begin{aligned} G(w) &= P(W \leqslant w) = P(-\sqrt{2w} \leqslant Z \leqslant \sqrt{2w}) = 2P(0 \leqslant Z \leqslant \sqrt{2w}) \\ &= 2(P(Z \leqslant \sqrt{2w}) - P(Z \leqslant 0)) = 2P(Z \leqslant \sqrt{2w}) - 1 \\ &= 2\Phi(\sqrt{2w}) - 1. \end{aligned}$$

Thus

$$g(w) = G'(w) = 2\phi(\sqrt{2w}) \cdot (2w)^{\frac{1}{2}-1} = \frac{1}{\sqrt{\pi}} w^{\frac{1}{2}-1} e^{-w} = \frac{w^{\frac{1}{2}-1} e^{-w}}{\Gamma\left(\frac{1}{2}\right)}.$$

Therefore, $W = \frac{Z^2}{2} \sim G\left(\frac{1}{2}\right)$.

($\Leftarrow$) Suppose $-Z \overset{d}{=} Z$ and $\frac{Z^2}{2} \sim G\left(\frac{1}{2}\right)$. Let $W = \frac{Z^2}{2}$. We have $|Z| = \sqrt{2W}$, then

$$P(|Z| \leqslant z) = P(\sqrt{2W} \leqslant z) = P\left(W \leqslant \frac{z^2}{2}\right) = G\left(\frac{z^2}{2}\right).$$

Since $-Z \overset{d}{=} Z$, then

$$P(|Z| \leqslant z) = P(Z \leqslant z) - P(Z \leqslant -z) = 2P(Z \leqslant z) - 1 = 2\Phi(z) - 1.$$

Therefore,

$$\Phi(z) = \frac{1}{2} G\left(\frac{z^2}{2}\right) + \frac{1}{2}$$

and thus

$$\phi(z) = \frac{z}{2} g\left(\frac{z^2}{2}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Hence, $Z \sim \mathcal{N}(0,1)$. □

Property 4.26. If $Z \sim \mathcal{N}(0,1)$, then

$$\mathbb{E}[Z^n] = \begin{cases} 0, & \text{if } n = 2k-1 \\ (2k-1)!!, & \text{if } n = 2k \end{cases},$$

for some $k \in \mathbb{N}$.

Proof. If $n = 2k - 1$, then we have

$$-Z^n = (-Z)^n \overset{d}{=} Z^n \Rightarrow -\mathbb{E}[Z^n] = \mathbb{E}[-Z^n] = \mathbb{E}[Z^n]$$

and thus $\mathbb{E}[Z^n] = 0$.

If $n = 2k$, since $Z^2 = 2W, W \sim G\left(\frac{1}{2}\right)$, then

$$\mathbb{E}[Z^n] = 2^k \mathbb{E}[W^k] = 2^k \frac{\Gamma\left(\frac{1}{2} + k\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{2^k \frac{2k-1}{2} \Gamma\left(\frac{2k-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = \cdots = (2k-1)!!.$$

□

Example 4.14. $\mathrm{Var}[Z^2] = \mathbb{E}[Z^4] - (\mathbb{E}[Z^2])^2 = 3 - 1 = 2$.

Example 4.15. $\mathbb{E}[Z^{16}] = 15!!$.

4.12 χ^2 Distribution

Definition 4.20 (χ^2 Distribution)**.** $X \sim \chi^2_{(m)}$ iff

$$X = 2Z = \sum_{i=1}^{m} Z_i^2,$$

where $Z \sim G\left(\frac{m}{2}\right), Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.