Probability

6. Common Probability Laws

Ewen Gallic

Bachelor in Economics
International Program in Economics and Management
Academic Year 2026

amU FEG
Aix Marseille Université

Foreword

- This chapter introduces the most **common probability distributions**, both **discrete** and **continuous**, together with their main properties (densities, distribution functions, moments, and key interpretations).
- The material in this chapter is adapted from:
  - lecture slides by Mathieu Faure (Aix–Marseille University), kindly shared,
  - Chapter 3 of Hansen (2022).
- **Interactive visual illustrations** made by Bertille Picard (ENSAI) are available on her website (Probability Theory course), notably for the Normal distributionn using Desmos:
  - Normal density,
  - Inflection points of the Normal distribution.
- A Shiny application displaying the PDF and CDF of several standard distributions is available at: https://ewengallic.shinyapps.io/Common-Prob-Laws

1. Discrete Probability Laws

1.1. Bernoulli Distribution

Motivation

Jakob Bernoulli (Basel, Switzerland, 1654 – 1705), by Niklaus Bernoulli (Wikimedia).

- The **Bernoulli distribution** models the outcome of an experiment with only **two possible outcomes**:
  - success / failure,
  - yes / no,
  - event occurs / does not occur.
- For an individual randomly sampled from a population, it can describe:
  - if they are employed ($X = 1$) or not ($X = 0$),
  - if they will default on a loan ($X = 1$) or not ($X = 0$),
  - if they will buy a product ($X = 1$) or not ($X = 0$).
- In policy evaluation, it can also represent treatment assignment: $X = 1$ if an individual/firm/household receives the treatment, $X = 0$ otherwise.

- In all these examples, **randomness** is summarized by **one probability parameter**.

Formal Definition

Definition 1. Bernoulli Distribution

*A random variable X with support $\mathcal{D}_X = \{0, 1\}$ follows a **Bernoulli distribution** with parameter $p \in [0, 1]$ if*

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

Notation

We write $X \sim \mathcal{B}(p)$.

Probability Mass Function

Definition 2. Probability Mass Function of a Bernoulli Distribution

*The **probability mass function** of a variable X that follows a Bernoulli distribution with parameter $p \in [0,1]$ is*

$$\mathbb{P}(X = x) = \begin{cases} p & \textit{if} \quad x = 1, \\ 1 - p & \textit{if} \quad x = 0. \end{cases}$$

Equivalently, for $x \in \{0,1\}$,

$$\mathbb{P}(X = x) = p^x (1-p)^{1-x}.$$

Cumulative Distribution Function

Definition 3. Cumulative Distribution Function of a Bernoulli Distribution

*The **cumulative distribution function** of a variable X that follows a Bernoulli distribution with parameter $p \in [0, 1]$ is*

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \textit{if} \quad x < 0, \\ 1 - p & \textit{if} \quad 0 \leq x < 1, \\ 1 & \textit{if} \quad x \geq 1. \end{cases}$$

Visual Representation

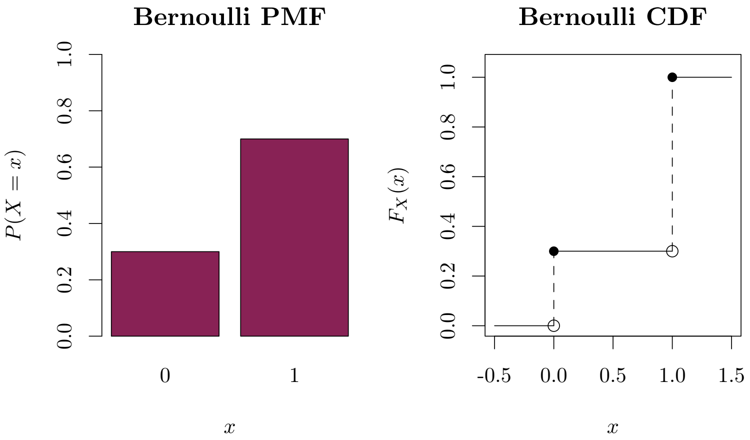


Figure 1: Probability Mass Function (left) and Cumulative Distribution Function (right) of a Bernoulli Distribution with $p = 0.7$.

Moments

Proposition 1. Moments of a Bernoulli Distribution

$$\mathbb{E}(X) = p$$
$$\mathrm{Var}(X) = p(1-p).$$

Proof.

- Expectation. By definition, for a discrete random variable,
$$\mathbb{E}(X) = \sum_x x\,\mathbb{P}(X = x) = 0 \cdot (1-p) + 1 \cdot p = p.$$
- Variance. Since $X \in \{0, 1\}$, we have $X^2 = X$. Thus,
$$\mathbb{E}(X^2) = \mathbb{E}(X) = p.$$
Hence,
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1-p).$$

Employment status

Let X denote employment status:

$$X = \begin{cases} 1 & \textit{if employed,} \\ 0 & \textit{otherwise.} \end{cases}$$

Suppose the employment probability is $p = 0.85$.

- $\mathbb{E}(X) = 0.85$: if we repeatedly select individuals at random and record their employment status, the average of the observed values of X will be close to 0.85.
- $\mathrm{Var}(X) = p(1-p) = 0.85 \times 0.15 = 0.1275$: this quantifies how much the observed values of X typically fluctuate around their average across repeated random selections of individuals.

1.2. Binomial Distribution

Motivation

- The **Binomial distribution** models the number of successes obtained when **repeating the same Bernoulli experiment** a fixed number of times.
- It is appropriate when:
  - the experiment is repeated a fixed number n of times,
  - each repetition has two possible outcomes (success / failure),
  - the probability of success p is constant,
  - repetitions are independent.
- The Binomial distribution is used to describe:
  - the number of employed individuals among n randomly selected workers,
  - the number of defaults in a portfolio of n identical loans,
  - the number of consumers who purchase a product among n contacted individuals,
  - the number of treated individuals among n randomly assigned units.

Formal Definition

Definition 4. Binomial Distribution

Let $X_1, \ldots, X_n$ be independent and identically distributed Bernoulli random variables with parameter p. The random variable

$$X = \sum_{i=1}^{n} X_i$$

*follows a **Binomial distribution** with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$.*

Notation

We write $X \sim \mathcal{B}(n, p)$.

Probability Mass Function

Definition 5. Probability Mass Function of a Binomial Distribution

Let $X \sim \mathcal{B}(n,p)$. *For* $k \in \{0, 1, \ldots, n\}$,

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

- $\binom{n}{k}$ counts the number of ways to obtain k successes among n trials.
- $p^k(1-p)^{n-k}$ is the probability of any given such configuration.

Remark

You recognize Newton's binomial formula (see Chapter 3)!

Visual Representation

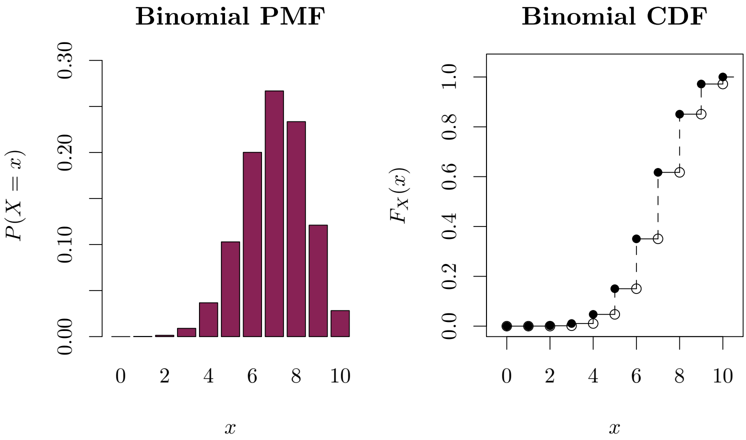


Figure 2: Probability Mass Function (left) and Cumulative Distribution Function (right) of a Binomial distribution with $n = 10$ and $p = 0.7$.

Moments

Proposition 2. Moments of a Binomial Distribution

Let $X \sim \mathcal{B}(n, p)$. *Then*

$$\mathbb{E}(X) = np,$$
$$\mathrm{Var}(X) = np(1-p).$$

Proof.

- Write $X = \sum_{i=1}^{n} X_i$ with $X_i \sim \mathcal{B}(p)$.
- By linearity of expectation,
$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np.$$
- Since the X_i are independent,
$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p).$$

Example

Employment in a sample of workers

Consider $n = 20$ *individuals selected independently at random from a population. For each individual, define*

$$X_i = \begin{cases} 1 & \textit{if employed,} \\ 0 & \textit{otherwise.} \end{cases}$$

Assume $\mathbb{P}(X_i = 1) = p = 0.85$, *and define* $X = \sum_{i=1}^{20} X_i$.

- $X \sim \mathcal{B}(20, 0.85)$.
- $\mathbb{E}(X) = 17$: across repeated samples of 20 individuals, the average number of employed individuals will be close to 17.
- $\mathrm{Var}(X) = 20 \times 0.85 \times 0.15 = 2.55$: this quantifies how much the number of employed individuals fluctuates around its average across repeated samples.

Sum of Two Independent Binomial Random Variables

Proposition 3. Stability under addition (same p)

Let $X \sim \mathcal{B}(n, p)$ and $Y \sim \mathcal{B}(n', p)$. If X and Y are independent, then

$$X + Y \sim \mathcal{B}(n + n', p).$$

Proof (1/2)

We compute the probability mass function of $X + Y$.

For $k \in \{0, 1, \ldots, n + n'\}$,

$$\mathbb{P}(X + Y = k) = \sum_{i=0}^{k} \mathbb{P}(X = i,\, Y = k - i).$$

Since X and Y are independent,

$$\mathbb{P}(X + Y = k) = \sum_{i=0}^{k} \mathbb{P}(X = i)\, \mathbb{P}(Y = k - i).$$

Using the Binomial PMF,

$$\mathbb{P}(X + Y = k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \binom{n'}{k-i} p^{k-i} (1-p)^{n'-(k-i)}.$$

Proof (1/2)

We factor out the terms that do not depend on i:

$$\mathbb{P}(X+Y=k)=p^k(1-p)^{(n+n')-k}\sum_{i=0}^{k}\binom{n}{i}\binom{n'}{k-i}.$$

By Vandermonde's identity (see Chapter 3),

$$\sum_{i=0}^{k}\binom{n}{i}\binom{n'}{k-i}=\binom{n+n'}{k}.$$

Therefore,

$$\mathbb{P}(X+Y=k)=\binom{n+n'}{k}p^k(1-p)^{(n+n')-k},$$

which is the PMF of a $\mathcal{B}(n+n',p)$ distribution. □

Exercise

A bank grants $n = 15$ identical loans. Each loan defaults independently with probability $p = 0.1$.

1. *Define a suitable random variable and specify its distribution.*
2. *Compute the probability that exactly two loans default.*
3. *Compute the expected number of defaults.*
4. *Compute the variance of the number of defaults.*
5. *Explain how the distribution changes when n increases while p remains fixed.*

Solution: Modelling

1. We model defaults as binary random variables

   - For **each loan** $i = 1, \ldots, 15$, define

$$X_i = \begin{cases} 1 & \text{if loan } i \text{ defaults,} \\ 0 & \text{otherwise.} \end{cases}$$

   - Each X_i follows a Bernoulli distribution with parameter $p = 0.1$.
   - The **total number of defaults** is

$$X = \sum_{i=1}^{15} X_i.$$

Since the loans default independently,

$$X \sim \mathcal{B}(15, 0.1).$$

Solution: $\mathbb{P}(X = 2)$

2 The probability that there are exactly k defaults is

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

Hence, for two defaults:

$$\begin{aligned}\mathbb{P}(X = 2) &= \binom{15}{2} (0.1)^2 (0.9)^{13} \\ &= 105 \times 0.01 \times (0.9)^{13} \\ &\approx 0.267.\end{aligned}$$

Solution: $\mathbb{E}(X)$, $\mathrm{Var}(X)$

3. For a Binomial random variable $X \sim \mathcal{B}(n, p)$,

$$\mathbb{E}(X) = np.$$

Here,

$$\mathbb{E}(X) = 15 \times 0.1 = 1.5.$$

Hence, if we repeatedly consider portfolios of 15 identical loans and record the number of defaults, the average number of observed defaults will be close to 1.5.

4. For a Binomial random variable $X \sim \mathcal{B}(n, p)$,

$$\mathrm{Var}(X) = np(1 - p).$$

Here,

$$\mathrm{Var}(X) = 15 \times 0.1 \times 0.9 = 1.35.$$

This variance quantifies how much the number of defaults fluctuates around its average across repeated portfolios.

Solution: Effect of Increasing n

Suppose the number of loans n increases while the default probability p remains fixed.

- The expected number of defaults increases linearly:

$$\mathbb{E}(X) = np.$$

- The variance also increases linearly:

$$\mathrm{Var}(X) = np(1-p).$$

- Relative fluctuations decrease: the standard deviation grows like $\sqrt{n}$, while the mean grows like n.

As a result, the distribution becomes more concentrated **relative to its mean**, and the Binomial distribution becomes increasingly symmetric and bell-shaped.

1.3. Poisson Distribution

Motivation

Siméon Denis Poisson (1781 – 1840), by François-Séraphin Delpech (Wikimedia).

- The **Poisson distribution** models the number of times an event occurs over a fixed interval of time or space.
- It is appropriate when:
  - events occur randomly and independently,
  - the probability of an event is small over a short interval,
  - events occur at a constant average rate.
- Examples:
  - the number of insurance claims filed in a day,
  - the number of customer arrivals at a service desk per hour,
  - the number of defaults in a large portfolio over a short period,
  - the number of accidents occurring on a road segment in a year.

- Randomness comes from **counting the number of events** occurring in a given interval.

Formal Definition

Definition 6. Poisson Distribution (Probability Mass Function)

*A random variable X follows a **Poisson distribution** with parameter $\lambda > 0$ if, for $k \in \mathbb{N}$,*

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

- λ represents the **average number of events** per interval.
- The support is infinite: X can take any nonnegative integer value.

Notation

We write $X \sim \mathcal{P}(\lambda)$.

Visual Representation

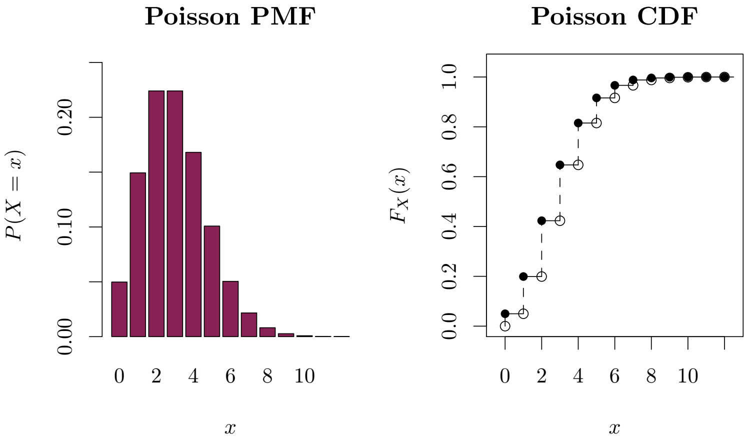


Figure 3: Probability Mass Function (left) and Cumulative Distribution Function (right) of a Poisson distribution with $\lambda = 3$.

Moments

Proposition 4. Moments of a Poisson Distribution

Let $X \sim \mathcal{P}(\lambda)$. *Then*

$$\mathbb{E}(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda.$$

Reminders.

- The exponential function admits the following **power series expansion**:

$$e^{\lambda} = \sum_{m=0}^{+\infty} \frac{\lambda^m}{m!} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots, \qquad \lambda \in \mathbb{R}.$$

- Equivalently,

$$e^{-\lambda} = \sum_{m=0}^{+\infty} \frac{(-\lambda)^m}{m!}.$$

Proof: Expectation of a Poisson Random Variable

Let $X \sim \mathcal{P}(\lambda)$ with $\lambda > 0$. By definition,

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{+\infty} k\,\mathbb{P}(X = k) \\
&= \sum_{k=0}^{+\infty} k\,e^{-\lambda}\,\frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} k\,\frac{\lambda^k}{k!} \quad \text{(term is } 0 \text{ at } k = 0\text{)} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \quad \text{(use } k/k! = 1/(k-1)!\text{)} \\
&= e^{-\lambda}\lambda \sum_{m=0}^{+\infty} \frac{\lambda^m}{m!} \quad \text{(set } m = k-1\text{)} \\
&= e^{-\lambda}\lambda e^{\lambda} \quad \text{(using the power series expansion)} \\
&= \lambda
\end{aligned}$$

Proof. Variance of a Poisson Random Variable

$$
\begin{aligned}
\mathbb{E}(X^2) &= e^{-\lambda} \sum_{n=0}^{+\infty} n^2 \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n=0}^{+\infty} [n(n-1)+n] \frac{\lambda^n}{n!} \quad (\text{using } n^2 = n(n-1)+n) \\
&= e^{-\lambda} \textstyle\sum_{n=0}^{+\infty} n(n-1) \frac{\lambda^n}{n!} + e^{-\lambda} \textstyle\sum_{n=0}^{+\infty} n \frac{\lambda^n}{n!}
\end{aligned}
$$

First term: since $\dfrac{n(n-1)}{n!} = \dfrac{1}{(n-2)!}$,

$$
e^{-\lambda} \sum_{n=0}^{+\infty} n(n-1) \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=2}^{+\infty} \frac{\lambda^n}{(n-2)!} = e^{-\lambda} \lambda^2 \sum_{m=0}^{+\infty} \frac{\lambda^m}{m!} = \lambda^2.
$$

Variance of a Poisson Random Variable

Second term: since $\frac{n}{n!} = \frac{1}{(n-1)!}$,

$$e^{-\lambda} \sum_{n=0}^{+\infty} n \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^n}{(n-1)!} = e^{-\lambda} \lambda \sum_{m=0}^{+\infty} \frac{\lambda^m}{m!} = \lambda.$$

Therefore,

$$\mathbb{E}(X^2) = \lambda^2 + \lambda.$$

Hence,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

$\square$

Sum of Two Independent Poisson Random Variables

Proposition 5. Stability under addition

Let $X \sim \mathcal{P}(\lambda)$ and $Y \sim \mathcal{P}(\lambda')$. If X and Y are independent, then

$$X + Y \sim \mathcal{P}(\lambda + \lambda').$$

Proof (1/2) (Homework)

Let $k \in \mathbb{N}$. By independence,

$$
\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{i=0}^{k} \mathbb{P}(X=i,\, Y=k-i) \\
&= \sum_{i=0}^{k} \mathbb{P}(X=i)\mathbb{P}(Y=k-i) \\
&= \sum_{i=0}^{k} \left(\frac{\lambda^i}{i!}e^{-\lambda}\right)\left(\frac{(\lambda')^{k-i}}{(k-i)!}e^{-\lambda'}\right) \quad \text{(Poisson PMF)} \\
&= e^{-(\lambda+\lambda')} \sum_{i=0}^{k} \frac{\lambda^i (\lambda')^{k-i}}{i!(k-i)!} \\
& e^{-(\lambda+\lambda')} \frac{1}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda^i (\lambda')^{k-i} \quad \left(\text{use } \frac{1}{i!(k-i)!} = \frac{1}{k!}\binom{k}{i}\right)
\end{aligned}
$$

Proof (2/2)

By the binomial formula,

$$\sum_{i=0}^{k} \binom{k}{i} \lambda^i (\lambda')^{k-i} = (\lambda + \lambda')^k.$$

Therefore,

$$\mathbb{P}(X + Y = k) = \frac{(\lambda + \lambda')^k}{k!} e^{-(\lambda+\lambda')},$$

which is the PMF of $\mathcal{P}(\lambda + \lambda')$. $\square$

Example

Insurance claims

Suppose that insurance claims arrive at an average rate of $\lambda = 2$ per day. Let X denote the number of claims observed in one day.

- $X \sim \mathcal{P}(2)$.
- $\mathbb{E}(X) = 2$: if we repeatedly observe the number of claims per day, the average count will be close to 2.
- $\mathrm{Var}(X) = 2$: the number of claims fluctuates around its average with variance equal to 2.

Exercise

Calls arrive at a call center at an average rate of $\lambda = 5$ per hour.

1. *Define a suitable random variable and specify its distribution.*
2. *Compute the probability that exactly three calls arrive in one hour.*
3. *Compute the probability that at most two calls arrive in one hour.*
4. *Compute the expectation and variance.*
5. *Explain how the distribution changes when λ increases.*

Solution: Modelling, $P(X = 3)$

1. Let X be the number of calls received in one hour. Since calls arrive at an average rate of $\lambda = 5$ per hour, we model

$$X \sim \mathcal{P}(5).$$

2. For a Poisson random variable $X \sim \mathcal{P}(\lambda)$,

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Here, $\lambda = 5$ and $k = 3$, so

$$\mathbb{P}(X = 3) = \frac{5^3}{3!} e^{-5} = \frac{125}{6} e^{-5} \approx 0.1404.$$

Solution: At least two calls, $\mathbb{E}(X)$, $\mathrm{Var}(X)$, effect of Increasing λ

3 The probability that at most two calls arrive in one hour is

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2) \\ &= e^{-5}\left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!}\right) = e^{-5}\left(1 + 5 + \frac{25}{2}\right) \approx 0.1247.\end{aligned}$$

4 $\mathbb{E}(X) = \mathrm{Var}(X) = \lambda = 5$: if we repeatedly observe one-hour periods, the average number of calls per hour will be close to 5, and the counts will fluctuate around that average with variance 5.

5 When λ increases:

- the distribution shifts to the right since $\mathbb{E}(X) = \lambda$ increases.
- the distribution becomes more spread out in absolute terms since $\mathrm{Var}(X) = \lambda$ increases.
- the PMF becomes more symmetric (less right-skewed) as λ grows.

1.4. Geometric Distribution

Motivation

- The **Geometric distribution** models the **waiting time until the first success** in a **sequence of independent Bernoulli trials**.
- It is appropriate when:
  - each trial has two outcomes (success / failure),
  - trials are independent,
  - the probability of success is constant across trials.
- Typical economic examples:
  - the number of job applications sent until receiving the first offer,
  - the number of calls made until a customer answers,
  - the number of inspections until a firm is found in violation,
  - the number of periods until a consumer makes a purchase.

Formal Definition

Definition 7. Geometric Distribution

Consider a sequence of independent Bernoulli trials with success probability $p \in (0,1)$. Let X be the number of trials required to obtain the first success. Then X follows a ***Geometric distribution*** *with parameter p if, for $k = 1, 2, \dots,$*

$$\mathbb{P}(X = k) = (1-p)^{k-1}p.$$

Notation

We write $X \sim \mathcal{G}(p)$.

Probability Mass Function

Definition 8. Probability Mass Function of a Geometric Distribution

Let $X \sim \mathcal{G}(p)$. For $k = 1, 2, \ldots,$

$$\mathbb{P}(X = k) = \underbrace{(1-p) \cdot (1-p) \cdot (1-p)}_{\text{failue } k-1 \text{ times}} \cdot \underbrace{p}_{1 \text{ success at the end}} = (1-p)^{k-1} p.$$

- The PMF decreases geometrically with k.
- Larger values of p lead to shorter waiting times on average.

Visual Representation

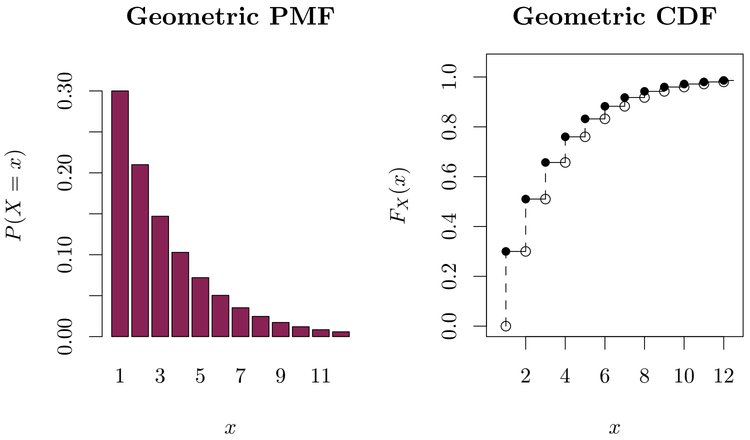


Figure 4: Probability Mass Function (left) and Cumulative Distribution Function (right) of a Geometric distribution with $p = 0.3$.

Moments

Proposition 6. Moments of a Geometric Distribution

Let $X \sim \mathcal{G}(p)$. Then

$$\mathbb{E}(X) = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Proof of the Expectation (1/2)

By definition,

$$\mathbb{E}(X) = \sum_{k \in \mathbb{N}^*} k\mathbb{P}(X = k) = \sum_{k=1}^{+\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{+\infty} k(1-p)^{k-1}.$$

Let $q = 1 - p$, then,

$$\mathbb{E}(X) = p \boxed{\sum_{k=1}^{+\infty} k\, q^{k-1}}.$$

To compute this sum, we start from the geometric series. Recall that for $|q| < 1$, the following converges:

$$\boxed{\sum_{k=0}^{+\infty} q^k} = \boxed{\frac{1}{1-q}}.$$

Proof of the Expectation (2/2)

Since $|q| < 1$, the series defines a smooth function of q, and we may differentiate term by term:

$$\frac{d}{dq}\sum_{k=0}^{+\infty} q^k = \sum_{k=0}^{+\infty} \frac{d}{dq} q^k = \sum_{k=0}^{+\infty} k\, q^{k-1} = \sum_{k=1}^{+\infty} k\, q^{k-1}$$

Since

$$\frac{d}{dq}\sum_{k=0}^{+\infty} q^k = \frac{d}{dq}\left(\frac{1}{1-q}\right) = -\frac{-1}{(1-q)^2} = \frac{1}{(1-q)^2},$$

Then,

$$\sum_{k=1}^{+\infty} k\, q^{k-1} = \frac{1}{(1-q)^2}.$$

Hence,

$$\mathbb{E}(X) = p\sum_{k=1}^{+\infty} k\, q^{k-1} = p\frac{1}{(1-q)^2} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

Proof of the Variance (1/3)

We begin with computing

$$\mathbb{E}(X^2) = \sum_{k=1}^{+\infty} k^2\, \mathbb{P}(X = k) = p \sum_{k=1}^{+\infty} k^2 q^{k-1}.$$

Use the identity $k^2 = k(k-1) + k$:

$$\sum_{k=1}^{+\infty} k^2 q^{k-1} = \sum_{k=1}^{+\infty} k(k-1) q^{k-1} + \underbrace{\sum_{k=1}^{+\infty} k q^{k-1}}_{=\frac{1}{(1-q)^2} \text{ for } |q|<1}.$$

We still need to compute $\sum_{k=1}^{+\infty} k(k-1) q^{k-1}$.

Proof of the Variance (2/3)

Again, for $|q| < 1$, we have (known result)

$$\sum_{k=0}^{+\infty} q^k = \frac{1}{1-q}.$$

If we differentiate twice term by term:

$$\frac{d^2}{dq^2}\left(\sum_{k=0}^{+\infty} q^k\right) = \sum_{k=2}^{+\infty} k(k-1)q^{k-2}, \qquad \text{and} \qquad \frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) = \frac{2}{(1-q)^3}.$$

Hence,

$$\sum_{k=2}^{+\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3}.$$

Proof of the Variance (3/3)

Multiplying both sides by q gives

$$\sum_{k=2}^{+\infty} k(k-1)q^{k-1} = q\,\frac{2}{(1-q)^3}.$$

Therefore,

$$\mathbb{E}(X^2) = p\left(\frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2}\right) = p\left(\frac{2q}{p^3} + \frac{1}{p^2}\right) = \frac{2q}{p^2} + \frac{1}{p}.$$

Since $q = 1-p$, this can be written as

$$\mathbb{E}(X^2) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}.$$

Finally, using $\mathbb{E}(X) = \frac{1}{p}$,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Example

Job applications

Suppose each job application leads to an offer with probability $p = 0.2$, independently of previous applications. Let X be the number of applications sent until the first offer is received.

- $X \sim \mathcal{G}(0.2)$.
- $\mathbb{E}(X) = 5$: if we repeatedly go through the application process, the average number of applications until the first offer will be close to 5.
- $\text{Var}(X) = \frac{1-0.2}{0.2^2} = 20$: the waiting time fluctuates substantially around its average.

Exercise

Each inspection on a firm detects a violation with probability $p = 0.1$, independently of previous inspections.

1. *Define a suitable random variable and specify its distribution.*
2. *Compute the probability that the first violation is detected on the third inspection.*
3. *Compute the probability that at most three inspections are needed.*
4. *Compute the expectation and variance.*
5. *Explain how the distribution changes when p increases.*

Solution: Modelling, $\mathbb{P}(X = 3)$

1. Each inspection is a Bernoulli trial with success probability $p = 0.1$ ("success" is defined as detecting a violation). Let X be the number of inspections needed until the first violation is detected. Therefore,

$$X \sim \mathcal{G}(0.1), \qquad \mathbb{P}(X = k) = (1 - 0.1)^{k-1}\, 0.1, \quad k = 1, 2, \ldots$$

2. The probability that the first violation is detected on the third inspection is

$$\mathbb{P}(X = 3) = (1 - p)^{3-1} p = (0.9)^2 \times 0.1 = 0.081$$

Solution: $\mathbb{P}(X \leq 3)$

3. The probability that at most three inspections are needed writes

$$\mathbb{P}(X \leq 3) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3).$$

Equivalently, it is easier to use the complement:

$$\mathbb{P}(X \leq 3) = 1 - \mathbb{P}(X > 3).$$

$\mathbb{P}(X > 3)$ means "no detection in the first 3 inspections", i.e. three failures:

$$\mathbb{P}(X > 3) = (1 - p)^3 = (0.9)^3.$$

Thus,

$$\mathbb{P}(X \leq 3) = 1 - (0.9)^3 = 1 - 0.729 = 0.271.$$

Solution: $\mathbb{E}(X)$, $\mathrm{Var}(X)$

4. The expectation and variance are

$$\mathbb{E}(X) = \frac{1}{p} = 10, \qquad \mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{0.9}{0.1^2} = 90.$$

This means that across repeated inspection processes, the average number of inspections required until the first detection will be close to 10, with substantial variability.

5. When p increases:
   - the waiting time decreases on average since $\mathbb{E}(X) = 1/p$ decreases,
   - the distribution puts more mass on small values (e.g. $X = 1, 2$),
   - the variance decreases since $\mathrm{Var}(X) = (1-p)/p^2$ decreases.

1.5. Hypergeometric Distribution

Motivation

- The **Hypergeometric distribution** models the number of successes obtained when sampling **without replacement** from a finite population.
- It is appropriate when:
  - the sample size is fixed,
  - each draw results in success or failure,
  - draws are **not independent**.
- For example:
  - the number of defective products in a quality-control sample,
  - the number of employed individuals in a small surveyed group,
  - the number of audited firms found non-compliant,
  - the number of subsidized households in a randomly selected village.

Formal Definition

Definition 9. Hypergeometric Distribution and its Probability Mass Function

Consider a finite population of size N, containing K successes and $N-K$ failures. A sample of size n is drawn uniformly at random ***without replacement****.*
Let X denote the number of successes in the sample. Then X follows a ***Hypergeometric distribution****. For integers k such that*

$$0 \leq k \leq K, \quad \text{and} \quad 0 \leq n-k \leq N-K, \quad \text{and} \quad \mathbb{P}(X=k)=0 \text{ otherwise},$$

the probability mass function is

$$\mathbb{P}(X=k) = \frac{\overbrace{\binom{K}{k}\binom{N-K}{n-k}}^{\text{favourable samples}}}{\underbrace{\binom{N}{n}}_{\text{possible samples of size } n}}.$$

Remarks

Notation

We write

$$X \sim \textit{HGeom}(N, K, n).$$

Convention

We adopt the convention

$$\binom{a}{b} = 0 \quad \textit{if } b > a.$$

This allows the probability mass function of the Hypergeometric distribution to be written without explicitly restricting the support.
With this convention, the formula defines a probability distribution, although we do not verify here the properties of non-negativity and unit total mass.

Visual Representation

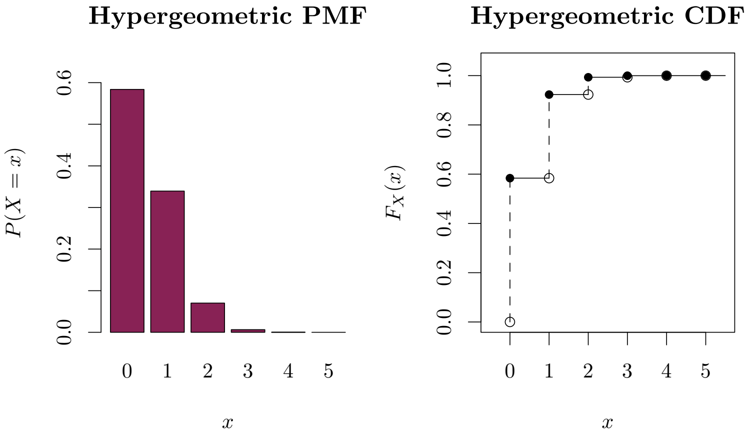


Figure 5: Probability Mass Function (left) and Cumulative Distribution Function (right) of a Hypergeometric distribution HGeom(N, K, n) with $N = 100$, $K = 10$, and $n = 5$.

Moments

Proposition 7. Moments of a Hypergeometric Distribution (Admitted)

Let $X \sim$ HGeom(N, K, n). Then

$$\mathbb{E}(X) = n\frac{K}{N}, \qquad \mathrm{Var}(X) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}.$$

Example

Quality control

A factory produces $N = 100$ items, of which $K = 10$ are defective. A quality inspector randomly selects $n = 5$ items without replacement.
Let X be the number of defective items in the sample.

- $X \sim \mathcal{H}(100, 10, 5)$.
- $\mathbb{E}(X) = 5 \times \frac{10}{100} = 0.5$.
- On average, half a defective item is found per inspection sample.

Exercise

A population consists of 50 individuals, among whom 20 are employed. A survey randomly selects 10 individuals without replacement. We are interested in modelling the number of employed individuals among the 10 surveyed.

1. *Define a suitable random variable and specify its distribution.*
2. *Compute the probability that exactly 4 employed individuals are selected.*
3. *Compute the expectation and variance.*
4. *Compare this model to a Binomial model with $p = K/N$.*

Solution: Modelling, $\mathbb{P}(X = 4)$

1. The population has size $N = 50$, with $K = 20$ employed and $N - K = 30$ not employed. A sample of size $n = 10$ is drawn uniformly at random *without replacement*. Let X be the number of employed individuals in the sample. Therefore,

$$X \sim \text{HGeom}(50, 20, 10).$$

2. For $X \sim \mathcal{H}(N, K, n)$,

$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Here, $N = 50$, $K = 20$, $n = 10$, and $k = 4$, so

$$\mathbb{P}(X = 4) = \frac{\binom{20}{4}\binom{30}{6}}{\binom{50}{10}} = \frac{4845 \cdot 593,775}{10,272,278,170} = \frac{92,491}{330,256} \approx 0.28.$$

Solution: Expectation and Variance

3. For an hypergeometric distribution,

$$\mathbb{E}(X) = n\frac{K}{N}, \qquad \mathrm{Var}(X) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}.$$

Here, $\frac{K}{N} = \frac{20}{50} = 0.4$, so

$$\mathbb{E}(X) = 10 \times 0.4 = 4.$$

In addition,

$$\mathrm{Var}(X) = 10 \times 0.4 \times 0.6 \times \frac{50-10}{50-1} = 2.4 \times \frac{40}{49} = \frac{96}{49} \approx 1.959.$$

Solution: Comparison with the Binomial Model

4. If we (incorrectly) assumed sampling *with replacement*, we would model

$$X \sim \mathcal{B}(n, p), \qquad p = \frac{K}{N} = 0.4,$$

so that

$$\mathbb{E}(X) = np = 10 \times 0.4 = 4, \qquad \mathrm{Var}(X) = np(1-p) = 10 \times 0.4 \times 0.6 = 2.4.$$

Sampling without replacement induces negative dependence between draws, which reduces the variance:

$$\mathrm{Var}_{\text{hyper}}(X) = \mathrm{Var}_{\text{binom}}(X) \times \frac{N-n}{N-1}, \qquad \frac{N-n}{N-1} = \frac{40}{49} < 1.$$

Thus, the Hypergeometric model has the same expectation but a smaller variance than the Binomial approximation.

1.6. Discrete Uniform Distribution

Motivation

- The **Discrete Uniform distribution** models a situation in which a finite number of outcomes are **equally likely**.
- It is appropriate when:
  - the set of possible outcomes is finite,
  - no outcome is favoured over another.
- For example:
  - the outcome of a fair die,
  - the result of a lottery draw with equally likely numbers,
  - the index of a randomly selected individual in a fixed list,
  - a random assignment among a finite number of identical options.

Formal Definition

Definition 10. Discrete Uniform Distribution and its Probability Mass Function

Let $a < b$ be two integers, and consider the finite set

$$\mathcal{D}_X = \{a, a+1, \dots, b\}.$$

A random variable X follows a ***Discrete Uniform distribution*** *on $\mathcal{D}_X$ if*

$$\mathbb{P}(X = k) = \frac{1}{b - a + 1} \quad \text{for all } k \in \mathcal{D}_X, \qquad \mathbb{P}(X = k) = 0 \text{ otherwise.}$$

Notation

We write

$$X \sim \mathcal{U}\{a, b\}.$$

Visual Representation

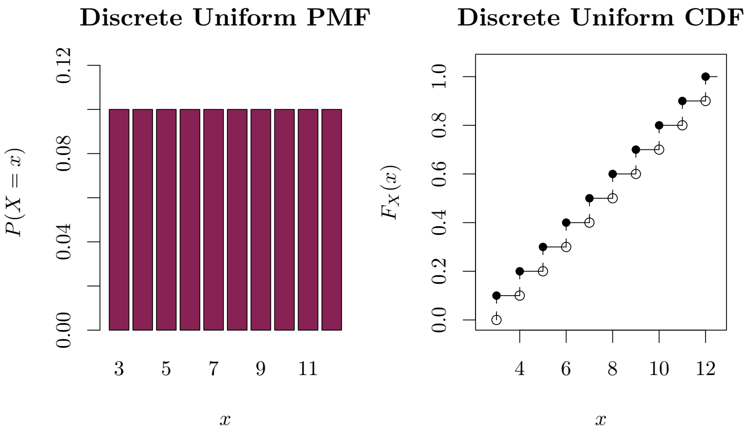


Figure 6: Probability Mass Function (left) and Cumulative Distribution Function (right) of a Discrete Uniform distribution on $\{a, \ldots, b\}$, with $a = 3$, $b = 12$.

Moments

Proposition 8. Moments of a Discrete Uniform Distribution

Let $X \sim \mathcal{U}\{a, , b\}$. *Then*

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

Proof of the Expectation

Let $X \sim \mathcal{U}\{a, , b\}$, where $a < b$ are integers. Then

$$\mathbb{P}(X = k) = \frac{1}{b - a + 1}, \qquad k = a, a + 1, \dots, b.$$

By definition,

$$\mathbb{E}(X) = \sum_{k=a}^{b} k\,\mathbb{P}(X = k) = \frac{1}{b - a + 1} \sum_{k=a}^{b} k.$$

The sum of consecutive integers is

$$\sum_{k=a}^{b} k = \frac{(a + b)(b - a + 1)}{2}.$$

Therefore,

$$\mathbb{E}(X) = \frac{1}{b - a + 1} \cdot \frac{(a + b)(b - a + 1)}{2} = \frac{a + b}{2}.$$

Proof of the Variance (1/2)

Let $X \sim \mathcal{U}\{a, b\}$ and set $n = b - a + 1$ (so the support has n values).

Write $X = a + Y$, where $Y \sim \mathcal{U}\{0, n-1\}$. Since adding a constant does not change variance,

$$\mathrm{Var}(X) = \mathrm{Var}(Y).$$

We compute $\mathbb{E}(Y)$ and $\mathbb{E}(Y^2)$:

$$\mathbb{E}(Y) = \frac{1}{n}\sum_{j=0}^{n-1} j = \frac{n-1}{2}, \qquad \mathbb{E}(Y^2) = \frac{1}{n}\sum_{j=0}^{n-1} j^2.$$

Using the identity

$$\sum_{j=0}^{n} j^2 = \frac{(n)(n+1)(2n+1)}{6},$$

we get

$$\mathbb{E}(Y^2) = \frac{1}{n} \cdot \frac{(n-1)n(2n-1)}{6} = \frac{(n-1)(2n-1)}{6}.$$

Proof of the Variance (2/2)

Therefore,

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{(n-1)(2n-1)}{6} - \left(\frac{n-1}{2}\right)^2 = \frac{n^2-1}{12}.$$

Since $n = b - a + 1$, we conclude

$$\mathrm{Var}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

$\square$

Example: Fair die

Let X denote the outcome of a fair six-sided die. Then

$$X \sim \mathcal{U}\{1,6\}.$$

- $\mathbb{E}(X) = \frac{1+6}{2} = 3.5.$
- The outcomes are symmetrically distributed around the mean.

Exercise

A random integer is drawn uniformly from the set $\{3, 4, \ldots, 12\}$.

1. *Define a suitable random variable and specify its distribution.*
2. *Compute the probability that the number drawn is equal to* 7.
3. *Compute the expectation and variance.*
4. *Compute the probability that the number drawn is greater than or equal to* 10.

Solution: Modelling, $\mathbb{P}(X = 7)$

1. Let X denote the random integer drawn from $\{3, 4, \ldots, 12\}$. Since all values are equally likely,
$$X \sim \mathcal{U}\{3, \ldots, 12\}.$$

2. The probability mass function is
$$\mathbb{P}(X = k) = \frac{1}{12 - 3 + 1} = \frac{1}{10} \quad \text{for } k \in \{3, \ldots, 12\}.$$
In particular,
$$\mathbb{P}(X = 7) = \frac{1}{10}.$$

Solution: Expectation and Variance, $\mathbb{P}(X \geq 10)$

3. For a Discrete Uniform distribution on $\{a, \ldots, b\}$,

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

Here, $a = 3$ and $b = 12$, so

$$\mathbb{E}(X) = \frac{3+12}{2} = 7.5, \qquad \mathrm{Var}(X) = \frac{10^2 - 1}{12} = \frac{99}{12}.$$

4. The event $\{X \geq 10\}$ corresponds to the values $\{10, 11, 12\}$. Since all outcomes are equally likely,

$$\mathbb{P}(X \geq 10) = \frac{3}{10}.$$

2. Continuous Probability Laws

2.1. Continuous Uniform Distribution

Motivation

- The **Continuous Uniform distribution** models a situation in which all values in a **continuous interval** are equally likely.
- It is appropriate when:
  - the set of possible outcomes is an interval of real numbers,
  - no subinterval is favoured over another.
- For example:
  - the arrival time of a customer within a fixed time window,
  - a random point chosen uniformly on a segment,
  - a starting time chosen uniformly between two hours.

Formal Definition

Definition 11. Continuous Uniform Distribution and its Density

Let $a < b$ be real numbers. A random variable X follows a ***Continuous Uniform distribution*** *on $[a, b]$ if its probability density function is*

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \textit{otherwise.} \end{cases}$$

Notation

We write

$$X \sim \mathcal{U}(a, b).$$

Distribution Function

Proposition 9. Cumulative Distribution Function of a Uniform Distribution

The cumulative distribution function of $X \sim \mathcal{U}(a, b)$ is given by

$$F_X(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases}$$

- F_X is linear on $[a, b]$,
- probabilities correspond to **lengths of subintervals**.

Proof.

Case 1. For $x < a$, since $f_X(x) = 0$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = 0.$$

Case 2. For $a \leq x \leq b$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{a}^{x} f_X(t)\,dt = \int_{a}^{x} \frac{1}{b-a}\,dt = \left[\frac{1}{b-a}t\right]_a^x = \frac{x-a}{b-a}.$$

Case 3. For $x > b$, the entire support $[a, b]$ is included in $(-\infty, x]$, hence

$$F_X(x) = \int_{a}^{b} f_X(t)\,dt = 1.$$

$\square$

Visual Representation

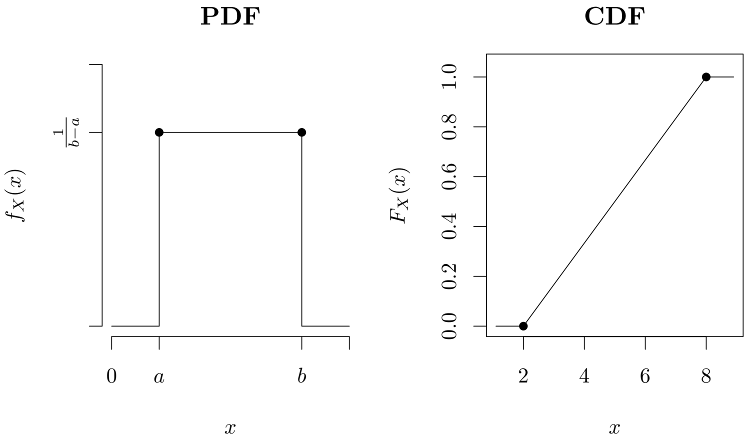
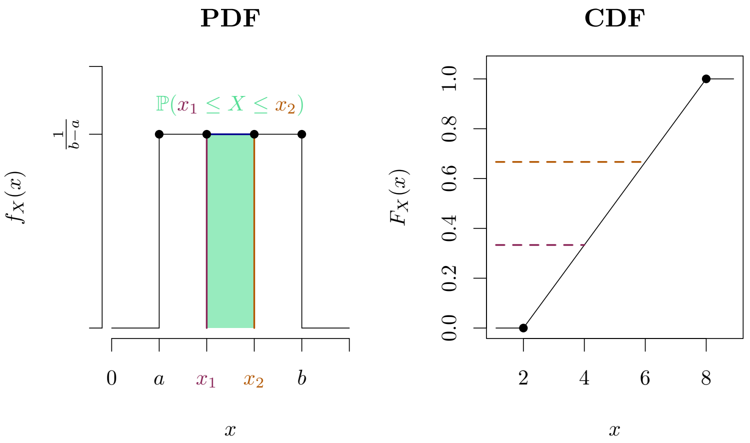


Figure 7: Probability Density Function (left) and Cumulative Distribution Function (right) of a Continuous Uniform distribution on $[a, b]$, with $a = 2$ and $b = 8$.

Geometric Reasoning

Note that the PDF is a rectangle. Hence, computing $\mathbb{P}(x_1 \leq X \leq x_2)$ *is easy!*



Moments

Proposition 10. Moments of a Continuous Uniform Distribution

Let $X \sim \mathcal{U}(a, b)$. *Then*

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Proof of the Expectation

By definition,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx.$$

We compute

$$\int_a^b x\, dx = \left[\frac{x^2}{2}\right]_a^b = \frac{b^2 - a^2}{2}.$$

Therefore,

$$\mathbb{E}(X) = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{1}{b-a} \cdot \frac{(b-a)(b+a)}{2} = \frac{a+b}{2}.$$

Proof of the Variance

Using $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, we first compute

$$\mathbb{E}(X^2) = \int_a^b x^2 f_X(x)\, dx = \frac{1}{b-a}\int_a^b x^2\, dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$$
$$= \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)}.$$

Therefore,

$$\text{Var}(X) = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$
$$= \frac{4a^2+4ab+4b^2-3(a^2+b^2+2ab)}{12}$$
$$= \frac{(a-b)^2}{12}.$$

Example: Uniform arrival time

Suppose a customer arrives uniformly between 9:00 and 10:00. Let X denote the arrival time (in hours).

Then

$$X \sim \mathcal{U}(9, 10).$$

- $\mathbb{E}(X) = 9.5$,
- any n-minute interval (e.g., 10-minute) has the same probability.

Exercise

A point is chosen uniformly at random on the interval $[2, 8]$.

1. *Define a suitable random variable and specify its distribution.*
2. *Compute* $\mathbb{P}(X \leq 5)$.
3. *Compute the expectation and variance.*
4. *Compute* $\mathbb{P}(4 \leq X \leq 6)$.

Solution: Modelling, $\mathbb{P}(X \leq 5)$

1. Let X denote the random point on $[2, 8]$. Then

$$X \sim \mathcal{U}(2, 8).$$

2. Since probabilities correspond to interval lengths,

$$F_X(x) = \mathbb{P}(X \leq x) = \frac{x - a}{b - a}.$$

With $a = 2$ and $b = 8$ and $x = 5$,

$$\mathbb{P}(X \leq 5) = \frac{5 - 2}{8 - 2} = \frac{3}{6} = \frac{1}{2}.$$

Solution: Moments, $\mathbb{P}(4 \leq X \leq 6)$

3 For $X \sim \mathcal{U}(a, b)$,

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Here,

$$\mathbb{E}(X) = 5, \qquad \mathrm{Var}(X) = \frac{36}{12} = 3.$$

4

$$\mathbb{P}(4 \leq X \leq 6) = \frac{6-4}{8-2} = \frac{2}{6} = \frac{1}{3}.$$

2.2. Gaussian (Normal) Distribution

Motivation

Carl Friedrich Gauß(1777–1855), by Christian Albrecht Jensen (Wikimedia).

- The **Gaussian (Normal) distribution** is the most widely used continuous probability law.
- It naturally arises when:
  - many small, independent effects contribute additively,
  - variability is symmetric around a typical value,
  - extreme deviations are unlikely.
- For example:
  - measurement errors,
  - biological characteristics (e.g., height),
  - financial returns over short horizons,
  - aggregated economic shocks.

Formal Definition

Definition 12. Gaussian Distribution and its Density

Let $\mu \in \mathbb{R}$ and $\sigma > 0$. A random variable X follows a ***Gaussian distribution*** *with mean μ and variance σ^2 if its probability density function is*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$

Notation

We write

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

- We will admit that $\int_{\mathbb{R}} f_X(x)\, dx = 1$.

Visual Representation

- The density is **symmetric** around μ, where the max is reached and equals $\frac{1}{\sigma\sqrt{2\pi}}$.
- The parameter σ controls the **spread** of the distribution.
- Larger σ implies more dispersed outcomes.

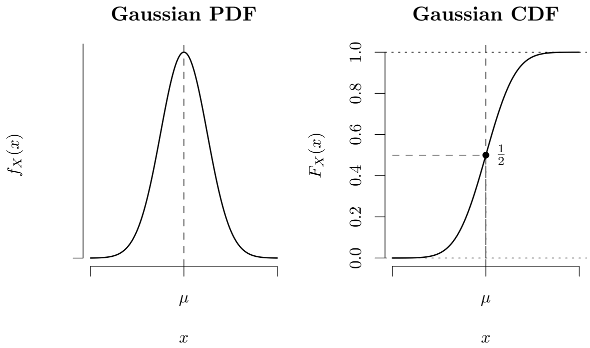


Figure 8: Probability density function of a Gaussian distribution with mean μ and variance σ^2.

First Derivative of the Gaussian Density

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Its probability density function is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Differentiating with respect to x,

$$f'_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \left(-\frac{x-\mu}{\sigma^2}\right) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = -\frac{x-\mu}{\sigma^3\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = -\frac{x-\mu}{\sigma^2} f_X(x).$$

Hence:

- $f'_X(x) = 0 \iff x = \mu$,
- f_X is increasing on $(-\infty, \mu)$,
- f_X is decreasing on $(\mu, +\infty)$.

Second Derivative of the Gaussian Density (1/2)

Starting from

$$f'_X(x) = -\frac{x-\mu}{\sigma^2}\, f_X(x) = -\frac{1}{\sigma^2} x f_X(x) + \frac{\mu}{\sigma^2} f_X(x),$$

we differentiate once more:

$$\begin{aligned} f''_X(x) &= -\frac{1}{\sigma^2}\left(1 \times f_X(x) + x\, f'_X(x)\right) + \frac{\mu}{\sigma^2} f'_X(x) \\ &= -\frac{1}{\sigma^2} f_X(x) + f'_X(x)\left(\frac{\mu}{\sigma^2} - \frac{x}{\sigma^2}\right) \\ &= -\frac{1}{\sigma^2} f_X(x) - \frac{x-\mu}{\sigma^2} f'_X(x). \end{aligned}$$

Second Derivative of the Gaussian Density (2/2)

Substituting the expression of $f'(x)$,

$$
\begin{aligned}
f''(x) &= -\frac{1}{\sigma^2} f_X(x) - \frac{x-\mu}{\sigma^2} f_X'(x) \\
&= -\frac{1}{\sigma^2} f_X(x) - \frac{x-\mu}{\sigma^2} \left(-\frac{x-\mu}{\sigma^2} f_X(x) \right) \\
&= -\frac{1}{\sigma^2} f_X(x) + \frac{(x-\mu)}{\sigma^2} \frac{(x-\mu)}{\sigma^2} f_X(x) \\
&= \left(\frac{(x-\mu)^2 - \sigma^2}{\sigma^4} \right) f_X(x).
\end{aligned}
$$

Hence:

- $f_X''(x) = 0 \iff |x - \mu| = \sigma$,
- **inflection points** at $\mu - \sigma$ and $\mu + \sigma$ (more details on the next slide),
- $f_X''(\mu) < 0$, confirming a unique maximum at μ.

Inflection Points and the Role of σ

From

$$f''_X(x) = \left(\frac{(x-\mu)^2 - \sigma^2}{\sigma^4}\right) f_X(x),$$

note that $f_X(x) > 0$ for all x, so **the sign of** $f''_X(x)$ is the sign of $(x-\mu)^2 - \sigma^2$.

- If $|x-\mu| < \sigma$, then $(x-\mu)^2 - \sigma^2 < 0$, hence $f''_X(x) < 0$: the density is **concave** (curving downward) around the peak.
- If $|x-\mu| > \sigma$, then $(x-\mu)^2 - \sigma^2 > 0$, hence $f''_X(x) > 0$: the density becomes **convex** (curving upward) in the tails.
- Therefore, the curve changes concavity exactly at $x = \mu - \sigma$ and $x = \mu + \sigma$: these are the **inflection points**.
- The **distance** from the **center** μ to each **inflection point** is σ.
- Increasing σ pushes the inflection points outward, making the bell curve **wider**: this helps to understand why σ **measures dispersion**.

Visual Representation

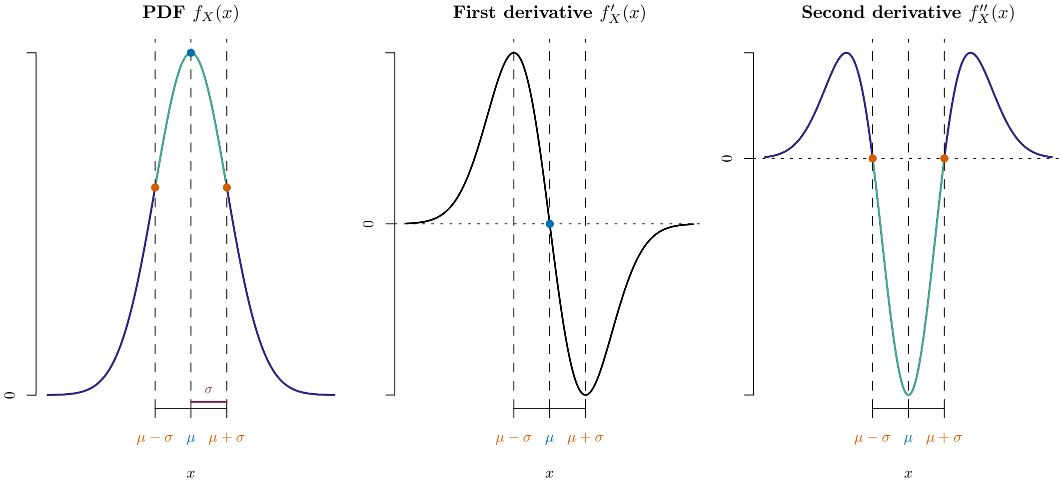


Figure 9: Gaussian Density and Inflection Points.

Standard Gaussian Distribution

Definition 13. Standard Normal Distribution

*A random variable Z follows a **standard Gaussian distribution** if*

$$Z \sim \mathcal{N}(0, 1).$$

Its density is

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

- Mean 0, variance 1, symmetric around 0.
- This distribution serves as a **reference distribution** (see the statistics course with Alain Paraponaris).

Notation

- *The CDF (F_Z) of a standard Gaussian distribution is usually denoted Φ.*
- *Its density function is usually denoted φ.*

Moments

Proposition 11. Moments of a Gaussian Distribution

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. *Then*

$$\mathbb{E}(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2.$$

Proof of the Expectation (1/2)

First, note that the existence of moments of any order is guaranteed by the fact that the function

$$x \longmapsto x^k e^{-x^2/2}$$

is integrable over $\mathbb{R}$ for all $k \in \mathbb{N}$ (you saw that with Pierre Michel).

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. By definition of the expectation,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Proof of the Expectation (2/2)

We perform the change of variables

$$y = \frac{x-\mu}{\sigma}, \qquad \text{so that } x = \sigma y + \mu, \quad dx = \sigma\, dy.$$

Hence,

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma y + \mu)\, e^{-y^2/2}\, dy.$$

We decompose the integral:

$$\mathbb{E}(X) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y\, e^{-y^2/2}\, dy + \mu \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2}\, dy}_{=1 \text{ (density of a } \mathcal{N}(0,1))}.$$

The first integral is zero since the integrand is an odd function.

Therefore, $\mathbb{E}(X) = \mu$.

Proof of the Variance (1/3)

Recall that
$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Since $\mathbb{E}(X) = \mu$, it remains to compute $\mathbb{E}(X^2)$:
$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)\, dx = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} x^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Again, use the change of variable $x = \mu + \sigma y$ (i.e., $y = (x-\mu)/\sigma$), so that $dx = \sigma\, dy$ and
$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\mu + \sigma y)^2 e^{-y^2/2}\, dy.$$

Expand:
$$(\mu + \sigma y)^2 = \mu^2 + 2\mu\sigma y + \sigma^2 y^2.$$

Thus
$$\mathbb{E}(X^2) = \frac{1}{\sqrt{2\pi}} \left[\mu^2 \int_{-\infty}^{+\infty} e^{-y^2/2}\, dy + 2\mu\sigma \int_{-\infty}^{+\infty} y e^{-y^2/2}\, dy + \sigma^2 \int_{-\infty}^{+\infty} y^2 e^{-y^2/2}\, dy\right].$$

Proof of the Variance (2/3)

Developing:

$$\mathbb{E}(X^2) = \mu^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2}\, dy + 2\mu\sigma \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} y e^{-y^2/2}\, dy + \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-y^2/2}\, dy.$$

We use:

- $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2}\, dy = 1$ (density),
- $\int_{-\infty}^{+\infty} y e^{-y^2/2}\, dy = 0$ (odd integrand),

For the **third term**, let $Z \sim \mathcal{N}(0,1)$. We have:

$$\mathbb{E}(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 f_Z(z)\, dz.$$

Proof of the Variance (3/3)

Let us define $I = \int_{-\infty}^{+\infty} z^2 f_Z(z)\, dz$, and

$$u(z) = z, \quad u'(z) = 1, \qquad v'(z) = ze^{-z^2/2}, \quad v(x) - e^{-z^2/2}.$$

Using integration by parts, we obtain

$$I = \int_{-\infty}^{+\infty} z^2 f_Z(z)\, dz = \left[-ze^{-z^2/2}\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-z^2/2} dz = 0 - 0 + \sqrt{2\pi} = \sqrt{2\pi}.$$

Therefore, the **third term** of $\mathbb{E}(X^2)$ is

$$\sigma^2 \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2$$

Hence,

$$\mathbb{E}(X^2) = \mu^2 + \sigma^2.$$

Finally,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mu^2 = \sigma^2.$$

□

Standardization

Proposition 12. Standardization

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ *and define*

$$Z = \frac{X - \mu}{\sigma}.$$

Then

$$Z \sim \mathcal{N}(0, 1).$$

- Allows probability computations using standard normal tables,
- reduces all Gaussian problems to the same reference case.

Proof (1/2)

The density of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

For $z \in \mathbb{R}$,

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(X \leq \mu + \sigma z) = F_X(\mu + \sigma z).$$

Since X admits a density f_X and F_X is differentiable, we may differentiate with respect to z (by the chain rule):

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_X(\mu + \sigma z) = \sigma\, f_X(\mu + \sigma z)$$

Proof (2/2)

Substituting the Gaussian density of X,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

we obtain

$$\begin{aligned} f_Z(z) &= \sigma \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mu+\sigma z-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \end{aligned}$$

This is the density of a $\mathcal{N}(0,1)$ distribution. Therefore,

$$Z \sim \mathcal{N}(0,1).$$

Sum of Two Independent Gaussian Variables

Proposition 13. Sum of Two Independent Gaussian Variables

Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be ***independent****.*
Then

1. $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$.
2. $\mathbb{E}(X_1 + X_2) = \mu_1 + \mu_2$,
3. $\mathrm{Var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2$,

- We will admit the first proposition. It can be proven using moment generating functions.

Proof for the Expectation and the Variance

Since expectation is linear,

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = \mu_1 + \mu_2.$$

For the variance, recall that for any two random variables,

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\ \mathrm{Cov}(X_1, X_2).$$

Since X_1 and X_2 are independent, we have

$$\mathrm{Cov}(X_1, X_2) = 0.$$

Therefore,

$$\mathrm{Var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2.$$

□

Example

Let $X_1 \sim \mathcal{N}(2, 1)$ and $X_2 \sim \mathcal{N}(1, 4)$ be independent.

- $\mathbb{E}(X_1 + X_2) = 2 + 1 = 3$,
- $\text{Var}(X_1 + X_2) = 1 + 4 = 5$,
- $X_1 + X_2 \sim \mathcal{N}(3, 5)$.

Exercise

Let X_1 and X_2 be two independent random variables such that

$$X_1 \sim \mathcal{N}(10, 4), \qquad X_2 \sim \mathcal{N}(5, 1).$$

1. *Compute* $\mathbb{E}(X_1 + X_2)$ *and* $\mathrm{Var}(X_1 + X_2)$.
2. *Determine the distribution of* $Y = X_1 + X_2$.
3. *Compute* $\mathbb{P}(X_1 \leq 12)$.
4. *Compute* $\mathbb{P}(X_1 + X_2 \geq 20)$.

Solution: Moments and Distribution

Let $X_1 \sim \mathcal{N}(10, 4)$ and $X_2 \sim \mathcal{N}(5, 1)$ be independent, and define

$$Y = X_1 + X_2.$$

1. By linearity of expectation,

$$\mathbb{E}(Y) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 10 + 5 = 15.$$

2. Since X_1 and X_2 are independent,

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 4 + 1 = 5.$$

3. Since the sum of two independent Gaussian variables is Gaussian,

$$Y = X_1 + X_2 \sim \mathcal{N}(15, 5).$$

Solution: Probability Computations

3. Compute $\mathbb{P}(X_1 \leq 12)$.

   Since $X_1 \sim \mathcal{N}(10, 4)$, we have $\sigma_1 = 2$ and

$$\mathbb{P}(X_1 \leq 12) = \mathbb{P}\left(\frac{X_1 - 10}{2} \leq \frac{12 - 10}{2}\right) = \mathbb{P}(Z \leq 1) = \Phi(1) \approx 0.8413.$$

4. Compute $\mathbb{P}(X_1 + X_2 \geq 20)$.

   From the previous slide, $Y = X_1 + X_2 \sim \mathcal{N}(15, 5)$, so $\sigma_Y = \sqrt{5}$:

$$\mathbb{P}(Y \geq 20) = \mathbb{P}\left(\frac{Y - 15}{\sqrt{5}} \geq \frac{20 - 15}{\sqrt{5}}\right) = \mathbb{P}\left(Z \geq \sqrt{5}\right) = 1 - \Phi(\sqrt{5}) \approx 0.0127.$$

2.3. Exponential Distribution

Motivation

- The **Exponential distribution** models the **waiting time until the occurrence of an event**.
- It is appropriate when:
  - events occur randomly in time,
  - the probability of occurrence does not depend on how long one has already waited.
- For example:
  - the time until failure of a device,
  - the waiting time before a customer arrives,
  - the time between two phone calls,
  - the duration until the next accident.

Formal Definition

Definition 14. Exponential Distribution and its Density

*Let $\lambda > 0$. A random variable X follows an **Exponential distribution** with parameter λ if its probability density function is*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Notation

We write

$$X \sim \mathit{Exp}(\lambda).$$

Distribution Function

Definition 15. Cumulative Distribution Function of an Exponential Distribution

The cumulative distribution function of $X \sim Exp(\lambda)$ is

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

- F_X is increasing and concave,
- the tail probability is

$$\mathbb{P}(X > x) = e^{-\lambda x}.$$

Visual Representation

- The density is **decreasing** on $[0, +\infty)$,
- larger values of λ correspond to **shorter waiting times**.

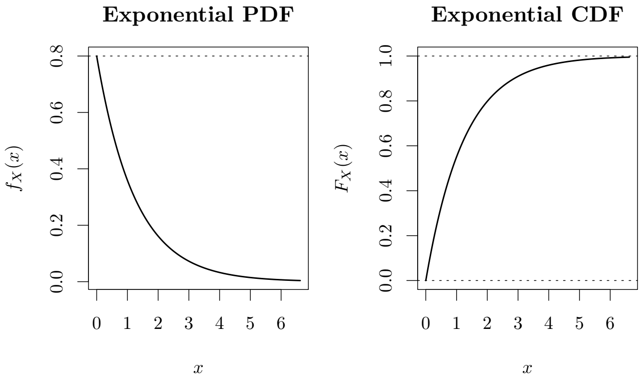


Figure 10: Probability density function (left) and cumulative distribution function (right) of an Exponential distribution $X \sim \text{Exp}(\lambda)$, with $\lambda = .8$.

Moments

Proposition 14. Moments of an Exponential Distribution

Let $X \sim$ Exp(λ). Then

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Proof of the Expectation

By definition,

$$\mathbb{E}(X) = \int_0^{+\infty} x\,\lambda e^{-\lambda x}\,dx.$$

Using integration by parts with

$$u(x) = x, \qquad u'(x) = 1, \qquad v'(x) = \lambda e^{-\lambda x}, \qquad v(x) = -e^{-\lambda x}$$

we obtain

$$\begin{aligned}
\mathbb{E}(X) &= [u(x)v(x)]_0^{+\infty} - \int_0^{+\infty} u'(x)v(x)\,dx \\
&= \left[-x\,e^{-\lambda\,x}\right]_0^{+\infty} - \int_0^{+\infty} -e^{-\lambda\,x}\,dx \\
&= \left[-\frac{1}{\lambda}e^{-\lambda\,x}\right]_0^{+\infty} \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Proof of the Variance (1/2)

Use the definition:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

By definition,

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 \, \lambda \, e^{-\lambda \, x} \, dx.$$

Using a first integration by parts with

$$u(x) = x^2, \qquad u'(x) = 2x, \qquad v'(x) = \lambda \, e^{-\lambda \, x}, \qquad v(x) = -e^{-\lambda \, x}$$

we obtain:

$$\mathbb{E}(X^2) = \left[-x^2 \, e^{-\lambda \, x}\right]_0^{+\infty} + \int_0^{+\infty} 2x \, e^{-\lambda \, x} \, dx = 2 \, \int_0^{+\infty} x \, e^{-\lambda \, x}.$$

Proof of the Variance (2/2)

Let $I = \int_0^{+\infty} x\, e^{-\lambda\, x}$. Using a second integration by parts, with

$$u(x) = x, \qquad u'(x) = 1, \qquad v'(x) = e^{-\lambda\, x}, \qquad v(x) = -\frac{1}{\lambda} e^{-\lambda\, x}$$

we obtain:

$$I = \left[-\frac{x}{\lambda} e^{-\lambda\, x} \right]_0^{+\infty} + \int_0^{+\infty} \frac{1}{\lambda} e^{-\lambda\, x}\, dx = \frac{1}{\lambda} \left[-\frac{1}{\lambda} e^{-\lambda\, x} \right]_0^{+\infty} = \frac{1}{\lambda}(0 - (-\frac{1}{\lambda})) = \frac{1}{\lambda^2}.$$

Hence,

$$\mathbb{E}(X^2) = 2\, I = \frac{2}{\lambda^2}.$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

$\square$

Memoryless Property

Proposition 15. Memorylessness

Let $X \sim$ *Exp*(λ). *Then, for all* $s, t \geq 0$,

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t).$$

- The future waiting time does not depend on the past,
- the Exponential distribution is the **only continuous distribution** with this property.

Proof. For $s, t \geq 0$,

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t).$$

□

Minimum of Two Independent Exponential Variables

Proposition 16. Minimum of Two Independent Exponentially Distributed Random Variables

Let $X_1 \sim Exp(\lambda_1)$ and $X_2 \sim Exp(\lambda_2)$ be ***independent****. Then*

$$\min(X_1, X_2) \sim \mathcal{E}(\lambda_1 + \lambda_2).$$

Proof

Let $Y = \min(X_1, X_2)$. For any $t \geq 0$,

$$\mathbb{P}(Y > t) = \mathbb{P}(\min(X_1, X_2) > t) = \mathbb{P}(X_1 > t \text{ and } X_2 > t).$$

Since X_1 and X_2 are independent,

$$\begin{aligned}\mathbb{P}(Y > t) &= \mathbb{P}(X_1 > t)\mathbb{P}(X_2 > t)\\ &= e^{-\lambda_1 t} e^{-\lambda_2 t}\\ &= e^{-(\lambda_1+\lambda_2)t}.\end{aligned}$$

Thus,

$$F_Y(y) = \mathbb{P}(Y \leq y) = 1 - \mathbb{P}(Y > y) = 1 - e^{-(\lambda_1+\lambda_2)t},$$

which corresponds to the CDF of an Exponentially distributed random variable with parameter $\lambda_1 + \lambda_2$. Therefore,

$$\min(X_1, X_2) \sim \text{Exp}(\lambda_1 + \lambda_2).$$

Example

Let X denote the lifetime (in years) of a component, and suppose

$$X \sim \mathcal{E}(0.5).$$

- $\mathbb{E}(X) = 2$ years,
- $\mathbb{P}(X > 3) = e^{-1.5} \approx 0.223$,
- the probability of surviving one more year does not depend on the current age.

Exercise

Let $X \sim \mathcal{E}(\lambda)$.

1. *Compute* $\mathbb{P}(X > 2/\lambda)$.
2. *Compute* $\mathbb{E}(X)$ *and* $\mathrm{Var}(X)$.

Solution

1. Compute $\mathbb{P}(X > 2/\lambda)$.

   The tail probability of the Exponential distribution is

   $$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = e^{-\lambda x},$$

   we obtain

   $$\mathbb{P}\left(X > \frac{2}{\lambda}\right) = e^{-\lambda \cdot \frac{2}{\lambda}} = e^{-2}.$$

2. Compute $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

   From the general formulas for the Exponential distribution,

   $$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

2.4. Gamma Distribution

Motivation

- The **Gamma distribution** generalizes the Exponential distribution.
- It models the **waiting time until the k-th event** occurs.
- It is appropriate when:
  - events occur randomly in time,
  - we are interested in the accumulation of several waiting times.
- For example:
  - the time until the k-th failure of a system,
  - the time until the k-th customer arrives,
  - aggregated service or repair times,
  - rainfall amounts over a fixed period.

Formal Definition

Definition 16. Gamma Distribution and its Density

Let $k > 0$ (shape parameter) and $\lambda > 0$ (rate parameter). A random variable X follows a ***Gamma distribution*** *with parameters (k, λ) if its probability density function is*

$$f_X(x) = \begin{cases} \dfrac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The function Γ is defined, for all $k > 0$, by

$$\Gamma(k) = \int_0^{+\infty} x^{k-1} e^{-x} \, dx.$$

It generalizes the factorial in the sense that

$$\Gamma(k) = (k-1)! \quad \textit{for all } k \in \mathbb{N}.$$

Special Cases

Notation

We write

$$X \sim \Gamma(k, \lambda).$$

- If $k = 1$, then
$$\Gamma(1, \lambda) = \text{Exp}(\lambda).$$
- If $k \in \mathbb{N}$, the distribution is called an **Erlang distribution**.
- If $\lambda = \frac{1}{2}$ and $k = \frac{j}{2}$ with $j \in \mathbb{N}$, then
$$X \sim \Gamma\left(\frac{j}{2}, \frac{1}{2}\right) \quad \Longleftrightarrow \quad X \sim \chi^2(j),$$
the **chi-square distribution** with j degrees of freedom (see the statistics course with Alain Paraponaris).
- For large k, the Gamma distribution becomes approximately Gaussian (by the Central Limit Theorem that will be presented in Chapter 8).

$f_X(x)$ Is a Proper Density Function

Proposition 17. Positivity and Normalization of the Gamma Density

Let $k > 0$ and $\lambda > 0$. The function

$$f_X(x) = \begin{cases} \dfrac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

is a probability density function.

Proof. First, we show the **non-negativity** of the density function.

For $x \geq 0$, each factor λ^k, x^{k-1}, and $e^{-\lambda x}$ is non-negative, and $\Gamma(k) > 0$ for all $k > 0$. Hence $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.

Proof (continued)

Second, we show the **normalization** condition.

Since $f_X(x) = 0$ for $x < 0$,

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = \int_0^{+\infty} \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}\,dx.$$

Use the change of variable $u = \lambda x$ (so $x = u/\lambda$ and $dx = du/\lambda$):

$$\begin{aligned}
\int_0^{+\infty} \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}\,dx &= \frac{\lambda^k}{\Gamma(k)} \int_0^{+\infty} \left(\frac{u}{\lambda}\right)^{k-1} e^{-u}\,\frac{du}{\lambda} \\
&= \frac{1}{\Gamma(k)} \underbrace{\int_0^{+\infty} u^{k-1} e^{-u}\,du}_{=\Gamma(k)} \\
&= 1.
\end{aligned}$$

Moments

Proposition 18. Moments of a Gamma Distribution

Let $X \sim \Gamma(k, \lambda)$. *Then*

$$\mathbb{E}(X) = \frac{k}{\lambda}, \qquad \mathrm{Var}(X) = \frac{k}{\lambda^2}.$$

- The parameter k controls the **shape**:
  - small k: highly skewed distribution,
  - large k: more symmetric.
- The parameter λ controls the **time scale**:
  - large λ: shorter waiting times,
  - small λ: longer waiting times.

Proof of the Expectation

By definition,

$$\mathbb{E}(X) = \int_0^{+\infty} x\, f_X(x)\, dx = \int_0^{+\infty} x\, \frac{\lambda^k}{\Gamma(k)}\, x^{k-1} e^{-\lambda x}\, dx = \frac{\lambda^k}{\Gamma(k)} \int_0^{+\infty} e^{-\lambda x} x^k\, dx.$$

Using an integration by parts with

$$u(x) = x^k, \qquad u'(x) = k\, x^{k-1} \qquad v'(x) = e^{-\lambda x} \qquad v(x) = -\frac{1}{\lambda} e^{-\lambda x}$$

we obtain:

$$\begin{aligned}
\mathbb{E}(X) &= \frac{\lambda^k}{\Gamma(k)} \left(\left[e^{-\lambda x} x^k \right]_0^{+\infty} + \int_0^{+\infty} k \frac{x^{k-1}}{\lambda} e^{-\lambda x}\, dx \right) \\
&= \frac{k}{\lambda} \int_0^{+\infty} \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}\, dx \\
&= \frac{k}{\lambda}.
\end{aligned}$$

Proof of the Variance (1/2)

First, consider

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}\, dx = \frac{\lambda^k}{\Gamma(k)} \int_0^{+\infty} x^{k+1} e^{-\lambda x}\, dx$$

Using a first integration by parts with

$$u(x) = x^{k+1}, \qquad u'(x) = (k+1)x^k, \qquad v'(x) = e^{-\lambda x}, \qquad v(x) = -\frac{1}{\lambda} e^{-\lambda x}$$

we obtain:

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{\lambda^k}{\Gamma(k)} \left(\left[-\frac{x^{k+1}}{\lambda} e^{-\lambda x} \right]_0^{+\infty} + \int_0^{+\infty} \frac{k+1}{\lambda} x^k e^{-\lambda x}\, dx \right) \\ &= \frac{\lambda^k}{\Gamma(k)} \cdot \frac{k+1}{\lambda} \int_0^{+\infty} x^k e^{-\lambda x}\, dx. \end{aligned}$$

Proof of the Variance (2/2)

Using a second integration by parts with

$$u(x) = x^k, \qquad u'(x) = kx^{k-1}, \qquad v'(x) = e^{-\lambda x}, \qquad v(x) = -\frac{1}{\lambda}e^{-\lambda x}$$

we obtain:

$$\begin{aligned}\mathbb{E}(X^2) &= \frac{\lambda^k}{\Gamma(k)} \cdot \frac{k+1}{\lambda} \left(\left[-\frac{x^k}{\lambda} e^{-\lambda x} \right]_0^{+\infty} + \int_0^{+\infty} \frac{k}{\lambda} x^{k-1} e^{-\lambda x} \, dx \right) \\ &= \frac{k+1}{\lambda} \frac{k}{\lambda} \underbrace{\int_0^{+\infty} \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \, dx}_{= \int_{\mathcal{D}_X} f_X(x)\, dx = 1} \\ &= \frac{k^2 + k}{\lambda^2}.\end{aligned}$$

Therefore:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{k^2+k}{\lambda^2} - \left(\frac{k}{\lambda}\right)^2 = \frac{k}{\lambda^2}.$$

Sum of Independent Exponential Variables

Proposition 19. Gamma Distribution as a Sum

Let $X_1, \ldots, X_k$ be independent random variables such that

$$X_i \sim \mathit{Exp}(\lambda), \quad i = 1, \ldots, k.$$

Then

$$X_1 + \cdots + X_k \sim \Gamma(k, \lambda).$$

Sketch of the proof.

- The result follows from repeated convolution of Exponential densities,
- or from the additivity of waiting times in a Poisson process,
- or using moment generating functions.
- With this result, we better understand why the Gamma distribution naturally appears as a model for waiting times until the k-th event.

Example

Suppose events occur at rate $\lambda = 2$ per hour. Let X be the time until the third event occurs.

Then

$$X \sim \Gamma(3, 2),$$

with

$$\mathbb{E}(X) = \frac{3}{2}, \qquad \mathrm{Var}(X) = \frac{3}{4}.$$

Sum of Independent Gamma Random Variables

Proposition 20. Additivity of the Gamma Distribution

Let $X_1 \sim \Gamma(k_1, \lambda)$ and $X_2 \sim \Gamma(k_2, \lambda)$ be ***independent*** *Gamma random variables with the same rate parameter λ.*
Then

$$X_1 + X_2 \sim \Gamma(k_1 + k_2, \lambda).$$

Remark

The proof of this result relies on convolution of densities or on moment generating functions, and is omitted here.

References I

Hansen, B. (2022). *Probability and statistics for economists*. Princeton University Press.