CHAPTER 5: RANDOM VARIABLES

6.1. DISCRETE RANDOM VARIABLES

In probability, an experiment may produce different outcomes. A random variable gives a numerical value to each outcome. This section explains discrete random variables, their probability distributions, their graphical representation, and their cumulative distribution functions.

6.1.1. Random Variable and Probability Distribution

Definition 1: Random variable

A random variable is a rule that assigns a real number to each possible outcome of a random experiment. It is usually denoted by a capital letter such as X or Y.

Example 1: Tossing two coins

Consider an experiment in which two fair coins are tossed. The possible outcomes are HH, HT, TH, and TT, where H means heads and T means tails. Let X be the number of heads obtained.

$\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

The values of X are:

If the outcome is TT, then X = 0.

If the outcome is HT or TH, then X = 1.

If the outcome is HH, then X = 2.

Therefore, X can take only the values 0, 1, and 2.

$\text{X}(\Omega) = \{0, 1, 2\}$

Definition 2: Discrete random variable

A random variable is called discrete if it can take a finite number of values or a countable list of values. In simple terms, its possible values can be listed one by one.

Examples: the number of heads in several coin tosses, the number of defective items in a sample, and the number of particles detected during a fixed time interval.

Definition 3: Probability distribution

The probability distribution of a discrete random variable describes all the possible values of the variable and the probability of each value.

For a discrete random variable X, the probability that X takes the value x is written P(X = x). Each probability must be between 0 and 1, and the sum of all the probabilities must be equal to 1.

In symbols:

$0 \leq P(X = x) \leq 1$ and $\Sigma P(X = x) = 1$

$X=x_i$	x_1	x_2	…	x_p
$P(X=x_i)$	$P(X=x_1)$	$P(X=x_2)$	….	$P(X=x_p)$

Example 2: Probability distribution for two coin tosses

The four outcomes HH, HT, TH, and TT are equally likely. Since X is the number of heads, we obtain the following distribution:

$\Omega = \{ HH, HT, TH, TT\}$

$X(\Omega) = \{0, 1, 2\}$

(X=0) mean that we have no H, this events is {TT}, hence $P(X=0) = \frac{1}{4}$

(X=1) mean that we have one H, this events is {HT, TH}, hence $P(X=1) = \frac{2}{4}$

(X=2) mean that we have two H, this events is { HH}, hence $P(X=1) = \frac{1}{4}$

Value of X	0	1	2
Probability	1/4	2/4 = 1/2	1/4

Check:

1/4 + 1/2 + 1/4 = 1. The probabilities therefore form a valid distribution.

6.1.2. Graphical Representation of a Discrete Distribution

A discrete probability distribution can be represented by a diagram made of vertical line segments. Each line is drawn above a possible value x, and its height is the probability P(X = x). A point may be placed at the top of each line. The distinction for a discrete variable, probabilities are assigned to individual values

Figure 1. Probability distribution for the number of heads in two fair coin tosses.

In this example, the line above 1 is the highest because P(X = 1) = 1/2. This means that obtaining exactly one head is more likely than obtaining zero heads or two heads.

6.1.3. Cumulative Distribution Function

Definition 4: Cumulative distribution function

The cumulative distribution function of X, denoted by F, gives the probability that X is less than or equal to a given real number x.

$$\mathbf{F(x) = P(X \leq x)}$$

To calculate F(x), we add the probabilities of all values of X that are less than or equal to x.

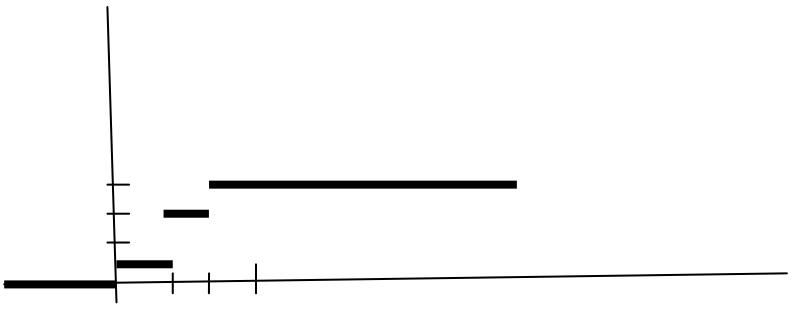
Example 3: Cumulative distribution for two coin tosses

Recall that X takes the values 0, 1, and 2, with probabilities 1/4, 1/2, and 1/4.

Condition on x	F(x)
$x < 0$	0
$0 \leq x < 1$	1/4
$1 \leq x < 2$	1/4 + 1/2 = 3/4
$x \geq 2$	1

For $x \geq 2$, $F(x) = P(X \leq 2) = P(X= 0) + P(X= 1) + P(X= 2) = 1$ (because all possible values of X are then included).

Thus, the cumulative distribution function is a step-shaped function. It never decreases, its values remain between 0 and 1, and it approaches 1 when x becomes sufficiently large.



Worked Exercise

A fair six-sided die is rolled once. Let X be the number shown on the upper face.

1. Is X discrete? 2. Give its probability distribution. 3.

Let $F(X)$ the distribution function defined by $F(a) = P(X \leq a)$. Calculate F(3).

Solution

1. X is discrete because it can take only the six values 1, 2, 3, 4, 5, and 6.

2. Since the die is fair, each face has probability 1/6. Thus, P(X = 1) = P(X = 2) = ... = P(X = 6) = 1/6.

3. F(3) = P(X ≤ 3) = P(X = 1) + P(X = 2) + P(X = 3) = 1/6 + 1/6 + 1/6 = 1/2.

Lecture Notes: Expected Value, Variance & Standardized Variables

Probability and Statistics | Course Materials in English

1. Key Terminology (Glossary)

The following bilingual table summarizes the main terms needed when teaching or presenting these statistical concepts in English:

French Term	English Term	Notation / Symbol
Variable aléatoire	Random variable	X, Y, Z
Espérance mathématique	Expected value (or Expectation)	E(X) or μ
Variance	Variance	Var(X) or V(X) or σ^2
Écart-type	Standard deviation	σ(X) or σ
Variable aléatoire centrée	Centered random variable	X - E(X)
Variable aléatoire réduite	Scaled / Standardized variable	(X - E(X)) / σ(X)
Variable centrée réduite	Standardized random variable	Z or Y

6.3 Standardized Random Variable

6.3.1 Centered Random Variable

A random variable *X* is **centered** by subtracting its expected value *E(X)*:

Formula:

$$\boldsymbol{X_centered} = \boldsymbol{X} - \boldsymbol{E(X)}$$

Property: The mean of any centered variable is always zero: E(X - E(X)) = 0.

1. Intuitive Concept

Imagine studying exam grades for a class, where the exam is scored out of 20. The class average is **12/20**.

- If a student scores 15, they are above average.
- If a student scores 9, they are below average.

Instead of analyzing the raw grades ($X = 9, 12, 15$), we can look at each student's **deviation from the mean**:

- The student with 15 is at (+3) points relative to the mean (15 - 12 = 3).
- The student with 9 is at (-3) points relative to the mean (9 - 12 = -3).
- The student with 12 is at (0) points relative to the mean (12 - 12 = 0).

This new value (the distance from the mean) represents the **centered variable**. The overall average of these new values is now **0**, because the positive deviations exactly balance out the negative deviations.

Why Do We Center a Random Variable?

The main purpose of centering a random variable (i.e., calculating ($X_{centered}$ = X - E(X)) is to change our reference point, allowing us to measure deviations from the mean rather than absolute values.

Here are the 4 main reasons why centering is useful:

1. Measuring Relative Deviations Instead of Absolute Values

Centering instantly tells you whether an outcome is above or below average:

- If ($X_{centered}$ > 0): The observation is **above the mean**.
- If ($X_{centered}$ < 0): The observation is **below the mean**.
- If ($X_{centered}$ = 0): The observation is **exactly at the mean**.

2. The Mandatory First Step for Standardization

To calculate a standardized (scaled) random variable, centering is the required first step:

$$\mathbf{Z} = \frac{\mathbf{X} - \mathbf{E(X)}}{\boldsymbol{\sigma(X)}}$$

Centering shifts the mean to zero(Le centrage amène la moyenne à 0).

- Dividing by the standard deviation scales the spread to 1(La division par l'écart-type amène l'écart-type à 1).

3. Simplifying Variance and Mathematical Calculations

In probability and statistics, working with centered variables significantly simplifies algebraic formulas:

- The formula for variance simply becomes the expected value of the squared centered variable:

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

- In linear regression and covariance, centering variables eliminates constant terms and streamlines correlation analysis.

4. Eliminating Bias and Shifts

In signal processing, finance, or machine learning, centering data removes systematic shifts or DC components, ensuring that analysis focuses purely on real data variability(Dans le traitement du signal, la finance ou l'apprentissage automatique (Machine Learning), centrer les données permet de supprimer la composante continue ou la dérive systématique (le biais) pour ne se concentrer que sur la variabilité réelle des données).

6.2 Expected Value and Variance

The **expected value** represents the long-run theoretical average outcome of a random variable. The **variance** measures the dispersion or spread of the random variable around its expected value.

Formulas & Mathematical Definitions

• **Expected Value (Discrete Case):** $E(X) = \sum [\, x_i \cdot P(X = x_i) \,]$

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

• **Standard Deviation:** $\sigma(X) = \sqrt{\mathrm{Var}(X)}$

Key Linearity Properties

• $E(aX + b) = a \cdot E(X) + b$

• $\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$

• $\sigma(aX + b) = |a| \cdot \sigma(X)$

6.3.2 Scaled / Standardized Random Variable

A centered variable is **scaled (or standardized)** by dividing by its standard deviation *σ(X)*:

$$\mathbf{Z} = \frac{\mathbf{X} - \mathbf{E(X)}}{\boldsymbol{\sigma}(\mathbf{X})}$$

Properties Of The Standardized Variable Z

• Mean of Z: $E(Z) = 0$

• Variance of Z: $\mathrm{Var}(Z) = 1$

• Standard Deviation of Z: $\sigma(Z) = 1$

6.4 Random Variables And Probability Distributions

6.4.1. Probability Density Function

A random variable is continuous when it can take any value in an interval of the real line. Examples include the life time of a component, the temperature of a body, the length of a rod, and the time between two radioactive detections**.** The distinction for a continuous variable, probabilities are calculated over intervals.

For a continuous random variable X, probabilities are described by a probability density function (PDF), usually denoted by f(x). The density is not itself the probability that X takes one exact value. Instead, probabilities are obtained by calculating areas under the density curve.

-Definition and properties

A function f is a **probability density function** of X if it satisfies the following conditions:

1- $f(x) \geq 0$ for every real number x.

2- The total area under the curve is equal to 1: the integral of f(x) over the whole real line is 1.

That is a probability density function must satisfy: $\int_{-\infty}^{+\infty} \boldsymbol{f(x)dx} = \mathbf{1}.$

3- For any $a \leq b$, the probability that X lies between a and b is the integral of f(x) from a to b.

4- For a continuous random variable, the probability of one exact value is zero: $P(X = a) = 0$. Consequently, $P(a \leq X \leq b)$, $P(a < X \leq b)$, and $P(a \leq X < b)$ are equal. that is

$$\boldsymbol{P(a \leq X \leq b)} = \boldsymbol{P(a < X \leq b)} = P(a \leq X < b)$$

- Cumulative Distribution Function

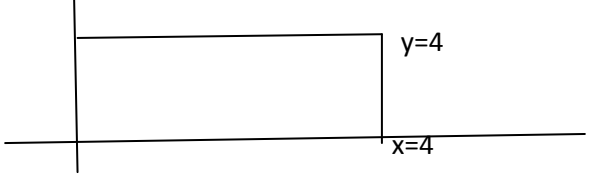
The cumulative distribution function (CDF) is $F(x) = P(X \leq x)$. For a continuous random variable, F(x) is the integral of the density from minus infinity to x. When the derivative exists, f(x) = F'(x). In other word $\int_{-\infty}^{x} f(x) = F(x)$ and $F'(x) = f(x)$

Example 1

Suppose X has density $f(x) = \frac{1}{4}$ for $0 \leq x \leq 4$, and f(x) =0 elsewhere

1. Check that f is a density.

The function is non-negative.



The total area is the area of a rectangle of width 4 and height 1/4.

$\int_{-\infty}^{+\infty} \boldsymbol{f(x)dx}$ = Area = 4 × 1/4 = 1= , so the density is valid.

Answer: f is a valid probability density function.

2. Calculate P(1 ≤ X ≤ 3). The required probability is the area between 1 and 3. P(1 ≤ X ≤ 3) = (3 − 1) × 1/4 = 2/4. P(1 ≤ X ≤ 3) = 0.5.

3. Calculate P(X = 2). A single point has zero width, so the area above that point is zero. P(X = 2) = 0.

3- Expectation and variance

If the relevant integrals exist, the expectation (mean) is E(X) = integral of x f(x) over the real line. The variance measures dispersion around the mean: Var(X) = E(X²) − [E(X)]², where E(X²) is the integral of x² f(x). The standard deviation is the square root of the variance. In other word

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\ ,\ E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x),\ \sigma = \sqrt{V(X)}$$

Example 2: mean and variance of the uniform distribution on [0, 4]

For f(x) = 1/4 on [0, 4],

E(X) = $\int_0^4 \frac{1}{4}$ x dx = (1/4) $\int_0^4$ x dx = $\frac{1}{4} \times 8$ =2.

E(X²) = $\int_0^4 \frac{1}{4} x^2$ dx = $\frac{1}{4} \times \int_0^4 x^2 dx$ = $\frac{1}{4}[\frac{1}{3} x^3]$ = $\frac{1}{4}$ ×(64/3) = 16/3.

Var(X) = 16/3 − 2² = 16/3 − 12/3 = 4/3.

E(X) = 2 , Var(X) = 4/3, and $\sigma = \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}}$

6.4.2. Bernoulli trial and Bernoulli distribution(تجربة برنولي وتوزيع برنولي)

1- Bernoulli trial

A Bernoulli trial is a random experiment with exactly two possible outcomes: success and failure. The word success simply identifies the outcome of interest; it does not necessarily mean something positive.

Let X = 1 if success occurs and X = 0 if failure occurs. If the probability of success is p, then the probability of failure is 1 − p. We write X ~ Bernoulli(p).

Probability law: P(X = 1) = p and P(X = 0) = 1 − p.

X=x_i	1	0
P(X = x_i)	p	1-p
x_i^2	1	0

E(X) = 1×p +0× $(p-1)$ = p

$Var(X) = p - p^2 = p(1-p)$

Example 3

one component is tested

A component is defective with probability 0.08. Define X = 1 if the component is defective and X = 0 otherwise. The Bernoulli model is appropriate because there are two outcomes for one test.

Here p = 0.08 and 1 − p = 0.92.

P(X = 1) = 0.08 and P(X = 0) = 0.92.

E(X) = 0.08 and Var(X) = 0.08 × 0.92 = 0.0736.

2- Binomial distribution

The binomial distribution counts the number of successes in a fixed number n of independent Bernoulli trials, when the probability p of success is the same at every trial.

Conditions:

1. The number of trials $\boldsymbol{n}$ is fixed in advance.
2. Each trial has two possible outcomes: success or failure.
3. The trials are independent.
4. The probability of success p is constant from one trial to another.
5. If X is the number of successes, we write $X \sim B(n, p)$.

$For\ k\ =\ 0, 1, \ldots, n,$ the probability is $\mathrm{P(X\ =\ }k) = \binom{k}{n} p^k (1-p)^{1-k}$

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

Mean: E(X) = np. Variance: Var(X) = np(1 − p).

Example 4: three tosses of a fair coin

A fair coin is tossed three times. Let X be the number of heads.

The trials are independent, n = 3 and p = 1/2. Thus X ~ B(3, 1/2).

The probability of exactly two heads is P(X = 2) = C(3,2)(1/2)²(1/2)¹.

Since C(3,2) = 3, the probability P(X = 2) = 0.375.

3/8 = 0.375.

E(X)= np = $3\times\frac{1}{2}$ = 1.5, and Var(X)= np(1 − p) = $\frac{3}{2}\frac{1}{2}=\frac{3}{4}$= 0.75.

Example:

A detector registers a particle with probability 0.2 on each independent trial. It is used five times. Determine the probability of at least one success

Let X be the number of detections. X ~ B(5, 0.2).

P(X ≥ 1) = 1 − P(X = 0) = 1 − (0.8)^5.

Therefore P(X ≥ 1) = 1 − 0.32768.

P(X ≥ 1) = 0.67232.