Probability Theory

Labix

May 11, 2025

Abstract

Notes for the basics of Probability Theory.

Contents

1 Foundations of Probability Theory

1.1 Definition of Probability

Definition 1.1.1: Probability Space

A probability space is a measure space $(\Omega, \mathcal{F}, P)$ where the measure P lands in $[0, 1]$.

Explicitly, a probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of the following data:

- $\Omega \neq \emptyset$ is a set called the sample space.
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a σ-algebra called events.
- $P : \mathcal{F} \to [0, 1]$ is a set function.

such that the following are true:

- $P(\Omega) = 1$.
- If $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$ are pairwise disjoint, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

Proposition 1.1.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $A, B \in \mathcal{F}$ be events. Then the following are true.

- $P(\Omega \setminus A) = 1 - P(A)$
- $A \subset B \implies P(A) \leq P(B)$

Proof. Let $A \subset B \subset \Omega$ be events in Ω.

- A and $\Omega \setminus A$ are disjoint and $P(\Omega) = P(A) + P(\Omega \setminus A)$ and $P(\Omega \setminus A) = 1 - P(A)$
- We have that A and $B \setminus A$ are disjoint. Thus $P(B) = P(A) + P(B \setminus A)$. Since $P(B \setminus A) \geq 0$, we have $P(A) \leq P(B)$.

$\square$

Definition 1.1.3: Uniform Probability Measure

Let Ω be a sample space. A probability measure P is uniform if fo all $a, b \in \Omega$,

$$P(\{a\}) = P(\{b\})$$

Theorem 1.1.4

Let Ω be a sample space and P a uniform probability measure of Ω. Then for all $A \subset \Omega$,

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

Proof. Suppose that A consists of $|A|$ distinct elements and the event space $|\Omega|$ contains $|\Omega|$ distinct elements. Since every singleton set is pairwise disjoint, we have $P(A) = |A|P(\{a\})$ for any $a \in A$. Similarly, we have $P(\Omega) = |\Omega|P(\{a\})$. Thus we have that $P(A) = \frac{|A|P(\Omega)}{|\Omega|}$ and $P(A) = \frac{|A|}{|\Omega|}$ $\square$

Theorem 1.1.5: Principle of Inclusion Exclusion

Let $A, B \subset \Omega$ be a sample space and P the probability measure.

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

Proof. Note that

$$\begin{aligned} A \cup (B \setminus A) &= A \cup (B \cap A^c) \\ &= (A \cup B) \cap (A \cup A^c) \\ &= A \cup B \end{aligned}$$

Note also that $A \cap (B \setminus A) = \emptyset$. Thus $P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$ $\square$

Theorem 1.1.6: Extended Principle of Inclusion Exclusion

Let $A_k \subset \Omega$ be a sample space and P the probability measure for all $k \leq n \in \mathbb{N}$. Then

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n}(-1)^{k+1} \sum_{1 \leq i_1 \leq \cdots \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

1.2 Multiplication Principle

Theorem 1.2.1: The Multiplication Principle

Suppose that Experiment A has a outcomes and Experiment B has b outcomes. Then the performing both A and B results in ab possible outcomes.

Theorem 1.2.2: Sampling with replacement - Ordered

In the case of sampling k balls with replacement from an urn containing n balls, there are $|\Omega| = n^k$ possible outcomes when the order of the objects matters, where $\Omega = \{(s_1, \dots, s_k) : s_i \in \{1, \dots, n\} \forall i \in \{1, \dots, k\}\}$.

Theorem 1.2.3: Sampling without replacement - Ordered

In the case of sampling k balls without replacement from an urn containing n balls, there are $|\Omega| = \frac{n!}{(n-k)!}$ possible outcomes when the order of the objects matters, where $\Omega = \{(s_1, \dots, s_k) : s_i \in \{1, \dots, n\} \forall i \in \{1, \dots, k\}, i \neq j \implies s_i \neq s_j\}$.

Theorem 1.2.4: Sampling without replacement - Unordered

In the case of sampling k balls without replacement from an urn containing n balls, there are $|\Omega| = \binom{n}{k}$ possible outcomes when the order of the objects does not matter, where $\Omega = \{\omega \subset \{1, \dots, n\} : |\omega| = k\}$.

Theorem 1.2.5: Sampling with replacement - Unordered

In the case of sampling k balls with replacement from an urn containing n balls, there are $|\Omega| = \binom{n+k-1}{k}$ possible outcomes when the order of the objects does not matter, where $\Omega = \{\omega \subset \{1, \dots, n\} : \omega$ is a k element multiset with elements from $\{1, \dots, n\}\}$.

1.3 Conditional Probability

Definition 1.3.1: Conditional Probability

Consider a probability space (Ω, P). Let $A, B \subset \Omega$ with $P(B) > 0$. Then the conditional probability of A given B, denoted by $P(A|B)$ is defined as

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

Theorem 1.3.2: Multiplication Rule

Let $n \in \mathbb{N}$. Then for any events $A_1, \dots, A_n$ such that $P(A_2 \cap \dots \cap A_n) > 0$, we have

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Proof. From the right hand side, we have

$$\begin{aligned}
&P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}) \\
&= P(A_1)\frac{P(A_2 \cap A_1)}{P(A_1)} \frac{P(A_3 \cap A_2 \cap A_1)}{P(A_2 \cap A_1)} \dots \frac{P(A_n \cap \dots \cap A_1)}{P(A_1 \cap \dots \cap A_{n-1})} \\
&= P(A_1 \cap \dots \cap A_n)
\end{aligned}$$

□

Theorem 1.3.3: Bayes' Rule

Let (Ω, P) be a probability measure. Let $A, B \subset \Omega$ with $P(A), P(B) > 0$. Then

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

Proof. We have that $P(A \cap B) = P(A|B)P(B)$ and $P(A \cap B) = P(B|A)P(A)$. □

Theorem 1.3.4: Law of Total Probability

Let (Ω, P) be a probability measure. Let $A_1, \dots, A_n$ be a partition of Ω with $P(A_i) > 0$ for all $i = 1, \dots, n$. Then for any $B \subset \Omega$,

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

Proof. Note that since $A_1, \dots, A_n$ is a partition, $B \cap A_1, \dots, B \cap A_n$ is also a parition.

$$\begin{aligned}
\sum_{k=1}^{n} P(A_k)P(B|A_k) &= \sum_{k=1}^{n} P(B \cap A_k) \\
&= P\left(\bigcup_{k=1}^{n} B \cap A_k\right) \\
&= P(B \cap \Omega) \\
&= P(B)
\end{aligned}$$

□

Theorem 1.3.5: General Bayes' Rule

Let (Ω, P) be a probability measure. Let $A_1, \dots, A_n$ be a partition of Ω with $P(A_i) > 0$ for all $i = 1, \dots, n$. Then for any $B \subset \Omega$ with $P(B) > 0$,

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

Proof. Apply Bayes' rule and apply the mulitplication rule. □

1.4 Independence of Events

Definition 1.4.1: Independent Events

Two events A, B are said to be independent if

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

Proposition 1.4.2

If A, B are independent, then A^c, B, A, B^c and A^c, B^c are independent.

Proof. We only proof the first and third item.

- Without loss of generality we prove the first and reader mirrors the second.

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) \\ &= P(B)(1 - P(A)) \\ &= P(B)P(A^c) \end{aligned}$$

- Note that $P(A \cap B) = P(A)P(B)$

$$\begin{aligned} P(A^c \cap B^c) &= 1 - P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

□

2 Probability Distributions

2.1 Random Variables and its Distribution

Definition 2.1.1: Random Variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(E, \mathcal{E})$ be a measurable space. An $(E, \mathcal{E})$ valued random variable is an $\mathcal{F}$-measurable function $X : \Omega \to E$.

Definition 2.1.2: Independent Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(E, \mathcal{E})$ be a measurable space. Let $X, Y : \Omega \to E$ be random variables. We say that X and Y are independent if for any $A, B \in \mathcal{E}$, we have that $X^{-1}(A)$ and $Y^{-1}(B)$ are independent events in $\mathcal{F}$.

Definition 2.1.3: Discrete and Continuous Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable.

- We say that X is discrete if $\text{im}(X)$ is a countable subset of $\mathbb{R}$.
- We say that X is continuous otherwise.

Recall that X is an $\mathcal{F}$-measurable function if $X^{-1}(B) \in \mathcal{F}$ for $B \in \mathcal{E}$.

Definition 2.1.4: Probability Distribution

Let $(\Omega, E, \mathbb{P})$ be a probability space. Let $(E, \mathcal{E})$ be a measurable space. Let $X : \Omega \to E$ be a measurable function. Define the probability distribution of X to be the pushforward measure $P \circ X^{-1} = P_X : \mathcal{E} \to [0, 1]$ defined by

$$P_X(A) = P(X^{-1}(A))$$

for $A \in \mathcal{E}$.

Definition 2.1.5: Probability Density Function

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Define the probability density function of X to be the Radon–Nikodym derivative

$$f_X = \frac{dX_* P}{d\mu}$$

where μ is the Lebesgue measure.

Recall that this means that f_X satisfies the property that

$$P_X(A) = \int_A f_X d\mu$$

for any measurable set $A \subseteq \mathbb{R}$. In particular, if $A = \{a\} \subseteq \mathcal{F}$, then we have

$$P_X(a) = f_X(a)$$

The probability distribution function has its input as every measurable subset of $\mathbb{R}$, while the probability density function takes input as individual points of $\mathbb{R}$. They are really the same thing because having its probability be determined on singletons is sufficient to determine the probability of every measurable subset.

Proposition 2.1.6

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a discrete random variable. Let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then the probability density function of $Y = g \circ X$ is given by

$$f_Y(y) = \sum_{x \in g^{-1}(y)} f_X(x)$$

Proposition 2.1.7

Suppose that X is a continuous random variable with density f_X and $g : \mathbb{R} \to \mathbb{R}$ is strictly monotone and differentiable with inverse function denoted g^{-1}, then $Y = g(X)$ has density

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}(g^{-1}(y))\right|$$

for all $y \in \mathbb{R}$

Example 2.1.8: Bernoulli Distribution

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. We say that X has a Bernoulli distribution if the probability density function of X is given by

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

for some $p \in [0, 1]$.

Example 2.1.9: Binomial Distribution

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. We say that X has a binomial distribution if the probability density function of X is given by

$$f_X(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for some $p \in [0, 1]$.

Definition 2.1.10: Poisson Distribution

A discrete random variable X is said to have Poisson Distribution with parameter $\lambda > 0$ if $\operatorname{im}(X) = \mathbb{N}_0$ and

$$p_X(x) = \frac{\lambda^x}{x!} e^{-x}$$

Definition 2.1.11: Geometric Distribution

A discrete random variable X is said to have Geometric Distribution with parameter $p \in (0, 1)$ if $\operatorname{im}(X) = \mathbb{N}_0$ and

$$p_X(x) = p(1 - p)^{x-1}$$

Let $I \subseteq \mathbb{R}$ be an interval. Recall that $\mathcal{B}(I)$ refers to the borel measurable subsets of I. Denote λ the Lebesgue measure on $\mathbb{R}^n$.

Example 2.1.12: Uniform Distribution

Let $[a,b] \subseteq \mathbb{R}$ be an interval. Let X be a random variable on the probability space $([a,b], \mathcal{B}([a,b]), P)$. We say that X has a uniform distribution if its probability density function is given by

$$f_X(A) = \frac{\lambda(A)}{b-a}$$

for $A \subseteq [a,b]$.

In particular, when $A = \{c\} \subseteq [a,b]$ is the one-point set, we have $P_X(c) = \frac{1}{b-a}$ so that the probability of any one point set is uniform.

Example 2.1.13

Let X be a uniform distribution on $[a,b]$. Then the probability density function of X is given by

$$F_X(x) = \frac{1}{b-a}$$

Example 2.1.14: Normal Distribution

Let X be a random variable on the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$. We say that X has a normal distribution if its probability density function is given by

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

for some $\mu \in \mathbb{R}$ and $\sigma > 0$.

Definition 2.1.15: Exponential Distribution

A conitnuous random variable X is said to have Exponential Distribution with parameter $\lambda > 0$ if its density function is given by

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

and its cumulative function given by

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0 \end{cases}$$

Definition 2.1.16: Gamma Distribution

A conitnuous random variable X is said to have Gamma Distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$ if its density function is given by

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

2.2 Cumulative Density Functions

Definition 2.2.1: Cummulative Distribution Function

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Define the cumulative distribution function $F_X : \mathbb{R} \to \mathbb{R}$ of X to be

$$F_X(x) = P_X(X \leq x)$$

Proposition 2.2.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the following are true.

- $f_X = \frac{dF_X}{dx}$.
- $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$

Proposition 2.2.3

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the following are true regarding the cumulative distribution function F_X.

- F_X is monotonically increasing: $x \leq y \implies F_X(x) \leq F_X(y)$
- F_x is right continuous: If (x_n) is a sequence such that $x_1 \geq \cdots \geq x_n \geq x_{n+1} \geq \cdots \geq x$ and $(x_n) \to x$, then $F_X(x_n) \to F_X(x)$
- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$

Proposition 2.2.4

Suppose that X is a random variable on a probability space $(\Omega, E, \mathbb{P})$ with cumulative distribution function F_X. If $a < b$, then $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

2.3 Multivariate Random Variables

Let $(\Omega, E, \mathbb{P})$ be a probability space. The definition of random variables and probability distribution is well-adapted to the case when the random variable X lands in $\mathbb{R}^n$. In this case, we may find the relationship between the probability density function of X and the probability density function of its individual components.

Definition 2.3.1: Joint Probability Mass Function

Let X, Y be discrete random variables. The joint probability mass function of X and Y is the function

$$p_{X,Y}(x, y) = P(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}) = P((X, Y) = (x, y))$$

for all $(x, y) \in \mathbb{R}^2$

Theorem 2.3.2

Let $p_{X,Y}$ be the joint probability mass function of two random variables X, Y.

- $p_X(x) = \sum_y p_{X,Y}(x, y)$
- $p_Y(y) = \sum_x p_{X,Y}(x, y)$

Definition 2.3.3: Joint Cumulative Distribution Function

Let X, Y be random variables. The joint cumulative distribution function of X and Y is the function

$$F_{X,Y}(x, y) = P(\{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y\}) = P(X \leq x, Y \leq y)$$

for all $(x, y) \in \mathbb{R}^2$

Theorem 2.3.4

Let $F_{X,Y}$ be the joint cumulative distribution function of two random variables X, Y.
- $\lim_{x,y\to-\infty} F_{X,Y}(x, y) = 0$
- $\lim_{x,y\to\infty} F_{X,Y}(x, y) = 1$
- $x \leq x'$ and $y \leq y'$ implies $F_{X,Y}(x, y) \leq F_{X,Y}(x', y')$
- $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y)$
- $F_Y(x) = \lim_{x\to\infty} F_{X,Y}(x, y)$

Definition 2.3.5: Jointly Continuous

Let X, Y be random variables. X and Y are jointly continuous if

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, dv\, du$$

for a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}^2$ satisfying
- $f_{X,Y}(u, v) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, dv\, du = 1$

We call $f_{X,Y}$ the joint density function of (X, Y).

Theorem 2.3.6

Let $F_{X,Y}$ be the joint cumulative distribution function of two random variables X, Y.
- $f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) & \text{if the derivative exists at } (x, y) \\ 0 & \text{otherwise} \end{cases}$
- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$
- $f_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$

Proposition 2.3.7: L

t $(\Omega, E, \mathbb{P})$ be a probability space. Let $(E, \mathcal{E})$ be a measurable space. Let $X, Y : \Omega \to \mathbb{R}$ be a random variables. Then the following are equivalent.
- X and Y are independent.
- $f_{(X,Y)} = f_X f_Y$.
- $F_{(X,Y)} = F_X F_Y$

2.4 Algebra of Random Variables

Proposition 2.4.1

Let $(\Omega, E, \mathbb{P})$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be a random variables. Then we have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(t, z - t)\, dt$$

Proposition 2.4.2

Let $X \approx$Poi(λ) and $Y \approx$Poi(μ) be independent. $X + Y \approx$Poi$(\lambda + \mu)$.

Proof.

$$\begin{aligned} p_{X+Y}(m) &= \sum_{k\in\mathbb{Z}} \frac{\lambda^k}{k!}e^{-k}\frac{\mu^{m-k}}{(m-k)!}e^{k-m} \\ &= \frac{1}{m!}e^{-m}\sum_{k=0}^{m} m!\frac{\lambda^k}{k!}\frac{\mu^{m-k}}{(m-k)!} \\ &= \frac{1}{m!}e^{-m}\sum_{k=0}^{m}\binom{m}{k}\lambda^k\mu^{m-k} \\ &= \frac{(\lambda+\mu)^m}{m!}e^{-m} \end{aligned}$$

□

Proposition 2.4.3

Let $X_1,\dots,X_n \approx$Bern(p) be independent. $X_1+\dots+X_n \approx$Bin(n,p).

Proof. We prove by induction. When $n=2$,

$$\begin{aligned} p_{X_1+X_2}(0) &= p_{X_1}(0)p_{X_2}(0) \\ &= 1-2p+p^2 \\ p_{X_1+X_2}(1) &= p_{X_1}(0)p_{X_2}(1)+p_{X_1}(1)p_{X_2}(0) \\ &= (1-p)(p)+p(1-p) \\ &= 2p(1-p) \\ p_{X_1+X_2}(2) &= p_{X_1}(0)p_{X_2}(2)+p_{X_1}(1)p_{X_2}(1)+p_{X_1}(2)p_{X_2}(0) \\ &= p^2 \\ p_{\text{Bin}(2,p)}(x) &= \binom{2}{x}p^x(1-p)^{n-x} \end{aligned}$$

For $x\in\{0,1,2\}$, the two probability density functions match thus for the case $n=2$, it is true. Now suppose that $X_1+\dots+X_{n-1}\approx$Bin$(n-1,p)$. Let $Y=$Bin$(n-1,p)+X_n$. For $m\in\{0,\dots,n\}$,

$$\begin{aligned} p_Y(m) &= \sum_{k\in\mathbb{Z}} p_{\text{Bin}(n-1,p)}(k)p_{X_n}(m-k) \\ &= \sum_{k=0}^{m} p_{\text{Bin}(n-1,p)}(k)p_{X_n}(m-k) \\ &= \sum_{k=0}^{m}\binom{n-1}{k}p^k(1-p)^{n-1-k}p_{X_n}(m-k) \\ &= \sum_{k=m-1}^{m}\binom{n-1}{k}p^k(1-p)^{n-1-k}p_{X_n}(m-k) \\ &= \binom{n-1}{m-1}p^{m-1}(1-p)^{n-m}p_{X_n}(1)+\binom{n-1}{m}p^m(1-p)^{n-1-m}p_{X_n}(0) \\ &= \binom{n-1}{m-1}p^m(1-p)^{n-m}+\binom{n-1}{m}p^m(1-p)^{n-m} \\ &= \binom{n}{m}p^m(1-p)^{n-m} \end{aligned}$$

Thus for the case $X_1+\dots+X_n$ it is true. □

Proposition 2.4.4

Let $X \approx \text{Bin}(m,p)$ and $Y \approx \text{Bin}(n,p)$ be independent. $X+Y \approx \text{Bin}(m+n,p)$.

Proof.

$$\begin{aligned} p_{X+Y}(t) &= \sum_{k\in\mathbb{Z}} p_X(k)p_Y(t-k) \\ &= \sum_{k=0}^{t} \binom{m}{k} p^k(1-p)^{m-k} \binom{n}{t-k} p^{t-k}(1-p)^{n-t+k} \\ &= \sum_{k=0}^{t} \binom{m}{k}\binom{n}{t-k} p^t(1-p)^{m+n-t} \\ &= p^t(1-p)^{m+n-t} \sum_{k=0}^{t} \frac{m!}{k!(m-k)!}\frac{n!}{(t-k)!(n-t+k)!} \end{aligned}$$

□

Proposition 2.4.5

Let $\lambda > 0$. Let $n \in \mathbb{N}$. Let $T_1, \dots, T_n$ be independent random variables with exponential distribution parameter λ. Then

$$Z = \sum_{k=1}^{n} T_k \approx \text{Gamma}(n,\lambda)$$

Proof. We prove by induction. When $n = 2$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{T_1}(x) f_{T_2}(z-x)\, dx \\ &= \int_0^z \lambda e^{-\lambda x}\lambda e^{-\lambda(z-x)}\, dx \\ &= \lambda^2 e^{-\lambda z}\int_0^z dx \\ &= \lambda^2 z e^{-\lambda z} \end{aligned}$$

Thus the case $n = 2$ is true. Suppose that it is true for $n = k-1$. Let $X \approx \text{Gamma}(n-1,\lambda)$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_{T_n}(z-x)\, dx \\ &= \int_0^z \frac{\lambda^{n-1}}{\Gamma(n-1)} x^{n-2} e^{-\lambda x}\lambda e^{-\lambda(z-x)}\, dx \\ &= \frac{\lambda^n}{\Gamma(n-1)} e^{-\lambda z}\int_0^z x^{n-2}\, dx \\ &= \frac{\lambda^n}{\Gamma(n-1)} e^{-\lambda z}\frac{1}{n-1} z^{n-1} \\ &= \frac{\lambda^n}{\Gamma(n)} z^{n-1} e^{-\lambda z} \end{aligned}$$

Thus we are done □

Proposition 2.4.6

Let $m, n \in \mathbb{N}$ and $\lambda > 0$. Let $X \approx$Gamma(m, λ) and $Y \approx$Gamma(n, λ) be independent. $X + Y \approx$Gamma$(m + n, \lambda)$.

Proof.

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \\
&= \int_0^z \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x} \frac{\lambda^n}{\Gamma(n)} (z-x)^{n-1} e^{-\lambda(z-x)}\, dx \\
&= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} e^{-\lambda z} \int_0^z x^{m-1}(z-x)^{n-1}\, dx \\
&= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} e^{-\lambda z} \int_0^z x^{m-1} \sum_{k=0}^{n-1} \binom{n-1}{k} z^{n-1-k}(-x)^k\, dx \\
&= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} e^{-\lambda z} \sum_{k=0}^{n-1} \binom{n-1}{k} z^{n-1-k}(-1)^k \int_0^z x^{m-1+k}\, dx \\
&= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} z^{m+n-1} e^{-\lambda z} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{1}{m+k}
\end{aligned}$$

□

Theorem 2.4.7

Suppose that $T_1, T_2, \dots$ are independent random variables with exponential distribution parameter λ. Define for $t \geq 0$,

$$N_t = \begin{cases} 0 & \text{if } T_1 > t \\ 1 & \text{if } T_1 \leq t < T_1 + T_2 \\ 2 & \text{if } T_1 + T_2 \leq t < T_1 + T_2 + T_3 \\ \dots & \end{cases}$$

Then, for any $t \geq 0$, we have that $N_t \approx$Poi(λt).

Definition 2.4.8: Poisson Process

The family of random variables $\{N_t : t \geq 0\}$ is said to be Poisson process of intensity λ if

- $N_0 = 0$
- for any $t_0, \dots, t_n$ with $0 = t_0 < t_1 < t_2 < \cdots < t_n$, the random variables N_{t_1}, $N_{t_2} - N_{t_1}$, $N_{t_3} - N_{t_2}$, …, $N_{t_n} - N_{t_{n-1}}$ are independent, and $N_{t_i} - N_{t_{i-1}} \approx$Poi$(\lambda(t_i - t_{i-1}))$

3 Expectation and Variance

3.1 Expectations

Definition 3.1.1: Expectations

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Define the expectation of X to be
$$E[X] = \int_\Omega X dP$$

Lemma 3.1.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then we have
$$E[X] = \int_\mathbb{R} x f_X(x)\, dx$$

Proposition 3.1.3: Law of the Unconscious Staticians

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_1, \dots, X_n : \Omega \to \mathbb{R}$ be random variables. Let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then we have
$$E[g \circ (X_1, \dots, X_n)] = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

Proposition 3.1.4

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Then the following are true.

- If X, Y are random variables and $a, b \in \mathbb{R}$, then
$$E[aX + bY] = aE[X] + bE[Y]$$
- If $P(X \geq Y) = 1$, then
$$E[X] \geq E[Y]$$

Proposition 3.1.5

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Then X, Y are independent if and only if
$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$
for any two functions $g, h : \mathbb{R} \to \mathbb{R}$.

3.2 Variance and Covariance

Definition 3.2.1: Variance

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Define the variance of X to be
$$\text{Var}(X) = E[(X - E[X])^2]$$

Lemma 3.2.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the following are true.

- $\mathrm{Var}(X) \geq 0$.
- $\mathrm{Var}(X) = 0$ if and only if $P_X(E[X]) = 1$.
- $\mathrm{Var}(X) = E[X^2] - E[X]^2$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ for any $a, b \in \mathbb{R}$.

Proposition 3.2.3

Suppose that $X_1, \dots, X_n$ are independent variables with finite variance. Then

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k)$$

Definition 3.2.4: Covariance

Let X, Y be two random variables. The covariance of X, Y is defined as

$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Proposition 3.2.5

Suppose that X, Y are random variables.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- If X, Y are independent, $\mathrm{Cov}(X, Y) = 0$
- $\mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$

Proposition 3.2.6: Variance of Sums

For random variables $X_1, \dots, X_n$, we have

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}\left(X_i, X_j\right)$$

Theorem 3.2.7

Given two random variables X and Y, we have

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

Theorem 3.2.8: Correlation Coefficient

The correlation coefficient between two random variables X and Y is given by

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Proposition 3.2.9

Let X and Y be random variables. We have

$$-1 \leq \rho(X, Y) \leq 1$$

Moreover, for any $a, b, c, d \in \mathbb{R}$ with $a, c > 0$, we have

$$\rho(aX + b, cY + d) = \rho(X, Y)$$

Proposition 3.2.10

Let X, Y be random variables.

- $\rho(X, X) = 1$
- $\rho(X, -X) = -1$
- X, Y are uncorrelated if $\rho(X, Y) = 0$

3.3 Moments

Definition 3.3.1: kth Moment

Let X be a random variable. For $k \in \mathbb{N}$ we define the kth moment of X as $E[X^k]$ whenever the expectation exists.

Definition 3.3.2: Moment Generating Function

The moment-generating function of a random variable X is the function M_X defined as

$$M_X(t) = E[e^{tX}]$$

for all $t \in \mathbb{R}$ for which the expectation is well defined.

Theorem 3.3.3

Assume that M_X exists in a neighbourhood of 0, that is, there exists $\epsilon > 0$ such that for all $t \in (-\epsilon, \epsilon)$ we have $M_X(t) < \infty$. Then for all $k \in \mathbb{N}$ the kth moment of X exists, and

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}$$

Proof. We have that $E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x)\,dx$ for any continuous cummulative probability. On the other hand,

$$\begin{aligned}
\frac{d^k}{dt^k} M_X(t)\bigg|_{t=0} &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx\bigg|_{t=0} \\
&= \int_{-\infty}^{\infty} \frac{\partial^k}{\partial t^k} e^{tx} f_X(x)\,dx\bigg|_{t=0} \\
&= \int_{-\infty}^{\infty} x^k e^{tx} f_X(x)\,dx\bigg|_{t=0} \\
&= \int_{-\infty}^{\infty} x^k f_X(x)\,dx
\end{aligned}$$

□

Proposition 3.3.4

Assume that all expectations in the statement are well defined.

- For any $a, b \in \mathbb{R}$, $M_{aX+b}(t) = e^{tb} M_X(at)$
- If X, Y are independent, then $M_{X+Y}(t) = M_X(t) M_Y(t)$

Theorem 3.3.5

Let X, Y be two random variables. Assume that the moment generating functions of X, Y exists and are finite on an interval of the form $(-\epsilon, \epsilon)$. Assume further that $M_X(t) = M_Y(t)$ for all $t \in (-\epsilon, \epsilon)$. Then X, Y have the same distribution.

Theorem 3.3.6

Let X be a non-negative random variable whose expectation is well defined. We then have

$$P(X \geq x) \leq \frac{E(X)}{x}$$

Theorem 3.3.7

Let X be a random variable whose variance is well defined. Then

$$P(|X - E(X)| \geq x) \leq \frac{\text{Var}(X)}{x^2}$$

for all $x > 0$

3.4 Conditional Expectations

Definition 3.4.1: Conditional Expectations on Subalgebras

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Let $\mathcal{H}$ be a σ-subalgebra of $\mathcal{F}$. Define $E[X \mid \mathcal{H}] : \Omega \to \mathbb{R}$ to be a random variable such that the following are true.

- $E[X \mid \mathcal{H}]$ is $\mathcal{H}$-measurable.
- For any $A \in \mathcal{H}$, we have $E[X \cdot 1_A] = E[E[X \mid \mathcal{H}] \cdot 1_A]$

Lemma 3.4.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. Let $\mathcal{H}$ be a σ-subalgebra of $\mathcal{F}$. Then the random variable $E[X \mid \mathcal{H}]$ exists and is unique up to almost surely equality.

Lemma 3.4.3

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Let $\mathcal{H}$ be a σ-subalgebra of $\mathcal{F}$. Then the following are true.

- Stability: If X is $\mathcal{H}$-measurable, then $E[XY \mid \mathcal{H}] = XE[Y \mid \mathcal{H}]$.
- Independence: If $\sigma(X)$ and $\mathcal{H}$ are independent, then $E[X \mid \mathcal{H}] = E[X]$.

Definition 3.4.4: Conditional Expectation on Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Define the conditional expectation of X on Y to be

$$E[X \mid Y] = E[X \mid \sigma(Y)]$$

Definition 3.4.5: Conditional Density

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Define the conditional density of X on the event $\{\omega \in \Omega \mid Y(\omega) = y\}$ by

$$f_{X \mid Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Lemma 3.4.6

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. Then we have

$$E[X \mid Y](\omega) = E[X \mid Y = Y(\omega)] = \int_{-\infty}^{\infty} x f_{X \mid Y}(x, Y(\omega))\, dx$$

4 Convergence of Random Variables

4.1 Different Notions of Convergences

Definition 4.1.1: Convergence in Mean Square

We say that a sequence of random variables $X_1, X_2, \dots$ converges in mean square to a random variable X if

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0$$

Definition 4.1.2: Convergence in Probability

We say that a sequence of random variables $X_1, X_2, \dots$ converges in probability to a random variable X if for every $\epsilon > 0$, we have

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

Definition 4.1.3: Convergence in Distribution

We say that a sequence of random variables $X_1, X_2, \dots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for every x in the set $C = \{x \in \mathbb{R} : F_X \text{ is continuous at } x\}$.

Proposition 4.1.4

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_n : \Omega \to \mathbb{R}$ be a sequence of random variables for $n \in \mathbb{N} \setminus \{0\}$. Let $X : \Omega \to \mathbb{R}$ also be a random variable. Then the following are true.

- If X_n converges in mean square to X, then X_n converges in probability to X.
- If X_n converges in probability to X, then X_n converges in distribution to X.

4.2 Law of Large Numbers

Theorem 4.2.1: Markov Inequality

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. If $E[X] < \infty$, then we have

$$P(|X| \geq a) \leq \frac{E[X]}{a}$$

for any $a > 0$.

Theorem 4.2.2: Chebyshev Inequality

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable. If $E[X^2] < \infty$, then we have

$$P(|X| \geq a) \leq \frac{E[X^2]}{a}$$

for any $a > 0$.

Theorem 4.2.3: Weak law of large numbers

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_n : \Omega \to \mathbb{R}$ for $n \in \mathbb{N} \setminus \{0\}$ be a sequence of independently identically distributed random variables with mean μ. Let $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then we have

$$\lim_{n\to\infty} P(|S_n - \mu| > \varepsilon) = 0$$

for all $\varepsilon > 0$. In other words, $(S_n)_{n\in\mathbb{N}\setminus\{0\}}$ converges in probability to μ.

Theorem 4.2.4: Law of large numbers in mean square

Let $X_1, X_2, \dots$ be a sequence of independent random variable, each with mean μ and variance σ^2. Then

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} \to \mu$$

in mean square.

4.3 Central Limit Theorem

Theorem 4.3.1: Central Limit Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_n : \Omega \to \mathbb{R}$ be a sequence of independent and identically distributed random variables for $n \in \mathbb{N} \setminus \{0\}$, each with mean μ and variance $\sigma^2 \neq 0$. Let $S_n = X_1 + \cdots + X_n$. Then the standardized version of S_n,

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\operatorname{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution as $n \to \infty$ to a Gaussian random variable with mean 0 and variance 1. That is,

$$\lim_{n\to\infty} P(Z_n \leq x) = \lim_{n\to\infty} F_{Z_n}(x) = F_Y(y) = \int_{-\infty}^{x} -\frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$