Probability

Probability laws

Benchikh Tawfik

Faculty of Medicine
1st Year Medicine

November 23, 2025

Plan de cours

1. Continuous Probability Laws: The Normal Distribution
2. Discrete Probability Distributions
3. Exercises

Normal Distribution: Key Facts

- The normal distribution is a **theoretical** distribution, an ideal mathematical model that never appears exactly in nature.
- However, many real datasets closely resemble this well-known bell-shaped curve: many observations lie near the mean, with fewer as we move away from it, symmetrically.
- It is fundamental in inferential statistics: the sample mean is a random variable that tends to follow a normal distribution as the sample size increases, even when the underlying population distribution is not normal.

Central Limit Theorem (1)

- **Central Limit Theorem (CLT).** The sum of a large number of random variables, regardless of their individual distributions, tends to follow an approximate normal distribution. Formally,

$$X = X_1 + X_2 + \cdots + X_n$$

  approaches a normal law as $n \longrightarrow \infty$.
- This explains why so many real-world distributions look bell-shaped: they describe phenomena resulting from the **addition** of a **large number** of **independent** sources of variation.
  - **Example:** Human height or weight can be viewed as the sum of many independent factors (genetics, environment, nutrition, climate, etc.).

Central Limit Theorem (2)

- Many distributions can be well approximated by a normal distribution whenever the variable can be written as the sum of many independent components.
- This is notably true for the **binomial distribution** (sum of n independent Bernoulli variables), whose shape becomes bell-like as n grows.
- The approximation remains valid even when the exact distribution of each X_i is unknown.

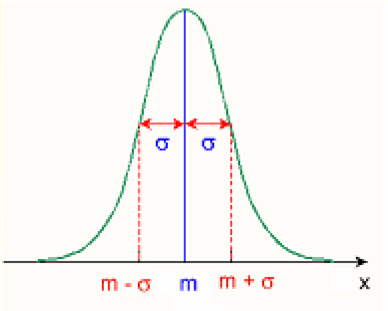
Standard Normal Distribution: Definition

- The most important continuous probability law.
- Density:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

- Expected value: $\mathbb{E}(X) = 0$.
- Variance: $\mathrm{Var}(X) = 1$.
- Standard deviation: $\sigma = 1$.

Shape of the $\mathcal{N}(0,1)$ Distribution



- Bell-shaped density curve
- Total area under the curve = 1
- Symmetric distribution
- Axis of symmetry = mean $m = 0$
- Maximum at the mean
- Standard deviation = distance between mean and inflection point

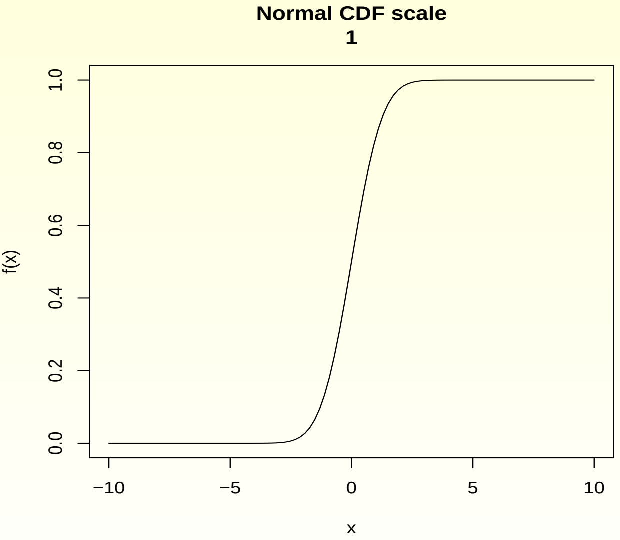
Cumulative Distribution Function

- The cumulative distribution function (CDF) plays a fundamental role.
- Recall: the CDF F is defined by:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

- Graphically, it corresponds to the area shaded in yellow under the curve.

Cumulative Distribution Function

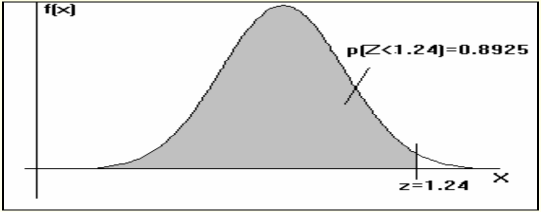


Probability Calculations for the Normal Distribution: Normal Table

- **A major drawback**: we do not know an antiderivative of the function e^{-x^2}. In other words, the cumulative distribution function $F(x)$ has no closed-form algebraic expression.
- Using numerical integration methods, one can compute the area under the curve for different values of x. This led to the creation of **normal tables** providing the values of $F(x)$, which allow us to determine the probabilities of events of interest.
- In practice, all probability calculations are reduced to the cumulative distribution function of the standard normal law $\mathcal{N}(0, 1)$, known as the **standard normal distribution**.

TABLE DE LA LOI NORMALE CENTREE

Lecture de la table: Pour z=1.24 (intersection de la ligne 1.2 et de la
on a la proportion P(Z < 1,24) = 0.8925



P(Z > 1,
P(Z > 2,
P(Z > 3,

Rappels:

1/ P(Z > z) = 1 - P(Z < z) et 2/ P(Z < -z) = P(Z > z)
Exemple: Sachant P(Z < 1,24) = 0,8925, on en déduit:
1/ (P(Z > 1,24) = 1 - P(Z < 1,24) = 1- 0,8925 = 0,1075
2/ P(Z < -1,24) = P(Z > 1,24) = 0,1075

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	[illegible]	[illegible]	[illegible]

Probability Calculations for the Normal Distribution: Normal Table

- The table provides, for various values of $x > 0$, the corresponding values of $F(x)$, i.e., $\mathbf{Pr}(X \leq x)$.
- For example, the table gives:
  - $F(0) = 0.5000$
  - $F(1) = 0.8413$
  - $F(0.5) = 0.6915$
  - $F(1.96) = 0.9750$
- By symmetry of the standard normal distribution:
  - $F(-1) = 1 - F(1) = 1 - 0.8413 = 0.1587$
  - $F(-0.5) = 1 - F(0.5) = 1 - 0.6915 = 0.3085$
  - $F(-1.96) = 1 - F(1.96) = 1 - 0.9750 = 0.0250$

Example: Computing Quantiles Using the Table

- Let $X \sim \mathcal{N}(0, 1)$.
- We want to find $\mathbf{Pr}(X \leq x)$ for a given quantile x (or conversely, find x for a given probability).
- The standard normal table gives the values of $F(x) = \mathbf{Pr}(X \leq x)$ for $x \geq 0$.

Example: Computing Quantiles Using the Table

- **What is $\Pr(X \leq 1.91)$?**
  - Split 1.91 into its first decimal and second decimal: $1.91 = 1.9 + 0.01$.
  - Look at the **row** corresponding to 1.9 and the **column** corresponding to 0.01. The value at their **intersection** gives the probability: **0.9719**.

Example: Computing Quantiles Using the Table

- **What is $\Pr(X \leq -1.91)$?**
  - The table does not directly list **negative quantiles**.
  - By symmetry of the normal density,

$$\Pr(X \leq -1.91) = \Pr(X \geq 1.91).$$

  - But the table only gives $\Pr(X \leq x)$, not $\Pr(X \geq x)$.
  - We use:

$$\Pr(X \geq 1.91) = 1 - \Pr(X \leq 1.91).$$

  Thus,

$$\begin{aligned}\Pr(X \leq -1.91) &= \Pr(X \geq 1.91) \\ &= 1 - 0.9719 = 0.0281.\end{aligned}$$

Using the Table

- **What is $\Pr(0.5 < X \leq 1.01)$?**
  - The event $(0.5 < X \leq 1.01)$ can be written as the difference between two cumulative probabilities:

$$\begin{aligned} \Pr(0.5 < X \leq 1.01) &= \Pr(X \leq 1.01) - \Pr(X \leq 0.5) \\ &= 0.8438 - 0.6915 = 0.1523. \end{aligned}$$

Using the Table

- **Find x such that $\mathbf{Pr}(X \leq x) = 0.8315$.**
  - Look for the value 0.8315 **inside the table**.
  - Read the corresponding row and column, and **add** them:

$$x = 0.9 + 0.06 = 0.96.$$

Using the Table

- **Find x such that $\Pr(X \leq x) = 0.2358$.**
  - ◇ This value does **not** appear inside the table. Since $0.2358 < 0.5$, the corresponding quantile must be **negative**.
  - ◇ Using symmetry:

$$\Pr(X \leq x) = 0.2358 = 1 - \Pr(X \leq -x).$$

Thus,

$$\Pr(X \leq -x) = 1 - 0.2358 = 0.7642.$$

From the table, $-x = 0.72$, so

$$x = -0.72.$$

General Normal Distribution

- Let μ be a real number, $\sigma > 0$, and let X be a standard normal random variable $X \sim \mathcal{N}(0, 1)$.
- The random variable

$$Z = \mu + \sigma X$$

obtained from X by a shift of origin and a change of scale has density, at any point z, given by

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

- We then say that Z follows a normal distribution with mean μ and variance σ^2, written

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

Expectation and Variance

- We have already seen how a shift of origin and a change of scale affect the expectation and variance:
  - ◇ $\mathbb{E}(Z) = \mathbb{E}(\mu + \sigma X) = \mu$
  - ◇ $\mathrm{Var}(Z) = \sigma^2$

Cumulative Distribution Function

- The probability that Z is less than or equal to a real number c is
$$F(c) = \int_{-\infty}^{c} f(z)\, dz.$$
- Thus, the probability that Z falls inside the interval $I = (a, b]$ is
$$\mathbf{Pr}(a < Z \leq b) = \int_{a}^{b} f(z)\, dz = F(b) - F(a).$$

Standardized Variable

Definition

The standardized variable associated with a random variable X with mean $\mathbb{E}(X)$ and variance $\sigma^2(X)$ is

$$X' = \frac{X - \mathbb{E}(X)}{\sigma(X)}.$$

- One checks that $\mathbb{E}(X') = 0$ and $\mathrm{Var}(X') = 1$.

Normal Distribution $\mathcal{N}(\mu, \sigma^2)$ and Probabilities

- Let $Z \sim \mathcal{N}(\mu, \sigma^2)$. We want to compute:

$$\mathbf{Pr}(a \leq Z \leq b) = ?$$

◇ Only the standard normal distribution $\mathcal{N}(0, 1)$ is tabulated.

◇ But if we standardize,

$$X = \frac{Z - \mu}{\sigma} \sim \mathcal{N}(0, 1),$$

we can compute probabilities for any normal distribution.

▷ Indeed,

$$\mathbf{Pr}(Z \leq z) = \mathbf{Pr}(\mu + \sigma X \leq z) = \mathbf{Pr}\left(X \leq \frac{z - \mu}{\sigma}\right) = F\left(\frac{z - \mu}{\sigma}\right).$$

◇ Therefore, knowing the CDF of the standard normal distribution is enough to compute any probability involving a normal variable $\mathcal{N}(\mu, \sigma^2)$.

Example: Computing Probabilities

- Let $Z \sim \mathcal{N}(20, 36)$.
- **What is $\mathbf{Pr}(Z \leq 22.4)$?**
  - ◇ We write

    $$\mathrm{Pr}(Z \leq 22.4) = \mathrm{Pr}\left(X \leq \frac{22.4 - 20}{6}\right),$$

    where $X \sim \mathcal{N}(0, 1)$.
  - ◇ Hence

    $$\mathrm{Pr}(Z \leq 22.4) = \mathrm{Pr}(X \leq 0.4) = 0.6554.$$

Example: Computing Quantiles

- Let $Z \sim \mathcal{N}(400, 15.4^2)$.
- Find the value of z such that $\mathbf{Pr}(Z > z) = 0.67$.
  - ◇ We have

$$\mathbf{Pr}(Z > z) \quad = \quad \mathbf{Pr}\left(\tfrac{Z-400}{15.4} > \tfrac{z-400}{15.4}\right) = \mathbf{Pr}(X > t) = 0.67 = F(0.44),$$

    so

$$\left|\frac{z-400}{15.4}\right| = |t| = 0.44.$$

  - ◇ Since $\mathbf{Pr}(X > t) = 0.67 > 0.5$, we must have $t < 0$. Thus $t = -0.44$, and

$$z = 400 + t \cdot 15.4 = 393.$$

Sum of Independent Normal Variables

- If X and Y are independent normal random variables with

$$X \sim \mathcal{N}(\mu_1, \sigma_1^2), \qquad Y \sim \mathcal{N}(\mu_2, \sigma_2^2),$$

then

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

- This result generalizes to the sum of any number of independent normal variables.
- Similarly,

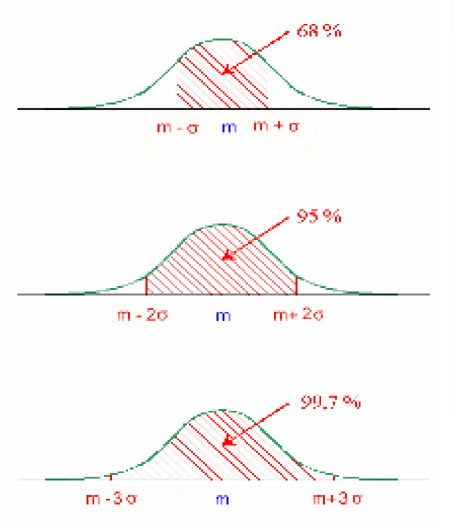
$$X - Y \sim \mathcal{N}(\mu_1 - \mu_2,\ \sigma_1^2 + \sigma_2^2).$$

Independent Normal Variables: Example

The height X of Algerian men is modeled by a normal distribution $\mathcal{N}(172, 196)$ (in cm).

1. What proportion of Algerian men are shorter than 160 cm?
2. What proportion measure more than 2 meters?
3. What proportion are between 165 and 185 cm?
4. If 10,000 Algerian men were ranked by height, what would be the height of the 9000th?
5. The height Y of Algerian women is modeled by $\mathcal{N}(162, 144)$. What is the probability that a randomly selected man is taller than a randomly selected woman?

Useful Rules of Thumb



- A normal variable has "95 chances out of 100" of lying between mean minus 2 standard deviations and mean plus 2 standard deviations (the exact value is 1.96).
- A normal variable is almost certainly between mean $\pm$ 3 standard deviations.

Discrete distributions describe random variables that take only integer values.

Bernoulli Distribution

- Consider an experiment with only two possible outcomes:
  - Examples: success/failure, presence/absence of a trait, male/female, alive/dead, etc.
- Let X be the random variable that assigns 0 to failure (e.g., unsuccessful surgery) and 1 to success. This variable is called a Bernoulli random variable.

Bernoulli Distribution: Expectation and Variance

- Let p denote the probability of the event "success":

$$\begin{aligned} \mathbb{P}(\{\text{success}\}) &= \mathbb{P}(X=1) = p, \\ \mathbb{P}(\{\text{failure}\}) &= \mathbb{P}(X=0) = 1-p = q, \end{aligned}$$

  with $p+q=1$.
- The distribution of X is:

X	0	1
$\mathbb{P}(X=x_i)$	$q=1-p$	p

- $\mathbb{E}(X) = \mu_X = p$.
- $\mathrm{Var}(X) = \sigma_X^2 = p\,q$.

Binomial Distribution: Definition

- Consider repeated, independent trials of the same Bernoulli experiment.
- Each trial has only two possible outcomes: success or failure.
- As before, p denotes the probability of success.
- We associate to this sequence of n trials the random variable X, which counts the **number of successes obtained**.
- This distribution is widely used, for example, in modelling allele and phenotype frequencies in genetics.

Binomial Distribution: Modelling

- n = number of trials.
- p = probability of success; $q = 1 - p$.
- x = **number of successes** observed over n **independent and identically distributed Bernoulli trials**. This defines the random variable X.

Example

- An urn contains two types of balls: white balls with proportion p and non-white balls with proportion $q = 1 - p$.
- Consider the experiment that consists of drawing n balls **with replacement**. Let X be the number of white balls obtained in n draws.
- The set of possible values is:

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

Example

- For each draw i, define the Bernoulli random variable X_i:

$$X_i = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } q = 1 - p. \end{cases}$$

- Then the total number of successes is

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

with each $X_i \sim \text{Bernoulli}(p)$.

Deriving the Distribution

- **Question:** What is the probability that exactly k successes are obtained? That is, what is $\mathbb{P}(X = k)$ for $k \in \{0, \ldots, n\}$?
- Recall that $C_n^k = \dfrac{n!}{k!(n-k)!}$ counts the number of ways to obtain k successes.
- The probability of a specific arrangement with k successes is $p^k q^{n-k}$.
- Therefore,

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k}, \qquad k = 0, 1, \ldots, n.$$

- One can verify:

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1.$$

Definition

Binomial Distribution

A random variable X follows a Binomial distribution with parameters n and p if:

- $X(\Omega) = \{0, 1, \ldots, n\}$,
- and for every $k \in \{0, \ldots, n\}$,

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k}, \qquad q = 1 - p.$$

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

Expectation and Variance

- If $X \sim \mathcal{B}(n,p)$, then:

◇ **Expectation:**

$$\mathbb{E}(X) = n\,p.$$

◇ **Variance:**

$$\mathrm{Var}(X) = n\,p\,q.$$

Example

- A fair coin is tossed 6 times. Call "heads" a success. Thus: $p = \frac{1}{2}$, $n = 6$.
  a) Probability of obtaining exactly 2 heads:

$$\mathbb{P}(X = 2) = C_6^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = \frac{15}{64}.$$

  b) Probability of obtaining at least 4 heads:

$$\mathbb{P}(X \geq 4) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) = \frac{15}{64} + \frac{6}{64} + \frac{1}{64} = \frac{11}{32}.$$

Example: Multiple Choice Test

- A test consists of 4 multiple choice questions, each with 5 possible answers, only one of which is correct. To pass the test, a student must answer at least 3 questions correctly.
- What is the probability that a student who did not study passes the exam?

Example: Solution

- If the student guesses randomly, each answer has probability $\frac{1}{5}$ of being correct. Thus: $p = \frac{1}{5}$.
- Each question corresponds to a Bernoulli trial. Hence:

$$X \sim \mathcal{B}(4, \tfrac{1}{5}), \qquad X = \text{number of correct answers.}$$

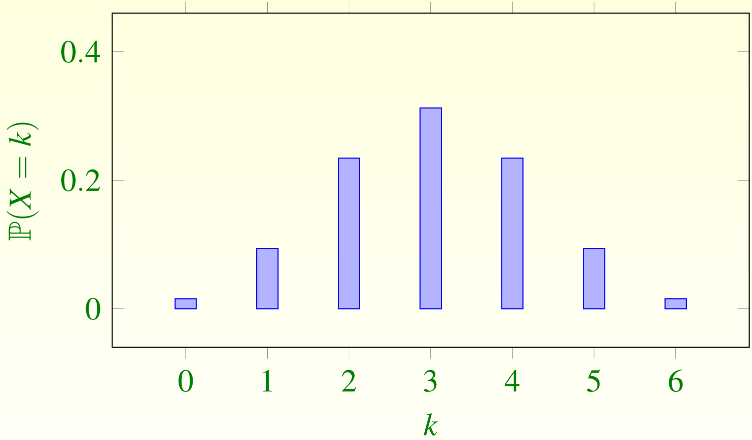
- Passing the exam corresponds to:

$$X \geq 3, \qquad \mathbb{P}(X \geq 3) = 0.0272.$$

Example: Biological Application

- What is the probability of finding $x = 0, 1, 2,$ or 3 girls in families with $n = 3$ children?

Binomial PMF (example): $X \sim \mathcal{B}(6, \frac{1}{2})$



This chart illustrates the PMF used in the coin example: e.g. $\mathbb{P}(X = 2) = 15/64$ and $\mathbb{P}(X \geq 4) = 11/32$.

Poisson Distribution

- A distribution used to model the occurrence of events that are:
  - **rare** (law of rare events),
  - **independent**.
- The random variable X represents the **number of events occurring during a given time period**.

Poisson Distribution: Applications

- It applies to counts such as: the number of accidents, the occurrence of rare anomalies (exceptional diseases), waiting-line models, the number of bacterial colonies on a Petri dish, quality-control defects in pharmaceutical manufacturing, etc.

Definition

Definition

- Let X be the random variable representing the number of independent occurrences of a rare event in an infinite population (**taking values** $0, 1, 2, \ldots$).
- The probability of observing exactly k occurrences is

$$\mathbf{Pr}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad \lambda > 0,\ k \in \mathbb{N}.$$

- We write $X \rightsquigarrow \mathcal{P}(\lambda)$; the distribution depends on a single positive parameter λ.

Properties

1. The support is $X(\Omega) = \mathbb{N}$, and
$$\sum_{k=0}^{\infty} \mathbf{Pr}(X = k) = 1.$$

2. If X_1 and X_2 are independent Poisson variables with parameters λ_1 and λ_2, then their sum is Poisson:
$$X_1 + X_2 \rightsquigarrow \mathcal{P}(\lambda_1 + \lambda_2).$$

Expectation and Variance

If $X \rightsquigarrow \mathcal{P}(\lambda)$, then:

1. The expectation is:

$$\mu_X = \mathbb{E}(X) = \lambda.$$

2. The variance is:

$$\sigma_X^2 = \mathrm{Var}(X) = \lambda.$$

Remark

- If we know $\mathbf{Pr}(X = 0) = p$ (probability of observing no event), then
$$p = \mathbf{Pr}(X = 0) = e^{-\lambda}.$$

- Hence $\lambda = -\ln(p)$, and we obtain
$$\begin{array}{rcccl}
\mathbf{Pr}(X = 1) & = & e^{-\lambda}\frac{\lambda}{1!} & = & p\,\lambda, \\
\mathbf{Pr}(X = 2) & = & e^{-\lambda}\frac{\lambda^2}{2!} & = & \mathbf{Pr}(X = 1)\,\frac{\lambda}{2}, \\
\vdots & & \vdots & & \vdots \\
\mathbf{Pr}(X = k) & = & e^{-\lambda}\frac{\lambda^k}{k!} & = & \mathbf{Pr}(X = k-1)\,\frac{\lambda}{k}.
\end{array}$$

- Probabilities can therefore be computed recursively.

Poisson Distribution: Example

- If in a given region there are on average 10 cases of a certain rare disease in a fixed period, then the number of observed cases follows a Poisson distribution with parameter $\lambda = 10$.

Poisson Tables

- At the intersection of the column λ and the row k appears the probability that a Poisson random variable with parameter λ takes the value k:

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

Approximating a Binomial Distribution by a Poisson Law

- Let $X \rightsquigarrow \mathcal{B}(n, p)$.
- If p is small (**typically** $p < 0.1$), n **large** ($n > 50$), and $0 \leq np \leq 10$,
- then the binomial distribution can be approximated by $Y \rightsquigarrow \mathcal{P}(\lambda = np)$.
  - $\Pr(X = k) \approx \Pr(Y = k)$,
  - $\Pr(Y > n)$ is small but nonzero.
- Poisson probabilities are often easier to compute than binomial ones.

Connection with the Binomial Law

- For large samples, when counting the number of occurrences of a "rare event", the binomial law $\mathcal{B}(n, p)$ (with n large and p small) can be approximated by a Poisson law with parameter $\lambda = np$.
- Typical examples include: number of vaccine-induced accidents, number of homicides or suicides in a large city over a year, etc.

Binomial-Poisson Approximation: Remark

1. Since X is binomial, its values cannot exceed n, whereas the Poisson approximation allows larger values. However, the probabilities beyond n are extremely small.
2. When $5 \leq np \leq 10$, both approximations (Poisson and normal) may be used. The Poisson approximation is more accurate when $q = 1 - p$ is close to 1.

Approximation: Example

- The probability of a mutation in an individual is 10^{-4}. How many individuals must be examined to be "almost certain" to observe at least one mutation?
- Let n be this sample size and let X be the number of mutants. Then $X \rightsquigarrow \mathcal{B}(n, 10^{-4})$. Since p is very small and n large, we may approximate X by $\mathcal{P}(\lambda = n \cdot 10^{-4})$.

Example (continued)

- If we interpret "almost certain" as probability ≥ 0.95:

$$\mathbf{Pr}(X \geq 1) \geq 0.95 \quad \Longleftrightarrow \quad \mathbf{Pr}(X = 0) < 0.05.$$

- Since $\mathbf{Pr}(X = 0) = e^{-\lambda}$,

$$\begin{aligned} e^{-n \cdot 10^{-4}} < 0.05 \quad & \Longleftrightarrow \quad n \geq \frac{-\ln(0.05)}{10^{-4}} \\ & \Longleftrightarrow \quad n \geq 29958. \end{aligned}$$

Normal Approximation of a Binomial Distribution

- If $X \rightsquigarrow \mathcal{B}(n, p)$, then mean $m = np$ and variance $\sigma^2 = npq$.
- If $n > 30$, $np \geq 5$, and $nq \geq 5$,

$$\frac{X - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1),$$

and

$$\mathbf{Pr}\left(\frac{X - m}{\sigma} < x\right) \approx F(x).$$

- With continuity correction:

$$\mathbf{Pr}(X = k) \approx \mathbf{Pr}(k - 0.5 \leq Y \leq k + 0.5).$$

- Probabilities $\mathbf{Pr}(Y < 0)$ and $\mathbf{Pr}(Y > n)$ are small but nonzero.

Normal Approximation of a Poisson Distribution

- If $X \rightsquigarrow \mathcal{P}(\lambda)$,
- and λ is large ($\lambda \geq 15$),
- then we may approximate

$$X \approx \mathcal{N}(\lambda, \lambda),$$

where the mean is λ and the standard deviation is $\sigma = \sqrt{\lambda}$.

- With continuity correction:

$$\mathbf{Pr}(X = k) \approx \mathbf{Pr}(k - 0.5 \leq Y \leq k + 0.5).$$

Exercise 1

In a family, the probability that a child is left-handed is 0.25.

a) What probability distribution is followed by the random variable X, the number of left-handed children in a family of n children? What type of distribution is it? Give the formula for $\mathbb{P}(X = k)$.

b) What is the probability of having exactly 2 left-handed children in a family of 9 children?

c) What is the probability of having exactly 6 right-handed children in this family?

d) What is the probability of having at most 6 right-handed children in this family?

Exercise 2

In a human population, the proportion of left-handed individuals is 1%. What is the probability of observing at least 4 left-handed people in a sample of 230 individuals?

Exercise 3

In a company, an average of 4 workplace accidents occur per month. Let X be the random variable representing the number of workplace accidents in this company.

1. Give the possible values of X.
2. Give the probability distribution of X.
3. Compute the probability that $X \geq 1$.

Exercise 4

In a Mexican city, the average number of people infected each day by the H1N1 swine flu virus is about 20. Let X be the random variable counting the number of people infected each day. We assume that one day corresponds to one time unit.

1. What is the probability distribution of X?
2. Can X be approximated by a normal distribution?
3. Using the standard normal table, compute the probability that the number of infected tourists on the day of their arrival is between 5 and 20.

Exercise 5

The age at which a child says their first words is normally distributed with mean 12 months and standard deviation 2.5 months.

1. What proportion of children say their first words before the age of 9 months?
   $(A)\ < 0.01 \qquad (B) = 0.11507 \qquad (C) = 0.13786 \qquad (D) = 0.86214 \qquad (E) > 0.88$
2. Determine the age above which only 2% of children say their first words.
   $(A) < 6 \qquad (B) = 6.875 \qquad (C) = 17.125 \qquad (D) = 17.15 \qquad (E) > 18$
3. The newborn Mohamed is 9.5 months old and has not yet spoken his first word. What is the probability that he will say his first word after 9.5 months?
   $(A) < 0.1 \qquad (B) = 0.15866 \qquad (C) = 0.84134 \qquad (D) = 0.94295 \qquad (E) = 1$

Exercise 5 (continued)

4. The newborn Fatima is 9.5 months old and has not yet spoken her first word. What is the probability that she will say her first word between 9.5 and 10 months?
   $(A) = 0.06323 \quad (B) = 0.44039 \quad (C) = 0.0532 \quad (D) = 0.33531 \quad (E) > 0.6$

5. The median of the random variable "Age" is:
   $(A) = 0 \quad (B) = 12 \quad (C) = 13.25 \quad (D) = 14 \quad (E) = 15$

Bibliographie

- Cours Probabilité, Benchikh Tawfik, http://www.univ-sba.dz/lsps/images/pdf/cours/
- Probabilités et Biostatistique. www.chups.jussieu.fr/polys/biostats/CoursProba1.pdf
- FMPMC Pitié-Salpêtrière-probabilités, cours 1 et 2. www.chups.jussieu.fr/polys/biostats/coursProba-1-2/
- Notions de Probabilités, Roch Giorgi, LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France http://lertim.fr