Contents

Chapter 1

Unit 1

1.1 Introduction

Science is the art of dealing with uncertainty. There doesn't exist any absolute certainty, each scientific truth has a non-zero probability of being false.
Hypothesis can be verified through experiments.

- Experiment: process where measurements are made
- Measurements: random (stochastic [non-determined intrinsic system])variables (stochastic), determined by the experiment

A certain measurement can embrace a wide range of possibles values and will be assigned a probability for each one.

Statistical inference: The properties of a system can be inferred from experimental results.

1.2 Probability

Probability theory is a branch of pure mathematics and it's based on axioms from which properties and theorems can be derived.

Let $\mathcal{E}$ an event space, let A be an event (a subset of $\mathcal{E}$), then

Axiom 1. $P(A) \geq 0$

Axiom 2. $P(\mathcal{E}) = 1$

Axiom 3. *If A and B have no elements in common (they're mutually exclusive or disjoint), then*

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$$

As of them, it can be derived the following

Theorem 1.2.1. *Let $\overline{A}$ be the complementary of A, such that*

$$A \cup \overline{A} = \mathcal{E}, A \cap \overline{A} = \varnothing \Rightarrow P(\overline{A}) = 1 - P(A)$$

Theorem 1.2.2. $0 \leq P(A) \leq 1$

Theorem 1.2.3. $P(\varnothing) = 0$

Theorem 1.2.4. *Mutually exclusive (or disjoint) events*

$$P(A + B + C + \dots) = P(A) + P(B) + P(C) + \cdots = P(A \cup B \cup C \cup \dots)$$

Theorem 1.2.5. *If $A \subset B \Rightarrow P(A) \leq P(B)$*

Theorem 1.2.6. *If A and B aren't disjoint, then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

It can be easily seen using Venn's diagrams, the last term is due to forbidden double counting.

1.3 Probability's interpretations

1.3.1 Objective interpretation

It isn't trivial to assign probabilities to different subsets of an event space. Those probabilities can be assigned as **relative frequencies**:

- Is **objective**
- States that the concept of probability can only be applied on problems which can be repeated many times (set of measurements of an experiment)

An *experiment* measures one or more magnitudes.
A *variable* is the result of the measurement (discrete or continuous).

- The same value is not obtained when repeating the experiment, even under the same conditions.
- Random (or stochastic) variable

Definition 1.3.1. Objective definition of the **recurrence frequency**, where an experiment is done N times and we register A event n times, then the probability is

$$P(A) = \lim_{N\to\infty} \frac{n}{N}$$

As we see, the probability P can't be determined strictly as $N \to \infty$, even though depending on the context makes sense when talking about *convergence*.

The sample (or event) space $\mathcal{E}$ is defined by every value of the random/stochastic variable.

Definition 1.3.2. Discrete variable:

- $\mathcal{E}$ can embrace a finite number of events.
- An event X_i has an associated probability of $P(X_i)$.
- If the variable isn't numeric (say head/tails), a number can be assigned to it.

Definition 1.3.3. Continuous variable:

- $\mathcal{E}$ can embrace an infinite number of events.
- An event is the subset of sample points in $[x, x + dx]$
- The probability of x is $P(x \in [x, x + dx]) = f(x)dx$.
  where $\mathbf{f(x)}$ is the *probability density* of x.

Example 1.3.1. An experiment which consists of measuring the length (x) and weight (y) of loaves made by a bakery. The sample space $\mathcal{E}$ is the positive quadrant of the xy plane.

Example 1.3.2. An experiment which consistes of throwing up a coin. The result can be head or tails, we can assign 0 to head and 1 to tails, the sample set $\mathcal{E}$ is $\{0, 1\}$.

Example 1.3.3. In example 1.3.1., let's divide the sample space in intervals and also, we know that x and y values are restricted below certain thresholds. The sample space $\mathcal{E}$ now is finite.

1.3.2 Subjective interpretation

The **subjective (or Bayesian)** probability states that the probability is a subjective concept. Instead of events, let's talk about hypothesis.

The probability of the hypothesis is the degree of belief we have on it, which we can express as a *number*, a quantity assigned for the purpose of representing a state of knowledge, or a state of belief.

This concept of probability embraces the previous one of relative frequency and can be applied to more general problems.

Example 1.3.4. Wondering about the probability of tomorrow raining is justified, either tomorrow rains or either it doesn't, but there's only one tomorrow, we'll never be able to repeat the experiment.

In these cases, **Bayes' Theorem** is specially useful as it allows to modify an existing probability as of experimental data. Firstly, an a priori probability is specified, then it is updated thanks to new relevant data coming from experimentalists ... reinforces or weakens the starting hypothesis.

1.4 Probability distribution

Definition 1.4.1. A probability function is some function that may be used to define a particular probability distribution.

Definition 1.4.2. The **probability density function (PDF)** or **density** of a continuous random variable, is a function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Definition 1.4.3. The **cumulative distribution function (CDF)** of a real-valued random variable x evaluated at a is the probability that x will take a value less than or equal to x. Moreover it obeys

$$F(a) = \int_{-\infty}^{a} f(x)dx \qquad f(x) = \frac{dF(x)}{dx}$$

In case of discrete variable, the integrals turn into summations.

Proposition 1.4.1. *The cumulative distribution function satisfies that*

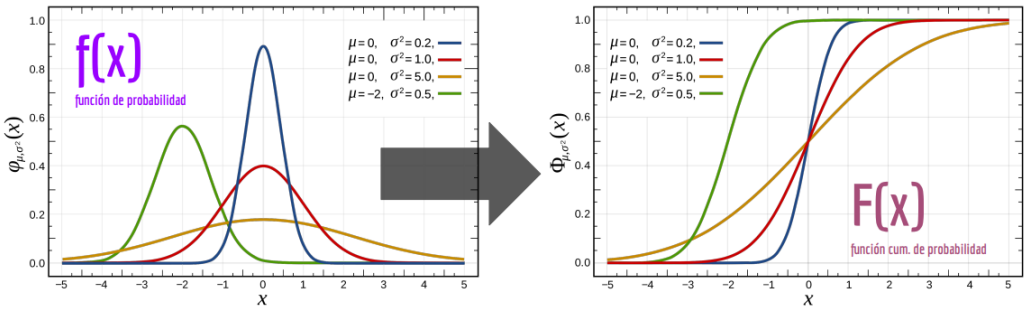
$$0 \leq F(x) \leq 1$$

Proposition 1.4.2. *If the CDF isn't decreasing (it's increasing or constant) then*

$$F(a) \leq F(b) \text{ if } a < b$$

Proposition 1.4.3. *Let the random variable be continuous, then*

$$P(a \leq x \leq b) = F(b) - F(a)$$



Proposition 1.4.4. *Let $F(x)$ be continuous and flat in x, then*

$$P(X = x) = 0$$

The **Kolmogorov-Smirnov** test is based on cumulative distribution functions and can be used to see if two empirical distributions are different or if an empirical distribution is different from an ideal distribution.

1.4.1 Important properties

Definition 1.4.4. The **expected value** of a random variable X, denoted $E(X)$ or $E[X]$, is a a generalization of the weighted average, and is intuitively the arithmetic mean of a large number of independent realizations of X. The expected value is also known as the expectation, mathematical expectation, mean, average, or first moment.
Let $u(X)$ be a function, where X has a the PDF $f(x)$, then the expected value is

$$E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx \qquad E[u(X)] = \sum_{i=1}^{n} u(X_i)P(X_i)$$

- The integral there is an improper integral, but usually has bounded limits becoming a definite integral.
- The expected value is a number, not a variable

It's immediate to derive the following properties.

Proposition 1.4.5.

$$E[u + v] = E[u] + E[v]$$

Proposition 1.4.6.

$$E[cu] = cE[u]$$

Proposition 1.4.7.

$$E[c] = c$$

Definition 1.4.5. The **moment of m-th order** of a random variable X with respect to the point c is defined as

$$\alpha_m^c = E[(x_c)^m] = \int_{-\infty}^{\infty} (x_c)^m f(x)dx$$

Definition 1.4.6. The **mean** is the expected value of the random variable x, namely, the moment of first order with respect to 0.

$$\mu = \int_{-\infty}^{\infty} xf(x)dx$$

Definition 1.4.7. The **variance** of a random variable x is the moment of second order with respect to the mean μ.

$$V[x] \equiv \sigma^2[x] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

Proposition 1.4.8.

$$V[cx] = c^2 V[x]$$

Proposition 1.4.9.

$$V[x] = E[x^2] - \mu^2$$

Definition 1.4.8. The **standard deviation** of a random variable x is the positive square root of the variance

$$\sigma = +\sqrt{\sigma^2}$$

Definition 1.4.9. The **statistical bias** of a random variable x is the third order momentum with respect to the mean, divided by the cubic standard deviation

$$\gamma_x = \frac{\alpha_3^\mu}{\sigma^3} = \frac{\int_{-\infty}^{\infty}(x-\mu)^3 f(x)dx}{\sigma^3}$$

This momentum tells us about the **symmetry** of the distribution with respect to the mean.

$$\begin{cases} = 0 & \text{symmetric} \\ \neq 0 & \text{asymmetric} \end{cases}$$

Definition 1.4.10. The **reduced (or normal) variable** of a random variable x is the following *dimensionless* quantity which normalizes the probability distribution

$$u = \frac{x-\mu}{\sigma}$$

where we obtain the following results

$$\begin{cases} E[u] = 0 \\ V[u] = 1 \end{cases}$$

Definition 1.4.11. The **median** x_m is the value separating the higher half from the lower half of a data sample, a population, or a probability distribution. For a data set, it may be thought of as "the middle" value. Indeed, the median is the value of x such that the probability of finding $x \leq x_m$ is equal to the probability of finding $x > x_m$

$$F(x_m) = \int_{-\infty}^{x_m} f(x)dx = 0,5$$

Definition 1.4.12. The **mode** x_M is the value that appears most often in a set of data values.[1] If X is a discrete random variable, the mode is the value x (i.e, X = x) at which the probability mass function takes its maximum value. In other words, it is the value that is most likely to be sampled.

$$\left.\frac{df(x)}{dx}\right|_{x=x_M} = 0$$

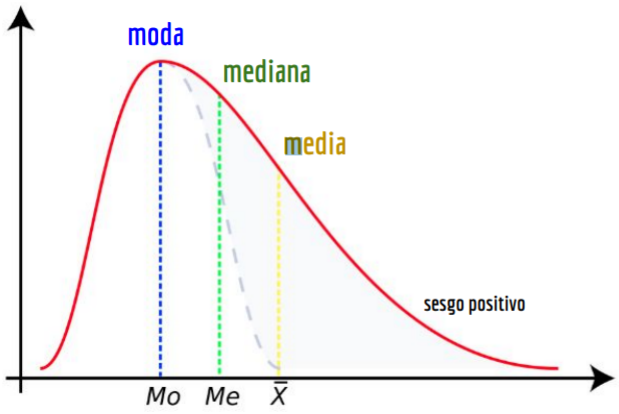
A probability distribution can be unimodal or multimodal (as the bimodal distribution).

Definition 1.4.13. The**percentile** (or a centile) is a score below which a given percentage of scores in its frequency distribution fall (exclusive definition) or a score at or below which a given percentage fall (inclusive definition).

$$F(x_q) = \int_{-\infty}^{x_q} f(x)dx = q$$

The *mean* is the half order percentile. The more frequent ones are the quartile and decile.

Many times the mean coincides with the median, and even the mode (symmetric probability distribution), other times it doesn't. Let's see the difference between the average wage and the most probable wage (asymmetric distribution).



2017
(www.ine.es)
20.600€ bruto
17.200€ bruto
13.500€ bruto

1.5 Change of variables

Let x be a random variable and $y(x)$ a function of this variable, we want to find the density probability of y, $g(y)dy$, namely, the probability of finding y in the interval $[y, y+dy]$, when $x \in [x, x+dx]$ from the original PDF of x, $f(x)$.

Definition 1.5.1. The PDF of $y(x)$ is

$$g(y)dy = \left| \int_{x(y)}^{x(y+dy)} f(x)dx \right.$$

Chapter 2

Summary

2.1 Probability

Let $\mathcal{E}$ an event space, let A be an event (a subset of $\mathcal{E}$), then

Axiom 4. $P(A) \geq 0 \qquad P(\mathcal{E}) = 1$

Axiom 5. *If A and B have no elements in common (they're mutually exclusive or disjoint), then*

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$$

As of them, it can be derived the following

Theorem 2.1.1. *Let $\overline{A}$ be the complementary of A, such that*

$$A \cup \overline{A} = \mathcal{E}, A \cap \overline{A} = \varnothing \Rightarrow P(\overline{A}) = 1 - P(A)$$

Theorem 2.1.2. $0 \leq P(A) \leq 1$

Theorem 2.1.3. $P(\varnothing) = 0$

Theorem 2.1.4. *Mutually exclusive (or disjoint) events*

$$P(A + B + C + \dots) = P(A) + P(B) + P(C) + \cdots = P(A \cup B \cup C \cup \dots)$$

Theorem 2.1.5. *If $A \subset B \Rightarrow P(A) \leq P(B)$*

Theorem 2.1.6. *If A and B aren't disjoint, then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

It can be easily seen using Venn's diagrams, the last term is due to forbidden double counting.

Definition 2.1.1. (Conditional probability). Let A and B be two events of the same event space $\mathcal{E}$, then the conditional probability of A knowing information B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Proposition 2.1.7. *Let A and B be two events of the same event space $\mathcal{E}$, then*

$$P(A|A) = 1$$
$$P(A|B) = 1 \Leftarrow B \subset A$$
$$P(A|B) = 0 \Leftarrow A \cap B = \varnothing$$

Definition 2.1.2. (Independent events). et A and B be two events of the same event space $\mathcal{E}$, then we say these events are independent if

$$P(A|B) = P(A) \Rightarrow P(A \cap B) = P(A)P(B) \qquad P(A), P(B) \neq 0$$

Note it is easy to see that if A is independent from B, then B is independent from A.

Observation 2.1.1. *We should not confuse independent events with mutually exclusive or disjoints events. The intersection of the previous two is zero. Nevertheless, independent events can have elements in common.*

Law 1. *(Total probability law). Let $B_i = \{B_1, \dots, B_n\}$ be a set of n disjoints events and let A be an event of the same event space $\mathcal{E}$ which might have elements in common with B_i. Then the Law of total probability states that*

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

Theorem 2.1.8. *(Bayes Theorem). Let $B_i = \{B_1, \dots, B_n\}$ be a set of n disjoints events and let A be an event of the same event space $\mathcal{E}$ and $P(A) > 0$. Then*

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A)} = P(A|B_i)\frac{P(B_i)}{\sum_j P(A|B_j)P(B_j)}$$

2.2 Distributions

2.2.1 Discrete variable

Definition 2.2.1. (Distribution Function). The distribution function of a random variable X is

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F(x) = P\{X \leq x\} \end{aligned}$$

Definition 2.2.2. (Discrete random variable). A random variable X is *discrete* if there exists a finite or numerable set $S \subset \mathbb{R}$ such that

$$P\{X \in S\} = 1$$

The set S is called the *support* of the distribution of X if $P\{X = x\} > 0$ for all $x \in S$. The probability function is

$$\begin{aligned} p : S &\longrightarrow [0, 1] \\ x &\longmapsto p(x) = P\{X = x\} \end{aligned}$$

Theorem 2.2.1. *The support and the probability density function of a discrete random variable automatically fixes the distribution.*

Definition 2.2.3. (Bernoulli distribution). Let X be a random discrete variable, we say this variable follows a *Bernoulli distribution* if it takes the value 1 for the probability of success and 0 for the probability of failure ($1 - p = q$). In this case we say $X \sim B(p)$. His support is $S = \{0, 1\}$ and its probability function is

$$p(k) = \begin{cases} p & k = 1 \\ q & k = 0 \end{cases}$$

Its expected value and variance are

$$\mu = \mathrm{E}[k] = p \qquad \sigma^2 = \mathrm{Var}[k] = p(p - 1)$$

Definition 2.2.4. (Binomial distribution). Let X be a discrete random variable, we say this variable follows a *binomial distribution* if it evaluates the number of successes in $n \in \mathbb{N}$ attempts with p as success probability ($1 - p = q$). In this case $X \sim B(n, p)$. Its support is $S = \{0, 1, \dots n\}$ and its probability function is

$$p(k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \forall k \in S$$

Its expected value and variance are

$$\mu = \mathrm{E}[k] = np \qquad \sigma^2 = \mathrm{Var}[k] = npq$$

Definition 2.2.5. (Discrete uniform distribution). Let X be a discrete random variable, we say it follows a *discrete uniform distribution* if its support is $S = \{x_1, x_2, \dots, x_n\}$ with $x_i \in \mathbb{R}$ different two to two and its probability function is

$$p(x_i) = \frac{1}{n} \quad \forall x_i \in S$$

In this case we say $X \sim Unif(\{x_1, \dots, x_n\})$.

Its expected value, variance and bias are

$$\mu = \mathrm{E}[X] = \frac{n + 1}{2} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{n^2 - 1}{12}$$

Definition 2.2.6. (Poisson distribution). Let X be a discrete random variable, we say it follows a *Poisson distribution* of parameter $\lambda > 0$ if its support is $S = \mathbb{N} \cup \{0\}$ and its probability function is

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \forall k \in S$$

In this case $X \sim Poiss(\lambda)$.

Its expected value, variance and bias are

$$\mu = \mathrm{E}[k] = \lambda \qquad \sigma^2 = \mathrm{Var}[k] = \lambda \qquad \mu c_3 = \sqrt{\lambda}$$

Definition 2.2.7. (Hypergeometric distribution). Let X be a discrete random variable. Let be a container with N balls where K are white. We say X follows an *hypergeometric distribution* if it computes the quantity of white balls taken in n throws without repetition. In this case we say $X \sim HGeom(N, K, n)$. Its support is $S = \{k \in \mathbb{N} \mid k \leq \min n, K, n-k \leq \min n, N-K\}$ and its probability function is

$$p(k) = P\{X = k\} = \frac{\binom{N}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \forall k \in S$$

Its expected value and variance are

$$\mu = \mathrm{E}[X] = n\frac{K}{N} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{nK}{N}\left(1 - \frac{K}{N}\right)\left(\frac{N-n}{N-1}\right)$$

2.2.2 Continuous variable

Definition 2.2.8. (Absolutely continuous random variable). We say X is an absolutely continuous random variable if there exists a probability density function (PDF) f such that

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(y)dy \quad \forall x \in \mathbb{R}$$

Definition 2.2.9. (Probability density function). Analogous to distribution function, we say a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a PDF if

$$f(x) \geq 0 \quad \forall x \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

Definition 2.2.10. (Cumulative probability function). Analogous to probability function, we say a function $F : \mathbb{R} \longrightarrow \mathbb{R}$ is a CDF if

$$F(x) = P\{X \leq x\} \quad \text{and} \quad f(x) = \frac{dF(x)}{dx}$$

Observation 2.2.1. *Let X be an absolutely continuous random variable with distribution function $F(x)$, then its probability density function is $f(x) = F'(x)$*

Definition 2.2.11. (Normal or Gaussian distribution). Let X be an absolutely continuous random variable, we say it follows a *normal distribution* if its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. In this case we say $X \sim N(\mu, \sigma^2)$ and if X follows a standard/normalized normal distribution if $X \sim N(0, 1)$.

Its expected value, variance and Γ_{FWHM} are

$$\mu = \mathrm{E}[k] \qquad \sigma^2 = \mathrm{Var}[k] \qquad \Gamma_{FWHM} = 2\sigma\sqrt{2\ln 2}$$

Definition 2.2.12. (Continuous uniform distribution.) Let X be an absolutely continuous random variable, we say it follows a *continuous uniform distribution* in an interval $[a, b]$ if its PDF is

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x)$$

In this case $X \sim Unif([a, b])$.

Its expected value and variance are

$$\mu = \mathrm{E}[X] = \frac{b+a}{2} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{(b-a)^2}{12}$$

Definition 2.2.13. (Exponential distribution). Let X be an absolutely continuous random variable, we say it follows an *exponential distribution* of parameter $\lambda > 0$ if its PDF is

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \quad \forall x \in (0, \infty)$$

In this case $X \sim \mathscr{E}xp(\lambda)$.

Its expected value and variance are

$$\mu = \mathrm{E}[X] = \lambda \qquad \sigma^2 = \mathrm{Var}[X] = \lambda^2$$

Definition 2.2.14. (Gamma distribution). Let X be an absolutely continuous random variable, we say it follows a *Gamma distribution* of parameters $\mu > 0$ and $k > 0$ if its PDF is

$$f(x) = \frac{\mu^k}{\Gamma(k)} x^{k-1} e^{-\mu x}$$

In this case $X \sim Gamma(\mu, k)$.

Its expected value and variance are

$$\mu = \mathrm{E}[X] = \frac{k}{\mu} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{k}{\mu^2}$$

Observation 2.2.2. *This last distribution is reduced to exponential distribution for $k = 1$, to χ^2 distribution with $m = 1/2$ and $k = n/2$ with n degrees of freedom and resembles Poisson's distribution with $\mu_P = \mu x$.*

Definition 2.2.15. (Chi-squared distribution). Let X be an absolutely continuous random variable, we say it follows a χ^2 *distribution* of parameter $n \in \mathbb{N}$ if its PDF is

$$f(x; n) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad x > 0$$

In this case $X \sim \chi^2(n)$.

Its expected value, variance and Γ_{FWHM} are

$$\mu = \mathrm{E}[X] = n \qquad \sigma^2 = \mathrm{Var}[X] = 2n$$

Definition 2.2.16. (t-Student distribution). Let be n independent variables X_i which come from the same distribution, with mean μ and σ unknown. Then we say X follows a *t-Student distribution* with $n - 1 = r$ degrees of freedom if its PDF is

$$f(t; n-1) \equiv f(t; r) = \frac{\Gamma\left(\frac{1}{2}(r+1)\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{1}{2}r\right)} \left(1 + \frac{t^2}{r}\right)^{-(r+1)/2}$$

In this case $X \sim$ t-Student(n).

Its expected value and variance are

$$\mu = \mathrm{E}[X] = 0 \qquad \sigma^2 = \mathrm{Var}[X] = \frac{r}{r-2}$$

Definition 2.2.17. (Cauchy [Lorentz] distribution). Let X be an absolutely continuous random variable, we say it follows a *Cauchy distribution* if its PDF is

$$f(x) = \frac{1}{\pi} \frac{\frac{1}{2}\Gamma}{(x-m)^2 + (\frac{1}{2}\Gamma)^2}$$

where m is the mean of the distribution and Γ is the FWHM. The distribution is *symmetric* with respect to m and its CPF is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^- 1\left(\frac{2(x-m)}{\Gamma}\right)$$

Its expected value and variance aren't defined as integral diverges.

Definition 2.2.18. (Standard Cauchy [Lorentz] distribution). Let X be an absolutely continuous random variable, we say it follows a *standard Cauchy distribution* if it follows a Cauchy distribution with $m = 0$ and $\frac{1}{2}\Gamma = 1$ so its PDF is

$$f(x) = \frac{1}{\pi(1+x)^2}$$

In this case we write $X \sim \mathrm{Cauchy}(0, 1)$.

Definition 2.2.19. (Landau distribution). Let X be an absolutely continuous random variable, we say it follows a *Landau distribution* if its PDF is

$$f(x) = \frac{1}{\pi}\int_0^\infty e^{-t(\ln t + xt)} \sin(\pi t) dt \approx \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}(x+e^{-x})}$$

In this case we write $X \sim$ Landau.

Its expected value and variance aren't defined as integral diverges.

Theorem 2.2.2. *(Transformation of random variables). Let X be an absolutely continuous random variable with density $f_X(x)$ and $\mathcal{U} = (a, b)$ such that $-\infty \leq a < b \leq \infty$ in an interval so that $P\{X \in \mathcal{U}\} = 1$. Let $h : \mathcal{U} \longrightarrow \mathcal{V}$ where $\mathcal{V} = (c, d)$ such that $-\infty \leq c < d \leq \infty$ and $h^{-}1 \in \mathcal{C}^1(\mathcal{V})$. Then, $Y = h(X)$ is another absolutely continuous random variable such that*

$$f_Y(y) = f_X(h^{-1}(y))\ |(h^{-1})'(y)| 1_{\mathcal{V}}(y) = f_X(h^{-1}(y)) \left| J\left(\frac{x}{y}\right) \right|$$

is its density function.

2.2.3 Expected value and variance

Definition 2.2.20. (Expected value). Let X be a discrete random variable such that

$$\sum_{k\in S} |k| P\{X = k\} < \infty$$

then we define the expected value of X as

$$\mathrm{E}[X] = \sum_{k\in S} kP\{X = k\}$$

Let X be an absolutely random variable such that

$$\int_{-\infty}^{\infty} |x| f(x) < \infty$$

then we define the expected value of X as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x)$$

Proposition 2.2.3. *(Main properties of expected value). The expected value of a variable X is linear, namely*

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y] \qquad \mathrm{E}[aX] = a\mathrm{E}[X]$$

for any random variable X and Y and any $a \in \mathbb{R}$. Moreover

$$E[a] = a \quad \forall a \in \mathbb{R}$$

Proposition 2.2.4. *(Expected value of a function $z(x)$). Let X be a random variable following a PDF $f(x)$, then a function $z(X)$ is also a random variable, with expected value*

$$\mathrm{E}[z(X)] = \sum_{k\in S} z(k)P(X = k) \qquad \mathrm{E}[z(X)] = \int_{-\infty}^{\infty} z(x)f(x)dx$$

Proposition 2.2.5. *(Transformation of random variables regarding expected value). Let $f(x)$ be a density function of X, let $z(x)$ be another function, then*

$$\mathrm{E}[z(x)] = \mathrm{E}[z(x(y))]$$

Definition 2.2.21. (Median). Let X be a discrete random variable, then we define the median x_m of X as the value with probability

$$P\{X \geq x_m\} = P\{X \leq x_m\} \geq \frac{1}{2}$$

Let X be an absolutely continuous random variable and F(x) its CDF, then we define the median of X as the value x_m for which

$$F(x_m) = \int_{-\infty}^{x_m} f(x)dx = \frac{1}{2}$$

Same number of values to the left and to the right of the median.

Definition 2.2.22. (Mode). Let X be a discrete random variable, then we define the mode x_M of X as the most repeated value.

Let X be an absolutely continuous random variable and $f(x)$ as its PDF, then we define the mode x_M of X as the value

$$\left.\frac{df(x)}{dx}\right|_{X=x_M}$$

Sometimes a distribution can be bimodal if it has two modes or multimodal if it has more than two.

Definition 2.2.23. (Variance and standard deviation). Let X be a random variable such that $\mathrm{E}[X^2] < \infty$, then we define the variance of X as

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

Moreover, we define the standard deviation of X as

$$\mathrm{Sd}[X] = \sqrt{\mathrm{Var}[X]}$$

Proposition 2.2.6. *(Main properties of variance). Let X and Y be random variables and $a \in \mathbb{R}$, then is satisfied*

$$\begin{gathered}\mathrm{Var}[X] \geq 0 \\ \mathrm{Var}[aX] = a^2\mathrm{Var}[X] \\ \mathrm{Var}[X \pm Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] \pm 2\mathrm{CoV}[X, Y] \\ \mathrm{Var}[a] = 0\end{gathered}$$

Proposition 2.2.7. *(Property of variance). Let $X_i = \{X_1, \ldots, X_n\}$ be N random variables, let a_i be a constant, then the variance of the sum of these random variables is*

$$\mathrm{Var}\left[\sum_{i=1}^{N} a_i X_i\right] = \sum_{i=1}^{N} a_i^2 \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{CoV}[X_i, X_j]$$

Definition 2.2.24. (Reduced variable). Let X be a random variable, then the *reduced (normal) variable* of X is defined as

$$u = \frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}}$$

Given this reduced variable, the expected value and the variance now are

$$\mathrm{E}[u] = 0 \qquad \mathrm{Var}[u] = 1$$

Definition 2.2.25. (Covariance). Let X and Y be two random variables such that $\mathrm{E}[|X|], \mathrm{E}[|Y|], \mathrm{E}[|XY|] < \infty$, then we define the covariance of X and Y as

$$\mathrm{CoV}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y]] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

Proposition 2.2.8. *(Main properties of covariance). Let X and Y be random variables and $a \in \mathbb{R}$, then is satisfied*

$$\begin{gathered}\mathrm{CoV}[X, Y] = \mathrm{CoV}[Y, X] \\ \mathrm{CoV}[aX, aY] = ab\mathrm{CoV}[X, Y] \\ \mathrm{CoV}[X + Y, Z] = \mathrm{CoV}[X, Z] + \mathrm{CoV}[Y, Z] \\ \mathrm{CoV}[X, a] = 0\end{gathered}$$

Definition 2.2.26. (Correlation coefficient). Let X and Y be random variables such that $\mathrm{E}[|X|], \mathrm{E}[|Y|], \mathrm{E}[|XY|] < \infty$, then we define the correlation coefficient of X and Y as

$$\rho_{XY} = \frac{\mathrm{CoV}[X, Y]}{\mathrm{Sd}[X]\mathrm{Sd}[Y]}$$

which is a dimensionless quantity $\rho_{XY} \in [-1, 1]$.

Observation 2.2.3. *X and Y are not correlated if and only if* $\mathrm{CoV}[X, Y] = 0$.

Proposition 2.2.9. *Let X and Y be independent random variables such that $E[|X|], E[|Y|], E[|XY|] < \infty$, then*

$$\mathrm{CoV}[X, Y] = 0$$

and, consequently, they are non-correlated

Observation 2.2.4. *The inverse of the previous Proposition doesn't have to be true, two random variables X and Y can have $\mathrm{CoV}[X, Y] = 0$ and still be correlated or dependent.*

Notation 2.2.1. *Despite X following or not a normal distribution, the expected value and the variance of X are usually written as μ and σ^2, respectively.*

2.2.4 Multidimensional continuous variable

Definition 2.2.27. (Probability density function). We say a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a PDF if

$$f(x,y) \geq 0 \quad \forall x \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

Definition 2.2.28. (Cumulative probability function). We say a function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a CDF if

$$F(x,y) = P\{X \leq x, Y \leq y\} \quad \text{and} \quad f(x,y) = \frac{d^2F(x)}{dxdy}$$

Definition 2.2.29. (Marginal distributions). Let X and Y be absolutely continuous random variables following the same PDF $f(x,y)$, then the marginal distributions of X and Y are

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy$$
$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

respectively. They are the PDF of each variable separately, each one is independent from the what's happens to the other.

Definition 2.2.30. (Conditional probability density). Let X and Y be absolutely continuous random variables of the same PDF $f(x,y)$, then the conditional probability density of an event is given by

$$f(y|x) = \frac{f(x,y)}{g(x)}$$

where $f(y|x)$ means x is fixed and $f(y|x)dy$ represents the probability of finding $y \in [y, y+dy]$ when x is fixed.

Definition 2.2.31. (Absolutely continuous independent variables). Let X and Y be absolutely continuous random variables of the same PDF $f(x,y)$ with marginals distributions $g(x)$ and $h(y)$, then we say X and Y are independent between them if

$$f(x,y) = g(x)h(y) \iff f(y|x) = h(y) \quad \text{and} \quad f(x|y) = g(x)$$

Proposition 2.2.10. *(Multivariable generalization). Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ absolutely continuous random variables following a distribution $f(\boldsymbol{x})$, then its marginal distributions and expected values are*

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$$
$$g_i(x_i) \int f(\boldsymbol{x}) \prod_{j \neq i} dx_j$$
$$\mathrm{E}[X_i] = \int_{-\infty}^{\infty} x_i f(\boldsymbol{x}) d\boldsymbol{x}$$

The covariance can be represented with a matrix whose elements are

$$\mathrm{C}_{ij} = \mathrm{CoV}[X_i, X_j] = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j]]$$

*or in vectorial notation, the **covariance matrix** or **error matrix** is*

$$\mathcal{C} = \mathrm{E}[(\boldsymbol{X} - \boldsymbol{E}[X])(\boldsymbol{X} - \boldsymbol{E}[X])^{\mathrm{T}}]$$
$$\mathrm{C}_{ii} = \mathrm{Var}[X_i]$$

The error propagation formula for a function $z(\boldsymbol{x})$ then is

$$\mathrm{Var}[z] = \sum_{i,j=1}^{n} \mathrm{C}_{ij} \left[\frac{\partial z}{\partial x_i} \frac{\partial z}{\partial x_j} \right]_{\boldsymbol{x}=\boldsymbol{E}[X]}$$

A non-linear transformation of covariance matrix is

$$\mathcal{C}_y = \mathcal{T} \mathcal{C}_x \mathcal{T}^{\mathrm{T}} \qquad \mathcal{T} = \left. \frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} \right|_{\boldsymbol{X}=\boldsymbol{E}[X]}$$

2.2.5 Central limit theorem

Theorem 2.2.11. *(Central limit theorem). Let* $\{X_n\}_{n\geq 1}$ *be a succession of independent random variables identically distributed such that* $\mathrm{E}[X_i^2] < \infty$. *Let* $\mathrm{E}[X_i] = \mu$,$\mathrm{Var}[X_i] = \sigma^2$ *and* $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Then*

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{n\to\infty} N(0, \sigma^2)$$

Observation 2.2.5. *Central limit theorem allows alternative definitions, each one equivalent to the rest. Among them, note*

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \xrightarrow{n\to\infty} N(0,1)$$

2.2.6 Random samples and population estimators/statistics

Definition 2.2.32. Let $n \in N$, let X be a random variable following a probability distribution F, then we say a sample of n elements is the set of n independent random variables x_i identically distributed according to F. Denoting it

$$\underline{x} = \{x_1, x_2, \ldots, x_n\}$$

Definition 2.2.33. (Sample mean). Let $\{x_1, x_2, \ldots, x_n\}$ be a sample of an experiment X, then we define the mean of the sample as

$$\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Definition 2.2.34. (Sample median). Let $\{x_1, x_2, \ldots, x_n\}$ be an even sample of an experiment X, then we define the median of the sample as

$$x_m = x_{(n+1)}/2$$

The Median of an odd sample is usually defined to be the mean of the two middle values

$$x_m = \frac{x_{n/2} + x_{(n/2)+1}}{2}$$

Definition 2.2.35. (Sample mode). Let $\{x_1, x_2, \ldots, x_n\}$ be a sample of an experiment X, then we define the mode of the sample to the most repeated value

$$x_M = \text{most repeated } x_i \text{ value}$$

A sample can be bimodal if it has two modes or multimodal it has more than two.

Proposition 2.2.12. *(Expected value and variance of the mean estimator). Let* X *an experiment with* $\mathrm{E}[X] = \mu$ *and* $\mathrm{Var}[X] = \sigma^2$, *and* $\overline{x}$ *the mean of the sample* $\{x_1, x_2, \ldots, x_n\}$ *of* X, *then*

$$\mathrm{E}[\overline{x}] = \mu \qquad \mathrm{Var}[\overline{x}] = \frac{\sigma^2}{n}$$

Definition 2.2.36. (Sample variance). Let $\{x_1, x_2, \ldots, x_n\}$ be a sample. Then we call

$$\widehat{\mathrm{Var}}[x] = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad \widehat{\mathrm{Var}}[x] = S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

the sample variance and S the sample standard deviation of a sample with an unknown mean $\mu \Rightarrow \overline{x}$ and with a known mean μ (having to compute $\overline{x}$), respectively.

Proposition 2.2.13. *(Expected value and variance of the variance estimator). Let X an experiment with $\mathrm{E}[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$, and S^2 the variance of the sample $\{x_1, x_2, \ldots, x_n\}$ of X, then*

$$\mathrm{E}[S^2] = \mathrm{Var}[X] \qquad \mathrm{Var}[s^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right) \qquad \mathrm{Var}[S^2] = \frac{2\sigma^4}{n}\ (Gaussian)$$

when we don't know the mean μ and when we know it, respectively.

Observation 2.2.6. *There are $n-1$ degrees of freedom as one random variable x_j can be found by knowing the rest of the elements and the mean of the sample. That's why it appears $1/(n-1)$ instead of $1/n$. However, a mathematical reason for this is the fact that the sample variance becomes a non-biased estimator with $1/(n-1)$ but a biased estimator with $1/n$.*

Definition 2.2.37. (Absolute and relative frequencies). Let the *absolute frequency* of an event i be the number of times n_i this event happens in an experiment.

If N is the number of total observations regarding the experiment, then the *relative frequency* of an event i is $f_i = n_i/N$.

2.3 Statistical inference

Definition 2.3.1. (Statistical inference). The statistical inference is a branch from statistics which focuses on deducing results from a population undergoing an study, from the analysis of several samples from the same population.

Definition 2.3.2. (Confidence intervals). Let X be an absolutely continuous random variable. Then we call a confidence interval of $1-\alpha$ to an interval (a,b) such that

$$P\{a < X \leq b\} = 1-\alpha$$

We will say the interval (a,b) is centred if its centred in the expected value of X.
Moreover, it can also be defined unilateral confidence intervals $(-\infty, b)$ and (a,∞) so that

$$P\{X \leq b\} = 1-\alpha \quad \text{and} \quad P\{X \geq a\} = 1-\alpha$$

respectively

Definition 2.3.3. (Statistic). We call an *statistic* a quantitative measure calculated from the data of a sample, which allows to estimate or contrast some characteristics of a population. It is common to denote an statistic of a sample $\{x_1, \ldots, x_n\}$ as

$$T = T(x_1, \ldots x_n)$$

Observation 2.3.1. *The sample mean μ and sample variance S^2 are examples of statistics.*

2.4 Parameter estimation

In this section we are gonna talk about parameter estimation which will be useful to specify which distribution follows a known sample. We'll see two important methods for parameter estimation, moments methods and maximum likelihood method.

Definition 2.4.1. (Statistical model). A statistical model is a family of probability distributions.

Example 2.4.1. We will mainly denote statistical models as $\{f(x;\theta)|\theta \in \Theta\} = \{f(x;\theta)\}$ where $f(x;\theta)$ is the probability density function (or probability function, in cases where the random variable is discrete; nevertheless, we can always denote it like $\{p(k;\theta)|\theta \in \Theta\} = \{p(k;\theta)\}$ in certain cases).

We use θ to denote the parameters of the distribution and note that it can also be expressed as a vector, even of infinite dimension, like $\theta = (\theta_1, \ldots, \theta_d)$. We use Θ to denote the domain of θ.

We call statistical models which follow a certain distribution simply by the name of the distribution, for example, Poisson's model is

$$\{p(k;\lambda) \mid \lambda > 0\} = \left\{ \frac{e^{-\lambda}\lambda^k}{k!} \;\middle|\; \lambda > 0 \right\}$$

and the normal model

$$\{f(x;\mu,\sigma) \mid \mu \in \mathbb{R}, \sigma > 0\} = \left\{ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \;\middle|\; \mu \in \mathbb{R}, \sigma > 0 \right\}$$

are statistical models.

Definition 2.4.2. (Parametric and regular model). Let $\{P_\theta \mid \theta \in \Theta\}$ be a statistical model then we say this model is *parametric* if θ is dimension-finite and therefore $\Theta \subset \mathbb{R}$. In this case we denote θ as $(\theta_1, \ldots, \theta_d)$.

We say an statistical model is *regular* if it can be differentiated under the integral sign with respect to θ three times.

Definition 2.4.3. (Estimator and estimate). An estimator $\hat{\theta}$ is an statistic (any quantity computed from values in a sample that is used for a statistical purpose) used to estimate an unknown parameter θ from the values of a sample.

We say $\hat{\theta}$ is an estimate if its value is enough close to θ.

Definition 2.4.4. (Inference problem). Given a sample $\underline{x} = \{x_1, x_2, \ldots, x_n\}$ and a model $\{f(x;\theta)\}$, an inference problem is an statistical problem consisting on finding an estimate and determining a confidence region for the parameter θ (for a fixed confidence).

2.4.1 Moments and its generative function

Before studying moments method let's remember what are moments and take advantage of them to deepen in the topic seeing its generative function and some interesting results.

Definition 2.4.5. (Moment of order r of a random variable). Given a random variable X such that $\mathrm{E}[|X|^r] < \infty$, we define the rth-order moment as

$$\mu_r = \mathrm{E}[X^r]$$

Observation 2.4.1. *The moment of order* 1 *is the expected value* $\mu_1 = \mathrm{E}[X]$ *so* $\mathrm{Var}[X] = \mu_2 - \mu_1^2$ *by definition.*

Definition 2.4.6. (Centred moment of order r of a random variable). Given a random variable X such that $\mathrm{E}[|X|^r] < \infty$, we define the centred moment of order r as

$$\mu c_r = \mathrm{E}[(X - \mathrm{E}[X])^r]$$

We call variance $\mu c_2 = \sigma^2$, bias to μc_3, kurtosis to μc_4, bias coefficient to $sk = \frac{\mu c_3}{\sigma^3}$, kurtosis coefficient to $ku = \frac{\mu c_4}{\sigma^4}$ and kurtosis excess to $ke = ku - 3$.

Observation 2.4.2. *The variance* $\mathrm{Var}[X]$ *is the centred moment of second order* μc_2.

Observation 2.4.3. *The bias of a PDF measures its asymmetry with respect to the expected value. Distributions like normal or t-Student are symmetric as* $\mu c_3 = 0$. *The term bias conflicts with the later bias definition so it can be also found as asymmetry.*

The kurtosis measures the "tailedness" of a PDF looking at the weight of the exterior tails. As kurtosis of $N(0, 1)$ *is* 3*, the excess of kurtosis measures the deviation with respect to this value. If the excess of kurtosis takes positives values, then tails weight more so the distribution will tend to have more extreme values.*

There are cases where neither bias or kurtosis can't be defined or aren't defined.

2.4.2 Maximum likelihood method

Definition 2.4.7. (Likelihood function). Given a sample $\underline{x} = \{x_1, x_2, \ldots, x_n\}$ of a model $f(x;\theta)$ (discrete or continuous) with $\theta = (\theta_1, \ldots, \theta_d)$, we define the likelihood function as

$$L(\theta) = L(\theta; \underline{x}) = \prod_{i=1}^{n} p(x_i; \theta)$$

Observation 2.4.4. *The likelihood function gathers the information of unknown parameters. Works with incomplete information. In the discrete case, it its the probability of observing the sample in function of parameters*

$$L(\theta) = P\{X_1 = x, X_2 = X_2, \ldots, X_n = x_n\} = \prod_{i=1}^{n} p(x_i; \theta)$$

Definition 2.4.8. (Likelihood coefficient). Let $L(\theta_1; \underline{x})$ and $L(\theta_2; \underline{x})$ be likelihood functions of the same sample and model of different unknown parameters θ_1 and θ_2, then the likelihood coefficient is

$$Q = \frac{\prod_{i=1}^{n} L(\theta_1; \underline{x})}{\prod_{i=1}^{n} L(\theta_2; \underline{x})}$$

If $Q > 1$, then it is easy to think that θ_1 value is more probable than θ_2.

Definition 2.4.9. (Maximum likelihood method). Given a sample $\underline{x} = \{x_1, x_2, \ldots, x_n\}$ and chosen a model $f(x;\theta)$, we call the maximum likelihood method to the inference problem which founds an approximate value for θ with

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \{L(\theta)\}$$

Definition 2.4.10. (Log-likelihood function). Let $L(\theta; \underline{x})$ be a likelihood function of a sample and a model, then we define the log-likelihood function as

$$l(\theta) = l(\theta; \underline{x}) = \log\left(L(\theta; \underline{x})\right)$$

as long as $L(\theta; \underline{x}) \neq 0$ for all $\theta \in \Theta$.

Proposition 2.4.1. *Let $L(\theta; \underline{x})$ and $l(\theta; \underline{x})$ be likelihood and log-likehood functions of a sample and a model, then the point where the functions reach the maximum are the same.*

Proof. It's obvious to see as logarithmic and exponential functions are strictly increasing functions. ■

2.5 Comparison and evaluation of estimators

Now that we have seen this two methods for estimating parameters, we are interested in knowing whether this estimates good enough.

In this section are going to be defined some basic properties estimators can satisfy, this way we're going to have the necessary tools to decide which estimators meet our requirements.

Definition 2.5.1. (Bias of an estimator). Let T be an estimator of θ, then the bias of T is

$$\mathrm{b}_\theta(T) = \mathrm{E}_\theta[T] - \theta \qquad \forall\theta$$

Definition 2.5.2. (Non-biased estimator). Let T be an estimator of θ, then we will say it is a non-biased estimator if

$$\mathrm{b}_\theta(T) = 0 \iff \mathrm{E}_\theta[T] = \theta \qquad \forall\theta$$

Observation 2.5.1. *In parameter estimation methods we consider the estimator to be a random variable. Its value changes from sample to sample.*

Definition 2.5.3. (Consistent estimator). Let $\{f(x;\theta)\}$ a model, $\underline{x} = \{x_1, \dots, x_n, \dots\}$ an infinite sample of density $f(x;\theta)$ and $\underline{x}_n = \{x_1, \dots, x_n,\}$ the set of first n observations of $\underline{x}$. Let $\{T_n(x) = T(\underline{x}_n)\}_{n\geq 1}$ the succession of estimators of parameter θ, then we say the estimator T_n is consistent if

$$T_n \xrightarrow{p} \theta \iff \lim_{n\to\infty} \mathrm{Var}[T_n] = 0$$

Definition 2.5.4. (Asymptotically non-biased estimator). Let $\{f(x;\theta)\}$ a model, $\underline{x} = \{x_1, \dots, x_n, \dots\}$ an infinite sample of density $f(x;\theta)$ and $\underline{x}_n = \{x_1, \dots, x_n,\}$ the set of first n observations of $\underline{x}$. Let $\{T_n(x) = T(\underline{x}_n)\}_{n\geq 1}$ the succession of estimators of parameter θ, then we say the estimator T_n is asymptomatically non-biased if

$$\mathrm{E}_\theta[T_n] \xrightarrow{n\to\infty} \theta$$

Proposition 2.5.1. *Under normal conditions, S^2 is a non-biased estimator of σ^2 while m_2 (remember $m_2 = S^2 \cdot (n-1)/n$ is an asymptotically non-biased estimator of σ^2.*

Definition 2.5.5. (Mean squared error). Let T be an estimator of θ parameter, then the mean squared error is

$$\mathrm{MSE}(T) = \mathrm{E}_\theta[(T-\theta)^2]$$

Definition 2.5.6. (Effectiveness). Let T be an estimator of θ parameter, then the efficiency is

$$\mathrm{eff}(T) = \frac{1}{\mathrm{MSE}(T)}$$

Definition 2.5.7. (Efficiency between estimators). Let T_1 and T_2 be two estimators of same parameter θ. We say T_1 is more efficient than T_2 if

$$\mathrm{MSE}(T_1) < \mathrm{MSE}(T_2)$$

Proposition 2.5.2. *Let T be an estimator of θ, then*

$$\mathrm{MSE}(T) = b^2(T) + \mathrm{Var}_\theta[T]$$

Corollary 2.5.3. *Let T be a non-biased estimator of θ. Then*

$$\mathrm{MSE}(T) = \mathrm{Var}_\theta[T]$$

Definition 2.5.8. (Observed and expected Fisher's Information). Let $J(\theta; \underline{x})$ be Fisher's information for a n-dimensional sample...

$$I(\theta) = \mathrm{Var}_\theta[S] = \mathrm{E}_\theta[J(\theta, \underline{x})] = E\left[\left(\frac{\partial}{\partial\theta}\sum_i \ln\left(f(x_i;\theta)\right)\right)^2\right] = \mathrm{E}[l'^2]$$

Theorem 2.5.4. *(Cramér-Rao inequality). Let T be an estimator of θ from a sample $\underline{x}$ of a regular model $\{f(x;\theta)\}$, then it is satisfied*

$$\mathrm{Var}_\theta[T] \geq \frac{(E'_\theta[T])^2}{I(\theta)} = \frac{\left(1 + \frac{\partial b}{\partial\theta}\right)^2}{I(\theta)}$$

Observation 2.5.2. *In case T is a non-biased estimator of θ, Cramér-Rao inequality is reduced to*

$$\mathrm{Var}_\theta[T] \geq \frac{1}{I(\theta)}$$

Definition 2.5.9. We will say an estimator T is efficient if it reaches Cramér-Rao bound.

Definition 2.5.10. (Efficiency). Let T be an estimator of θ parameter, then the efficiency is

$$\epsilon(T) = \frac{\mathrm{Var}_{CRF,\theta}[T]}{\mathrm{Var}_\theta[T]}$$

Proposition 2.5.5. *Let T an efficient and non-biased estimator of θ. Then T is the θ parameter estimator with least mean squared error.*