Chapitre 2

Random Variables

I - Definitions

Example 1 :

We toss a fair coin twice and are interested in the number of times the side "heads" appears. To compute the probabilities of the various outcomes, we introduce a variable X denoting the number of "heads" obtained. The variable X can take the values $0, 1, 2$.

Example 2 :

A dart is thrown at a circular target of radius 50 cm, and we are interested in the distance between the dart and the center. We introduce a variable X, defined as the distance between the impact point and the center of the target. This variable can take any real value between 0 and 50.

In both cases, X takes real values that depend on the outcome of the random experiment. The values taken by X are therefore random. Thus, X is called a *random variable*.

Example 3 :

Suppose the sample space Ω is the set of students. We select one student at random (see graphic 2.1). To each student we associate two numerical quantities :

1. their weight in kilograms,
2. their height in meters.

Let W and H be the functions that assign to every student the numerical values w (weight) and h (height), respectively. Both W and H are **random variables**, since they take an element of the sample space (a student) and return a real number.

The random variable W takes as input a student and outputs his weight. Similarly, H takes a student and outputs his height.

From these two random variables, we can construct a third one : the *body mass index* (BMI), defined by

$$B = \frac{W}{H^2}.$$

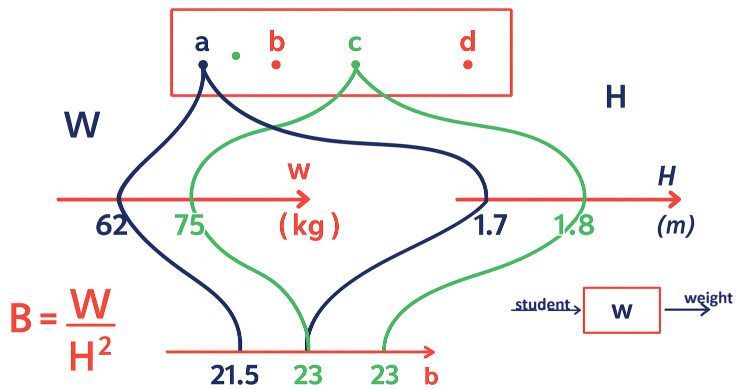


Figure 2.1 – Random variables associated with a randomly chosen student

Definition 1.1

Let Ω be a sample space associated with a random experiment, equipped with a probability measure. A *random variable* (r. v.) is a mapping

$$X : \Omega \longmapsto \mathbb{R}.$$

- ☞ A random variable is a function that assigns numerical values to random outcomes.
- ☞ By convention, a random variable is denoted by an uppercase letter (e.g., X), while the values it can take are denoted by lowercase letters $(x_1, x_2, \ldots, x_i, \ldots, x_n)$.
- ☞ The two random variables defined in the first two examples are of different types : the first is discrete, the second continuous.
- ☞ We can have several random variables defined on the same sample space.

II - Discrete Random Variables

Definition 2.1: Discrete random variable

A discrete random variable is a random variable that **takes only integer values**, in a finite or countably infinite subset of $\mathbb{Z}$.

To understand a random variable, it is essential to know which values occur most often and which are less frequent. Concretely, this amounts to determining the probabilities associated with each possible value of the variable.

Definition 2.2: Probability mass function

Associating to each possible value of a random variable the probability that the variable takes this value defines the *probability law* or *probability distribution* of the variable.

Let X be a discrete random variable taking values x_i. To compute $p(X = x_i)$, we collect all elementary events e_j such that $X(e_j) = x_i$, and we obtain

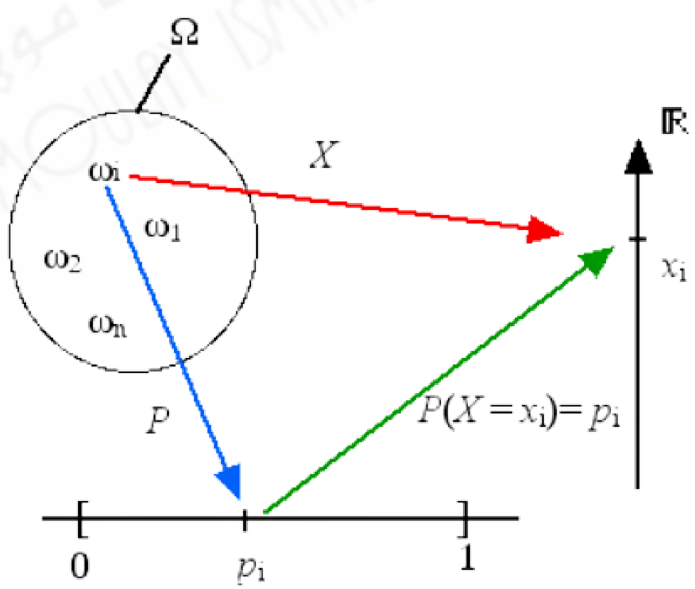
$$p(X = x_i) \;=\; p\big(\{\omega \in \Omega : X(\omega) = x_i\}\big).$$

The *probability mass function (pmf)* of X is the function

$$f : \mathbb{R} \longrightarrow [0, 1], \qquad f(x_i) = p(X = x_i),$$

and, of course,

$$\sum_i f(x_i) = 1.$$



Example 4 : *Case of Example 1.*

We have

$$\Omega = \left\{ \underbrace{(\text{tails}, \text{tails})}_{e_1}, \underbrace{(\text{tails}, \text{heads})}_{e_2}, \underbrace{(\text{heads}, \text{tails})}_{e_3}, \underbrace{(\text{heads}, \text{heads})}_{e_4} \right\},$$

and the variable

$$X = \text{“number of heads”}$$

takes the values

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

The pmf of X is :

$$\begin{aligned} f(0) &= p(X=0) = p(e_1) = \frac{1}{4}, \\ f(1) &= p(X=1) = p(e_2) + p(e_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \\ f(2) &= p(X=2) = p(e_4) = \frac{1}{4}, \\ f(x) &= 0 \quad \text{if } x \notin \{0, 1, 2\}. \end{aligned}$$

Its probability distribution can be presented in a table :

Value x_i	0	1	2	Total
$p(X = x_i)$	1/4	1/2	1/4	1

1) Cumulative distribution function (CDF)

In descriptive statistics (see the course “Probabilities and Descriptive Statistics” in semester 1), the notion of increasing cumulative frequencies is equivalent, in probability theory, to the cumulative distribution function.

Definition 2.3: Cumulative distribution function in the discrete case

The cumulative distribution function of a random variable X gives, for each real value x, the probability that X takes a value less than or equal to x. It is the sum of the probabilities of the values of X up to x. It is denoted F_X, and, when unambiguous, simply F.

$$\forall x \in \mathbb{R} : \quad F_X(x) = p(X \leqslant x) = \sum_{x_i \leqslant x} p(X = x_i).$$

The cdf is always increasing, takes values between 0 and 1, and is a very useful tool in theoretical work.

III - Continuous Random Variables

Definition 3.1: Continuous random variable

A random variable is said to be *continuous* if it can take all values of a finite or infinite interval.

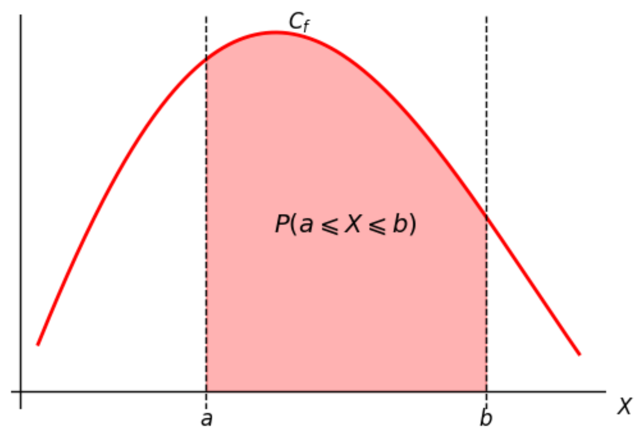
1) Probability density function (PDF)

Definition 3.2

A function f is a probability density function (PDF) if it has the following properties :

- ✓ It is nonnegative;
- ✓ It is integrable over $\mathbb{R}$,
  such that the probability that the random variable X lies between a and b, i.e., $p(a \leqslant X \leqslant b)$, equals the area between the x–axis, the graph of f, and the vertical lines $x = a$ and $x = b$,

$$p(a \leqslant X \leqslant b) = \int_a^b f(x)\, dx$$



- ✓ The total area under the graph is 1 :

$$\int_{\mathbb{R}} f(x)\, dx = 1.$$

Remark 3.3

Note that for a continuous random variable X we always have
$p(X = a) = 0$.
Indeed,
If f is continuous on an interval of the form $[a, a+h]$ with $h \to 0^+$, then
$p(a \leqslant X \leqslant a+h) = \int_a^{a+h} f(x)\,dx = h\,f(a+\theta h)$ with $(0 < \theta < 1)$ (Mean Value Theorem).
Thus, as $h \to 0$, $f(a+\theta h) \to f(a)$ and $hf(a+\theta h) \to 0$,
whence $p(a \leqslant X \leqslant a+h) \to p(X = a) = 0$.

2) Cumulative distribution function

Definition 3.4: Cdf in the continuous case

As for discrete random variables, one defines the cumulative distribution function F of a continuous random variable X :

$$F(x) = p(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt.$$

Property 3.5

1. F is continuous on $\mathbb{R}$, differentiable at every point where f is continuous, and $F' = f$.
2. F is increasing on $\mathbb{R}$.
3. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1.$
4. $p(a \leqslant X \leqslant b) = F(b) - F(a)$.

Proof:

1. Follows from $F(x) = \int_{-\infty}^{x} f(t)\,dt$.
2. Since $F' = f$, it is nonnegative on $\mathbb{R}$.
3. $\int_{-\infty}^{x} f(t)\,dt \to 0$ as $x \to -\infty$.
4. $p(a \leqslant X \leqslant b) = p(X < b) - p(X < a) = F(b) - F(a)$, hence $\int_{-\infty}^{b} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt = \int_{a}^{b} f(t)\,dt$.

■

Exercise 1

Let f be the function on $\mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x < 1, \\ \frac{2}{x^3} & \text{otherwise.} \end{cases}$$

1. Show that f is a probability density function for a random variable X.
2. Determine the cumulative distribution function F_X of X.

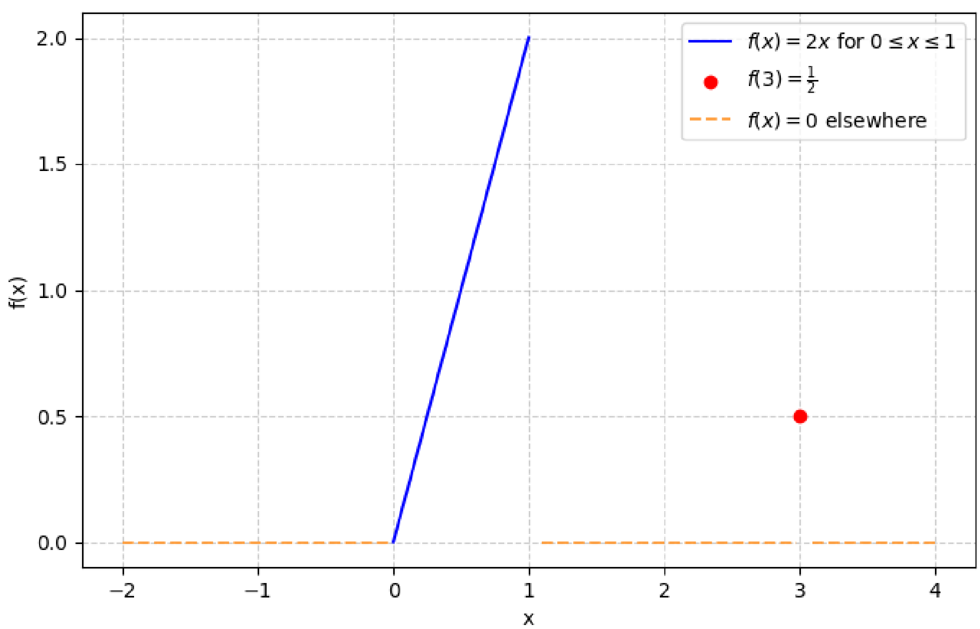
Example 5 :

Let f be the function on $\mathbb{R}$ defined by

$$f(x) = \begin{cases} 2x & \text{if } 0 \leqslant x < 1, \\ \frac{1}{2} & \text{if } x = 3, \\ 0 & \text{otherwise.} \end{cases}$$

1. Show that f is a probability density function for a random variable X.
2. Determine the cumulative distribution function F_X of X.

Solution :



1. f is clearly nonnegative, and piecewise integrable : on $]-\infty, 0[$, $]0,1[$, $]0,3[$ and $]3,+\infty[$ it is locally integrable. The right and left limits at 0, 1, and 3 exist and are finite. Moreover,

$$\int_{\mathbb{R}} f(t)\,dt = \cancel{\int_{-\infty}^{0} 0\,dt} + \int_0^1 2t\,dt + \cancel{\int_1^3 0\,dt} + \cancel{\int_3^{+\infty} 0\,dt} = 1.$$

2.
   - If $x < 0$, $F(x) = \int_{-\infty}^{x} 0\,dt = 0$,
   - If $0 \leqslant x \leqslant 1$, $F(x) = \int_{-\infty}^{0} 0\,dt + \int_0^x 2t\,dt = \left[t^2\right]_0^x = x^2$,
   - If $x > 1$, $F(x) = \int_{-\infty}^{0} 0\,dt + \int_0^1 2t\,dt + \underbrace{\int_1^x f(t)\,dt}_{=0} = 1$,

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ x^2 & \text{if } 0 \leqslant x \leqslant 1, \\ 1 & \text{if } x \geqslant 1. \end{cases}$$

Example 6 :

Let f be the function on $\mathbb{R}$ defined by

$$f(x) = \begin{cases} k\exp(-x) & \text{if } x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

1. Determine k so that f is the probability density function of a random variable X.
2. Determine the cumulative distribution function of X.
3. Compute $p(1 < X < 2)$.

Solution :

1. f must be nonnegative, so k must be positive. Moreover, f is Riemann-integrable on $\mathbb{R}$ (piecewise integrable : on $]-\infty, 0[$ and $]0, +\infty[$ it is locally integrable ; the limits $\lim_{0^+} f(x), \lim_{0^-} f(x)$ exist and are finite ; and $\lim_{A\to-\infty} \int_A^0 f(t)\,dt$ and $\lim_{B\to+\infty} \int_0^B ke^{-t}\,dt$ are finite).
   Furthermore, a probability density must satisfy $\int_{\mathbb{R}} f(t)\,dt = 1$,
   hence $\int_0^{+\infty} ke^{-t}\,dt = 1$. It follows that $\boxed{k = 1.}$
2. By definition, the cdf of X is
$$F(x) = \begin{cases} \int_0^x e^{-t}\,dt = 1 - e^{-x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$
3.
$$p(1 < X < 2) = F(2) - F(1) = \boxed{e^{-1} - e^{-2}} \approx 0.23.$$

IV - Descriptive Parameters of a Probability Distribution

1) Expectation of a probability distribution

Definition 4.1

1. Discrete case.
   - Let X be a discrete random variable taking finitely many values $x_1, x_2, \ldots, x_n$ with pmf f : $f(x_i) = p(X = x_i)$. The *expectation* of X, denoted $E(X)$, is
$$E(X) = \sum_{i=1}^{n} x_i f(x_i).$$
   - If X takes a countable set of values $x_1, x_2, \ldots, x_n, \ldots$, its *expectation* is
$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i),$$
     provided the series converges absolutely.
2. Continuous case. If X is continuous with density f, its *expectation* is
$$E(X) = \int_{\mathbb{R}} x\, f(x)\,dx,$$
   provided the function $x \mapsto xf(x)$ is integrable on $\mathbb{R}$.

2) Variance of a Probability Distribution

Definition 4.2

1. The *variance* of a random variable X is the mean of the squared deviations from its mean :

$$V(X) = E\left((X - E(X))^2\right).$$

The computation simplifies using :

König–Huygens formula

$$V(X) = E(X^2) - E(X)^2.$$

2. The *standard deviation* of X is the square root of its variance :

$$\sigma(X) = \sqrt{V(X)}.$$

In the case of a finite discrete random variable,

$$V(X) = \sum_{i=1}^{n}(x_i - E(X))^2 f(x_i) = \left(\sum_{i=1}^{n} x_i^2 f(x_i)\right) - E(X)^2.$$

In the case of a continuous random variable,

$$V(X) = \int_{\mathbb{R}} (x - E(X))^2 f(x)\,dx = \left(\int_{\mathbb{R}} x^2 f(x)\,dx\right) - E(X)^2.$$

3) Properties of expectation and variance

We summarize the main properties of these two parameters in a table.

Shift (origin)	Scale change	Affine transformation
$E(X+c) = E(X) + c$	$E(aX) = aE(X)$	$E(aX+c) = aE(X) + c$
$V(X+c) = V(X)$	$V(aX) = a^2V(X)$	$V(aX+c) = a^2V(X)$
$\sigma(X+c) = \sigma(X)$	$\sigma(aX) = \lvert a\rvert\sigma(X)$	$\sigma(aX+c) = \lvert a\rvert\sigma(X)$

Definition 4.3

- A random variable X is *centered* if its expectation is zero.
- A random variable X is *reduced* if its standard deviation equals 1.
- A centered and reduced random variable is said to be *standardized*.

4) A Pair of Random Variables

Marginal laws. From the joint law of (X, Y), one recovers the marginal laws :

$$p(X = x) = \sum_{y \in Y(\Omega)} p(X = x,\, Y = y), \qquad p(Y = y) = \sum_{x \in X(\Omega)} p(X = x,\, Y = y).$$

(If X or Y take infinitely many values, the corresponding series is absolutely convergent.)

Expectation. For any function $f : X(\Omega) \times Y(\Omega) \to \mathbb{R}$,

$$E[f(X,Y)] = \sum_x \sum_y f(x,y)\, p(X = x,\, Y = y).$$

In particular,

$$E(XY) = \sum_x \sum_y xy\, p(X = x, Y = y).$$

This directly yields the linearity :

$$E(X + Y) = E(X) + E(Y),$$

and, if X and Y are independent,

$$E(XY) = E(X)\, E(Y).$$

Proposition 4.4: Linearity of the Expectation

Let X and Y be real-valued random variables with finite expectations, and let $a, b \in \mathbb{R}$. Then

1) $E(aX + b) = a\, E(X) + b$,
2) $E(X + Y) = E(X) + E(Y)$.

Proof:
We give the proof in the discrete case. Let $p_{X,Y}(x, y)$ be the joint pmf, and

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

its marginals.

1) Affine change of variable.

$$E(aX + b) = \sum_x (ax + b)\, p_X(x) = a \sum_x x\, p_X(x) + b \sum_x p_X(x) = aE(X) + b.$$

2) Sum of two random variables.

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

■

Remark 4.5

The same properties hold for continuous variables.
If (X, Y) has joint density $f_{X,Y}$ and marginals

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy, \qquad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dx,$$

then, by Fubini's theorem,

$$E(X + Y) = E(X) + E(Y).$$

The identity $E(aX + b) = aE(X) + b$ follows from linearity of integrals.

4- a) Covariance

Definition 4.6: Covariance

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

(For infinite sums, the expectation is well-defined if the corresponding series are absolutely convergent.)

Covariance behaves linearly :

$$\mathrm{Cov}(aX + b,\ cY + d) = ac\, \mathrm{Cov}(X, Y).$$

If X and Y are independent, then

$$\mathrm{Cov}(X, Y) = 0.$$

⚠ The converse is false.

Proposition 4.7

1. If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.
2. In general,

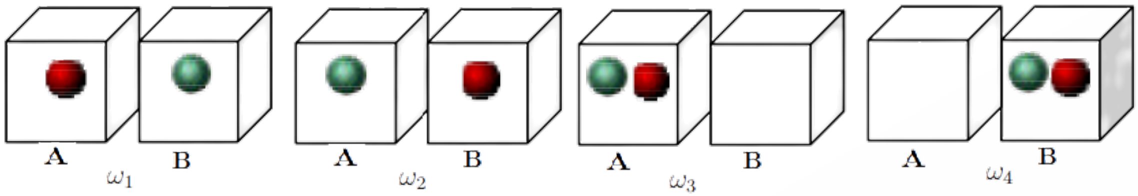
$$E(XY) = E(X)E(Y) + \mathrm{Cov}(X, Y), \qquad V(X + Y) = V(X) + V(Y) + 2\,\mathrm{Cov}(X, Y).$$

Example 7 :

Two balls, one red and one green, are placed at random into boxes A and B.
Define : X = number of balls in box A; Y = number of empty boxes.

1) **Sample space.**

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}.$$



2) **Distribution of X.**

$$X(\omega_1) = 1,\ X(\omega_2) = 1,\ X(\omega_3) = 2,\ X(\omega_4) = 0, \qquad X(\Omega) = \{0, 1, 2\}.$$

$$p(X = 0) = \tfrac{1}{4}, \qquad p(X = 1) = \tfrac{1}{2}, \qquad p(X = 2) = \tfrac{1}{4}.$$

$$E(X) = 1, \qquad V(X) = \tfrac{1}{2}.$$

3) **Distribution of Y.**

$$Y(\omega_1) = 0,\ Y(\omega_2) = 0,\ Y(\omega_3) = 1,\ Y(\omega_4) = 1, \qquad Y(\Omega) = \{0, 1\}.$$

$$p(Y = 0) = \tfrac{1}{2}, \qquad p(Y = 1) = \tfrac{1}{2},$$
$$E(Y) = \tfrac{1}{2}, \qquad V(Y) = \tfrac{1}{4}.$$

4) **Distribution of $X + Y$.**

$$X + Y(\Omega) = \{1, 3\},$$
$$p(X + Y = 1) = \tfrac{3}{4}, \qquad p(X + Y = 3) = \tfrac{1}{4},$$
$$E(X + Y) = \tfrac{3}{2}, \qquad V(X + Y) = \tfrac{3}{4}.$$

5) **Distribution of XY.**

$$XY(\Omega) = \{0, 2\},$$
$$p(XY = 0) = \tfrac{3}{4}, \qquad p(XY = 2) = \tfrac{1}{4},$$
$$E(XY) = \tfrac{1}{2}, \qquad V(XY) = \tfrac{3}{4}.$$

6) **Independence.** Although

$$E(XY) = E(X)E(Y),$$

we have

$$p(X = 0, Y = 0) = 0 \neq \tfrac{1}{4} \cdot \tfrac{1}{2} = \tfrac{1}{8}.$$

Hence X and Y are *not independent*.

V - Markov and Chebyshev Inequalities

When studying a random variable X, it is often difficult, or impossible, to know its exact PMF or PDF. Even with limited information (such as its mean or variance), we can still obtain useful bounds using **concentration inequalities**, which estimate how likely X is to take unusually large values or deviate from its mean.

1) Markov Inequality

- Provides bounds for non-negative random variables.
- If $E(X)$ is small, then large values of X must be rare.

Proposition 5.1: Markov Inequality

If $X \geqslant 0$ and $a > 0$, then

$$p(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

Proof:
For continuous X :

$$E(X) = \int_0^\infty x f_X(x)\,dx \geqslant \int_a^\infty x f_X(x)\,dx \geqslant a \int_a^\infty f_X(x)\,dx = a\,p(X \geqslant a).$$

Hence the result. (The discrete case is identical, with sums instead of integrals.) ■

2) Chebyshev Inequality

- Applies to any random variable with finite mean μ and variance σ^2.
- Quantifies how rarely X deviates far from its mean.

Proposition 5.2: Chebyshev Inequality

$$p(|X-\mu| \geqslant c) \leqslant \frac{\sigma^2}{c^2}, \qquad c > 0.$$

Proof:
Since $(X-\mu)^2 \geqslant 0$, applying Markov's inequality to it gives

$$p(|X-\mu| \geqslant c) = p((X-\mu)^2 \geqslant c^2) \leqslant \frac{E\left((X-\mu)^2\right)}{c^2} = \frac{\sigma^2}{c^2}.$$

■