1: PROBABILITY REVIEW

Marek Rutkowski

School of Mathematics and Statistics
University of Sydney

Semester 2, 2020

Outline

We will review the following notions:

1. Probability Measure
2. Equivalence of Probability Measures
3. Expectation of a Random Variable
4. Variance of a Random Variable
5. Examples of Discrete Distributions
6. Continuous Random Variables
7. Examples of Continuous Distributions
8. Conditional Distributions and Expectations

PART 1

PROBABILITY MEASURE

Sample Space

- We collect the possible states of the world and denote the set by Ω. Each states is called a **sample** or an **elementary event.**
- The sample space Ω is either **countable** or **uncountable**.
  - A toss of a coin: $\Omega = \{\text{Head}, \text{Tail}\} = \{H, T\}$.
  - The number of successes in a sequence of n identical and independent trials: $\Omega = \{0, 1, \ldots, n\}$.
  - The moment of occurrence of the first success in an infinite sequence of identical and independent trials: $\Omega = \{1, 2, \ldots\}$.
  - The lifetime of a light bulb: $\Omega = \{x \in \mathbb{R} \,|\, x \geq 0\}$.
- The choice of a sample space is arbitrary and thus any set can be taken as a sample space. However, practical considerations justify the choice of the most convenient sample space for the problem at hand.
- **Discrete** (either finite or infinite, but countable) sample spaces are easier to handle than general sample spaces.

Probability

Definition (Probability)

A map $\mathbb{P} : \Omega \mapsto [0, 1]$ is a **probability** on a discrete sample space Ω if

- P.1. $\mathbb{P}(\omega_k) \geq 0$ for all $k \in I$,
- P.2. $\sum_{k \in I} \mathbb{P}(\omega_k) = 1$.

Definition (Discrete Random Variable)

A function $X : \Omega \to \mathbb{R}$ on a discrete sample space $\Omega = (\omega_k)_{k \in I}$, where the set I is countable, is called a **discrete random variable**.

- Examples of random variables:
  - Prices of stocks.
  - Exchange rates.
  - Payoffs corresponding to portfolios.

Probability Measure

- Let $\mathcal{F} = 2^\Omega$ stand for the class of all subsets of the sample space Ω. The set 2^Ω is called the **power set** of Ω.
- Note that the **empty set** $\emptyset$ also belongs to any power set.

Definition (Probability Measure)

A map $\mathbb{P} : \mathcal{F} \mapsto [0,1]$ is called a **probability measure** on $(\Omega, \mathcal{F})$ if

- M.1. For any sequence $A_i \subset \mathcal{F}$, $i = 1, 2, \dots$ of events such that $A_i \cap A_j = \emptyset$ for all $i \neq j$ we have

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i \big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

- M.2. $\mathbb{P}(\Omega) = 1$.

- For any **event** $A \in \mathcal{F}$ we have $\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A)$.

Probability Measure on a Discrete Sample Space

- Note a probability $\mathbb{P} : \Omega \mapsto [0,1]$ on a discrete sample space Ω uniquely specifies probabilities of all events $A_k = \{\omega_k\}$.
- Notation: it is common to write $\mathbb{P}(\{\omega_k\}) = \mathbb{P}(\omega_k) = p_k$.
- The theorem shows that any probability on a discrete sample space Ω generates a unique probability measure on $(\Omega, \mathcal{F})$.

Proposition

Let $\mathbb{P} : \Omega \mapsto [0,1]$ be a probability on a discrete sample space Ω. Then the unique probability measure on $(\Omega, \mathcal{F})$ generated by $\mathbb{P}$ satisfies for all $A \in \mathcal{F}$

$$\mathbb{P}(A) = \sum_{\omega_k \in A} \mathbb{P}(\omega_k).$$

- The proof of the theorem presents no difficulties, since Ω is assumed to be discrete.

Example: Coin Flipping

Example (1.1)

- Let X be the number of "heads" when a **fair** coin is tossed twice.
- We choose the sample space Ω to be $\Omega = \{0, 1, 2\}$ where a number $k \in \Omega$ represents the number of "heads."
- A single flip of a coin is a **Bernoulli trial**.
- The probability measure $\mathbb{P}$ on Ω is defined as

$$\mathbb{P}(k) = \begin{cases} 0.25, & \text{if } k = 0, 2, \\ 0.5, & \text{otherwise.} \end{cases}$$

- We recognise here the binomial distribution with $n = 2$ and $p = 0.5$.

Example: Coin Flipping

Example (1.2)

- We now suppose that the coin is not a fair one.
- Let the probability of "head" be p for some $p \neq 0.5$.
- Then the probability measure $\mathbb{P}$ is given by

$$\mathbb{P}(k) = \begin{cases} q^2, & \text{if } k = 0, \\ 2pq, & \text{if } k = 1, \\ p^2, & \text{if } k = 2, \end{cases}$$

  where $q := 1 - p$ is the probability of "tail" appearing.
- We obtain the binomial distribution with $n = 2$ and $0 < p < 1$.

PART 2

EQUIVALENCE OF PROBABILITY MEASURES

Radon-Nikodym Density

Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a discrete sample space Ω.

Definition (Equivalence of Probability Measures)

We say that the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are **equivalent** and we write $\mathbb{P} \sim \mathbb{Q}$ if for all $\omega \in \Omega$ we have that: $\mathbb{P}(\omega) > 0 \;\Leftrightarrow\; \mathbb{Q}(\omega) > 0$.

Definition (Radon-Nikodym Density)

The **Radon-Nikodym density** of $\mathbb{Q}$ with respect to $\mathbb{P}$ is the random variable $L : \Omega \to \mathbb{R}_+$ given by

$$L(\omega) = \frac{\mathbb{Q}(\omega)}{\mathbb{P}(\omega)}.$$

Note that

$$\mathbb{E}_{\mathbb{Q}}\left(X\right) = \sum_{\omega \in \Omega} X(\omega)\,\mathbb{Q}(\omega) = \sum_{\omega \in \Omega} X(\omega)L(\omega)\,\mathbb{P}(\omega) = \mathbb{E}_{\mathbb{P}}\left(LX\right).$$

Example: Radon-Nikodym Density

- The sample space Ω is defined as $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$.
- Consider the probability measures $\mathbb{P}$ and $\mathbb{Q}$ on Ω given by

$$\begin{aligned}(\mathbb{P}(\omega_1), \mathbb{P}(\omega_2), \mathbb{P}(\omega_3), \mathbb{P}(\omega_4)) &= \left(\frac{1}{8}, \frac{3}{8}, \frac{2}{8}, \frac{2}{8}\right)\\ (\mathbb{Q}(\omega_1), \mathbb{Q}(\omega_2), \mathbb{Q}(\omega_3), \mathbb{Q}(\omega_4)) &= \left(\frac{4}{8}, \frac{1}{8}, \frac{2}{8}, \frac{1}{8}\right).\end{aligned}$$

- It is clear that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, that is, $\mathbb{P} \sim \mathbb{Q}$.
- The Radon-Nikodym density L of $\mathbb{Q}$ with respect to $\mathbb{P}$ equals

$$L = (L(\omega_1), L(\omega_2), L(\omega_3), L(\omega_4)) = \left(4, \frac{1}{3}, 1, \frac{1}{2}\right).$$

- Check that for any random variable X: $\mathbb{E}_{\mathbb{Q}}(X) = \mathbb{E}_{\mathbb{P}}(LX)$.

PART 3

EXPECTATION OF A RANDOM VARIABLE

Expectation of a Random Variable

Definition (Expectation)

Let X be a random variable on a discrete sample space Ω endowed with a probability measure $\mathbb{P}$. The **expectation** (the **expected value** or the **mean value)** of X is defined to be

$$\mathbb{E}_{\mathbb{P}}(X) = \mu := \sum_{k \in I} X(\omega_k)\,\mathbb{P}(\omega_k) = \sum_{k \in I} x_k p_k.$$

- Note that the expectation of a random variable can be seen as the weighted average.
- Since it is impossible to predict which event will occur in the future, the expected value could be helpful when making decisions.

Expectation Operator

- Any random variable defined on a finite set Ω admits the expectation.
- When the sample space Ω is countable, but infinite, we say that X is $\mathbb{P}$-**integrable** whenever $\mathbb{E}_{\mathbb{P}}\left(|X|\right) = \sum_{k \in I} |X(\omega_k)|\, \mathbb{P}(\omega_k) < \infty$.
- Then the expectation $\mathbb{E}_{\mathbb{P}}\left(X\right)$ is well defined and finite.

Theorem (1.1)

Let X and Y be two $\mathbb{P}$-integrable random variables and $\mathbb{P}$ be a probability measure on a discrete sample space Ω. Then for all $\alpha, \beta \in \mathbb{R}$

$$\mathbb{E}_{\mathbb{P}}(\alpha X + \beta Y) = \alpha\, \mathbb{E}_{\mathbb{P}}(X) + \beta\, \mathbb{E}_{\mathbb{P}}(Y).$$

Hence $\mathbb{E}_{\mathbb{P}}\left(\cdot\right) : \mathcal{X} \to \mathbb{R}$ is a linear map on the space $\mathcal{X}$ of $\mathbb{P}$-integrable random variables.

Expectation Operator

Proof of Theorem 1.1.

We note that for arbitrary real numbers α and β

$$\mathbb{E}_{\mathbb{P}}\left(|\alpha X + \beta Y|\right) \leq |\alpha|\,\mathbb{E}_{\mathbb{P}}\left(|X|\right) + |\beta|\,\mathbb{E}_{\mathbb{P}}\left(|Y|\right) < \infty$$

so that the random variable $\alpha X + \beta Y$ belongs to $\mathcal{X}$. Moreover

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left(\alpha X + \beta Y\right) &= \sum_{k \in I} \left(\alpha X(\omega_k) + \beta Y(\omega_k)\right) \mathbb{P}(\omega_k) \\
&= \sum_{\omega_k \in \Omega} \alpha X(\omega_k)\,\mathbb{P}(\omega_k) + \sum_{k \in I} \beta Y(\omega_k)\,\mathbb{P}(\omega_k) \\
&= \alpha \sum_{k \in I} X(\omega_k)\,\mathbb{P}(\omega_k) + \beta \sum_{k \in I} Y(\omega_k)\,\mathbb{P}(\omega_k) \\
&= \alpha\,\mathbb{E}_{\mathbb{P}}(X) + \beta\,\mathbb{E}_{\mathbb{P}}(Y).
\end{aligned}$$

□

Expectation: Coin Flipping

Example (1.3)

- A fair coin is tossed three times. The player receives one dollar each time "head" appears and loses one dollar when "tail" occurs.
- Let the random variable X represent the player's payoff.
- The sample space Ω is defined as $\Omega = \{0, 1, 2, 3\}$ where $k \in \Omega$ represents the number of times "head" occurs.
- The probability measure is given by

$$\mathbb{P}(k) = \begin{cases} 1/8, & \text{if } k = 0, 3, \\ 3/8, & \text{if } k = 1, 2. \end{cases}$$

- This is the binomial distribution with $n = 3$ and $p = 0.5$.

Expectation: Coin Flipping

Example (1.3 Continued)

- The random variable X is defined as

$$X(k) = \begin{cases} -3, & \text{if } k = 0, \\ -1, & \text{if } k = 1, \\ 1, & \text{if } k = 2, \\ 3, & \text{if } k = 3. \end{cases}$$

- Hence the player's expected payoff equals

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}(X) &= \sum_{k=0}^{3} X(k)\mathbb{P}(k) \\ &= \frac{-3}{8} + \left(\frac{-3}{8}\right) + \frac{3}{8} + \frac{3}{8} \\ &= 0. \end{aligned}$$

Expectation of a Function of a Random Variable

Function of a random variable.

- Let X be a random variable and $\mathbb{P}$ be a probability measure on a discrete sample space Ω. We define $Y = f(X)$ where $f : \mathbb{R} \to \mathbb{R}$ is an arbitrary function.
- Then Y is also a random variable on the sample space Ω endowed with the same probability measure $\mathbb{P}$. Moreover,

$$\mathbb{E}_{\mathbb{P}}(Y) = \mathbb{E}_{\mathbb{P}}(f(X)) = \sum_{k \in I} f(X(\omega_k))\mathbb{P}(\omega_k).$$

- If a random variable X is deterministic then

$$\mathbb{E}_{\mathbb{P}}(X) = X, \quad \mathbb{E}_{\mathbb{P}}(f(X)) = f(X).$$

PART 4

VARIANCE OF A RANDOM VARIABLE

Risky Investments

- When deciding whether to invest in a given portfolio, an agent may be concerned with the "risk" of his investment.

Example (1.4)

Consider an investor who is given an opportunity to choose between the following two options:

1. The investor either receives or loses 1,000 dollars with equal probabilities. This random payoff is denoted by X_1.
2. The investor either receives or loses 1,000,000 dollars with equal probabilities. We denote this random payoff as X_2.

Hence in both scenarios the expected value of the payoff equals 0

$$\mathbb{E}_{\mathbb{P}}\left(X_1\right) = \mathbb{E}_{\mathbb{P}}\left(X_2\right) = 0.$$

The following question arises: which payoff is preferred?

Variance of a Random Variable

Definition (Variance)

The **variance** of a random variable X on a discrete sample set Ω is defined as

$$Var\,(X) = \sigma^2 := \mathbb{E}_{\mathbb{P}}\big[\,(X-\mu)^2\,\big]$$

where $\mathbb{P}$ is a probability measure on Ω.

- Variance is a measure of the spread of a random variable about its mean and also a measure of uncertainty.
- In finance, it is common to identify the variance of the price of a security with its degree of "risk".
- Note that $Var\,(X) = \sigma^2 \geq 0$ and it equals 0 if and only if X is deterministic.

Variance of a Random Variable

Example (1.4 Continued)

- The variance of option 1 equals

$$Var\,(X_1) = \frac{(1000-0)^2}{2} + \frac{(-1000-0)^2}{2} = 10^6.$$

- The variance of option 2 equals

$$Var\,(X_2) = \frac{(10^6-0)^2}{2} + \frac{(-10^6-0)^2}{2} = 10^{12}.$$

- We say that X_2 is more risky than X_1.
- A **risk-averse agent** would prefer the first option over the second.
- A **risk-loving agent** would prefer the second option over the first.

Variance of a Random Variable

Theorem (1.2)

Let X be a random variable and $\mathbb{P}$ be a probability measure on a discrete sample space Ω. Then the following equality holds

$$Var\left(X\right)=\mathbb{E}_{\mathbb{P}}\left(X^2\right)-\mu^2.$$

Proof of Theorem 1.2.

$$\begin{aligned} Var\left(X\right) &= \mathbb{E}_{\mathbb{P}}\left[(X-\mu)^2\right]=\mathbb{E}_{\mathbb{P}}\left(X^2-2\mu X+\mu^2\right) \quad \text{(linearity)} \\ &= \mathbb{E}_{\mathbb{P}}\left(X^2\right)-2\mu\,\mathbb{E}_{\mathbb{P}}\left(X\right)+\mu^2=\mathbb{E}_{\mathbb{P}}\left(X^2\right)-\mu^2. \end{aligned}$$

□

Independence of Random Variables

Definition (Independence)

Two discrete random variables X and Y are called **independent** if for all $x, y \in \mathbb{R}$

$$\mathbb{P}\left(X = x, Y = y\right) = \mathbb{P}\left(X = x\right)\mathbb{P}\left(Y = y\right)$$

where $\mathbb{P}\left(X = x\right)$ is the probability of the event $\{X = x\}$.

- If X and Y are independent, then $\mathbb{E}_{\mathbb{P}}\left(XY\right) = \mathbb{E}_{\mathbb{P}}\left(X\right)\mathbb{E}_{\mathbb{P}}\left(Y\right).$

Theorem (1.3)

Let X and Y be independent discrete random variables. Then for arbitrary $\alpha, \beta \in \mathbb{R}$,

$$Var(\alpha X + \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y).$$

Independence of Random Variables

Proof of Theorem 1.3.

Let $\mathbb{E}_{\mathbb{P}}(X) = \mu_X$ and $\mathbb{E}_{\mathbb{P}}(Y) = \mu_Y$. From Theorem 1.1, we have

$$\mathbb{E}_{\mathbb{P}}(\alpha X + \beta Y) = \alpha \mu_X + \beta \mu_Y .$$

Using Theorem 1.2, we obtain

$$\begin{aligned}
Var(\alpha X + \beta Y) &= \mathbb{E}_{\mathbb{P}}\left\{(\alpha X + \beta Y)^2\right\} - (\alpha \mu_X + \beta \mu_Y)^2 \\
&= \alpha^2\, \mathbb{E}_{\mathbb{P}}\left(X^2\right) + 2\alpha\beta\, \mathbb{E}_{\mathbb{P}}(XY) + \beta^2\, \mathbb{E}_{\mathbb{P}}(Y) \\
&- (\alpha \mu_X + \beta \mu_Y)^2 \\
&= \alpha^2 \left(\mathbb{E}_{\mathbb{P}}\left(X^2\right) - \mu_X^2\right) + \beta^2 \left(\mathbb{E}_{\mathbb{P}}\left(Y^2\right) - \mu_Y^2\right) \\
&\qquad + 2\alpha\beta \left(\mathbb{E}_{\mathbb{P}}(X)\, \mathbb{E}_{\mathbb{P}}(Y) - \mu_X \mu_Y\right) \\
&= \alpha^2 Var(X) + \beta^2 Var(Y).
\end{aligned}$$

□

PART 5

EXAMPLES OF DISCRETE DISTRIBUTIONS

Discrete Probability Distributions

Example (Binomial Distribution)

- Let $\Omega = \{0, 1, 2, \ldots, n\}$ be the sample space and let X be the number of successes in n independent trials where p is the probability of success in a single Bernoulli trial.
- The probability measure $\mathbb{P}$ is called the **binomial distribution** if

$$\mathbb{P}(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, \ldots, n$$

where

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

- Then

$$\mathbb{E}_{\mathbb{P}}(X) = np \quad \text{and} \quad Var(X) = np(1-p).$$

Example (Poisson Distribution)

- Let the sample space be $\Omega = \{0, 1, 2, \dots\}$.
- We take an arbitrary number $\lambda > 0$.
- The probability measure $\mathbb{P}$ is called the **Poisson distribution** if

$$\mathbb{P}(k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0, 1, 2, \dots .$$

- Then

$$\mathbb{E}_{\mathbb{P}}(X) = \lambda = Var(X).$$

- The Poisson distribution can be obtained as the limit of the binomial distribution when n tends to infinity and np_n tends to $\lambda > 0$.

Discrete Probability Distributions

Example (Geometric Distribution)

- Let $\Omega = \{1, 2, 3, \dots\}$ be the sample space and X be the number of independent trials to achieve the first success.
- Let p stand for the probability of a success in a single trial.
- The probability measure $\mathbb{P}$ is called the **geometric distribution** if

$$\mathbb{P}(k) = (1-p)^{k-1} p \quad \text{for } k = 1, 2, 3, \dots .$$

- Then

$$\mathbb{E}_{\mathbb{P}}(X) = \frac{1}{p} \quad \text{and} \quad Var(X) = \frac{1-p}{p^2} = \frac{q}{p^2}.$$

PART 6

CONTINUOUS RANDOM VARIABLES

Continuous Random Variables

Definition (Continuous Random Variable)

A random variable X on the sample space Ω is said to have a **continuous distribution** if there exists a real-valued function f such that

$$\begin{aligned} f(x) &\geq 0, \\ \int_{-\infty}^{\infty} f(x)\, dx &= 1, \end{aligned}$$

and for all real numbers $a < b$

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

Then $f : \mathbb{R} \to \mathbb{R}_+$ is called the **probability density function (pdf)** of a continuous random variable X.

Continuous Random Variables

Assume that X is a continuous random variable.

- Note that a probability density function need not satisfy the constraint $f(x) \leq 1$.
- A function $F(x)$ is called a **cumulative distribution function (cdf)** of a continuous random variable X if for all $x \in \mathbb{R}$

$$F(x) = \mathbb{P}\,(X \leq x) = \int_{-\infty}^{x} f(y)\,dy.$$

- The relationship between the pdf f and cdf F

$$F(x) = \int_{-\infty}^{x} f(y)dy \quad \Leftrightarrow \quad f(x) = \frac{d}{dx}F(x).$$

Continuous Random Variables

- The expectation and variance of a continuous random variable X are defined as follows

$$\mathbb{E}_{\mathbb{P}}(X) = \mu := \int_{-\infty}^{\infty} x\, f(x)\, dx,$$
$$Var(X) = \sigma^2 := \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx,$$

  or, equivalently,

$$Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2 = \mathbb{E}_{\mathbb{P}}\left(X^2\right) - \left(\mathbb{E}_{\mathbb{P}}(X)\right)^2$$

- The properties of expectations of discrete random variables carry over to continuous random variables, with probability measures replaced by pdfs and sums by integrals.

PART 7

EXAMPLES OF CONTINUOUS DISTRIBUTIONS

Continuous Probability Distributions

Example (Uniform Distribution)

- We say that X has the **uniform distribution** on an interval (a, b) if its pdf equals

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

- It is clear that

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_a^b \frac{1}{b-a}\, dx = 1.$$

- We have

$$\mathbb{E}_{\mathbb{P}}(X) = \frac{a+b}{2} \quad \text{and} \quad Var(X) = \frac{(b-a)^2}{12}.$$

Continuous Probability Distributions

Example (Exponential Distribution)

- We say that X has the **exponential distribution** on $(0, \infty)$ with the parameter $\lambda > 0$ if its pdf equals

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

- It is easy to check that

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{0}^{\infty} \lambda e^{-\lambda x}\, dx = 1.$$

- We have

$$\mathbb{E}_{\mathbb{P}}(X) = \frac{1}{\lambda} \quad \text{and} \quad Var(X) = \frac{1}{\lambda^2}.$$

Continuous Probability Distributions

Example (Gaussian Distribution)

- We say that X has the **Gaussian (normal) distribution** with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if its pdf equals

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } x \in \mathbb{R}.$$

  We write $X \sim N(\mu, \sigma^2)$.
- One can show that

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1.$$

- We have

$$\mathbb{E}_{\mathbb{P}}(X) = \mu \quad \text{and} \quad Var(X) = \sigma^2.$$

Continuous Probability Distributions

Example (Standard Normal Distribution)

- If we set $\mu = 0$ and $\sigma^2 = 1$ then we obtain the **standard normal distribution** $N(0,1)$ with the following pdf

$$n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{for } x \in \mathbb{R}.$$

- The cdf of the probability distribution $N(0,1)$ equals

$$N(x) = \int_{-\infty}^{x} n(u)\, du = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \quad \text{for } x \in \mathbb{R}.$$

- The values of $N(x)$ can be found in the **cumulative standard normal table** (also known as the **Z table**).
- If $X \sim N\left(\mu, \sigma^2\right)$ then $Z := \frac{X-\mu}{\sigma} \sim N(0,1)$.

LLN and CLT

Theorem (Law of Large Numbers)

Assume that $X_1, X_2, \ldots$ *are independent and identically distributed random variables with mean* μ. *Then with probability one*

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \quad \text{as } n \to \infty$$

Theorem (Central Limit Theorem)

Assume that $X_1, X_2, \ldots$ *are independent and identically distributed random variables with mean* μ *and variance* $\sigma^2 > 0$. *Then for all* $x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left\{ \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x \right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = N(x).$$

PART 8

CONDITIONAL DISTRIBUTIONS AND EXPECTATIONS

Conditional Distributions and Expectations

Definition (Conditional Probability)

For two random variables X_1 and X_2 and an arbitrary set A such that $\mathbb{P}(X_2 \in A) \neq 0$, we define the **conditional probability**

$$\mathbb{P}(X_1 \in A_1 \,|\, X_2 \in A_2) := \frac{\mathbb{P}(X_1 \in A_1, X_2 \in A_2)}{\mathbb{P}(X_2 \in A_2)}$$

and the **conditional expectation**

$$\mathbb{E}_{\mathbb{P}}(X_1 \,|\, X_2 \in A) := \frac{\mathbb{E}_{\mathbb{P}}(X_1 \mathbb{1}_{\{X_2 \in A\}})}{\mathbb{P}(X_2 \in A)}$$

where $\mathbb{1}_{\{X_2 \in A\}} : \Omega \to \{0, 1\}$ is the **indicator function** of $\{X_2 \in A\}$, that is,

$$\mathbb{1}_{\{X_2 \in A\}}(\omega) = \begin{cases} 1, & \text{if } X_2(\omega) \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Discrete Case

- Assume that X and Y are discrete random variables

$$p_i = \mathbb{P}(X = x_i) > 0 \quad \text{for } i = 1, 2, \dots \quad \text{and} \quad \sum_{i=1}^{\infty} p_i = 1,$$

$$\widehat{p}_j = \mathbb{P}(Y = y_j) > 0 \quad \text{for } j = 1, 2, \dots \quad \text{and} \quad \sum_{j=1}^{\infty} \widehat{p}_j = 1.$$

Definition (Conditional Distribution and Expectation)

Then the **conditional distribution** equals

$$p_{X \,|\, Y}(x_i \,|\, y_j) = \mathbb{P}(X = x_i \,|\, Y = y_j) := \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{p_{i,j}}{\widehat{p}_j}$$

and the **conditional expectation** $\mathbb{E}_{\mathbb{P}}(X \,|\, Y)$ is given by

$$\mathbb{E}_{\mathbb{P}}(X \,|\, Y = y_j) := \sum_{i=1}^{\infty} x_i \, \mathbb{P}(X = x_i \,|\, Y = y_j) = \sum_{i=1}^{\infty} x_i \, \frac{p_{i,j}}{\widehat{p}_j}.$$

Discrete Case

- It is easy to check that

$$p_i = \mathbb{P}(X = x_i) = \sum_{j=1}^{\infty} \mathbb{P}(X = x_i \,|\, Y = y_j)\, \mathbb{P}(Y = y_j) = \sum_{j=1}^{\infty} p_{X|Y}(x_i \,|\, y_j)\, \widehat{p}_j.$$

- The expected value $\mathbb{E}_{\mathbb{P}}(X)$ satisfies

$$\mathbb{E}_{\mathbb{P}}(X) = \sum_{j=1}^{\infty} \mathbb{E}_{\mathbb{P}}(X \,|\, Y = y_j)\, \mathbb{P}(Y = y_j).$$

Definition (Conditional cdf)

The **conditional cdf** $F_{X|Y}(\cdot \,|\, y_j)$ of X given Y is defined for all y_j such that $\mathbb{P}(Y = y_j) > 0$ by

$$F_{X|Y}(x \,|\, y_j) := \mathbb{P}(X \leq x \,|\, Y = y_j) = \sum_{x_i \leq x} p_{X|Y}(x_i \,|\, y_j).$$

Hence $\mathbb{E}_{\mathbb{P}}(X \,|\, Y = y_j)$ is the mean of the conditional distribution.

Continuous Case

- Assume that the continuous random variables X and Y have the joint pdf $f_{X,Y}(x,y)$.

Definition (Conditional pdf and cdf)

The **conditional pdf** of Y given X is defined for every x such that $f_X(x) > 0$ and equals

$$f_{Y|X}(y \,|\, x) := \frac{f_{X,Y}(x,y)}{f_X(x)} \quad \text{for } y \in \mathbb{R}.$$

The **conditional cdf** of Y given X equals

$$F_{Y|X}(y \,|\, x) := \mathbb{P}(Y \leq y \,|\, X = x) = \int_{-\infty}^{y} \frac{f_{X,Y}(x,u)}{f_X(x)}\, du.$$

Continuous Case

Definition (Conditional Expectation)

The **conditional expectation** of Y given X is defined for all x such that $f_X(x) > 0$ by

$$\mathbb{E}_{\mathbb{P}}(Y \,|\, X = x) := \int_{-\infty}^{\infty} y \, dF_{Y|X}(y \,|\, x) = \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x, y)}{f_X(x)} \, dy.$$

- An important property of conditional expectation is that

$$\mathbb{E}_{\mathbb{P}}(Y) = \mathbb{E}_{\mathbb{P}}(\mathbb{E}_{\mathbb{P}}(Y \,|\, X)) = \int_{-\infty}^{\infty} \mathbb{E}_{\mathbb{P}}(Y \,|\, X = x) f_X(x) \, dx.$$

- Hence the expectation $\mathbb{E}_{\mathbb{P}}(Y)$ can be determined by first computing $\mathbb{E}_{\mathbb{P}}(Y|X)$ and then integrating with respect to the pdf of X.