1 Introduction

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

Axiom 2. *If A and B have no elements in common (they're mutually exclusive or disjoint), then*

$$A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B) \tag{1}$$

Theorem 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 1.2. $0 \leq P(A) \leq 1$

Theorem 1.3. $P(\varnothing) = 0$

Theorem 1.4. *Mutually exclusive (or disjoint) events*

$$P(A + B + C + \dots) = P(A) + P(B) + P(C) + \dots = \tag{2}$$

$$P(A \cup B \cup C \cup \dots) \tag{3}$$

Theorem 1.5. *If $A \subset B \Rightarrow P(A) \leq P(B)$*

Theorem 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 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

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4}$$

Proposition 1.7. *Let A and B be two events of the same event space $\mathcal{E}$, then*

$$\begin{aligned} &P(A|A) = 1 \\ &P(A|B) = 1 \Leftarrow B \subset A \\ &P(A|B) = 0 \Leftarrow A \cap B = \varnothing \end{aligned}$$

Definition 1.2 (Independent events)**.** Let A and B be two events of the same event space $\mathcal{E}$, then we say these events are independent if

$$P(A|B) = P(A) \Rightarrow P(A \cap B) = P(A)P(B), \tag{5}$$

$$P(A), P(B) \neq 0 \tag{6}$$

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

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*

$$P(A) = \sum_i P(A|B_i)P(B_i) \tag{7}$$

Theorem 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*

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A)} = \tag{8}$$

$$P(A|B_i)\frac{P(B_i)}{\sum_j P(A|B_j)P(B_j)} \tag{9}$$

Definition 1.3. (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 1.4. (Discrete random variable). A random variable X is *discrete* if there exists a finite or numerable set $S \subset \mathbb{R}$ such that

$$P\{X \in S\} = 1 \tag{10}$$

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

Definition 1.5. (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

$$p(k) = \begin{cases} p & k = 1 \\ q & k = 0 \end{cases} \tag{11}$$

Its expected value and variance are

$$\mu = \mathrm{E}[k] = p, \tag{12}$$

$$\sigma^2 = \mathrm{Var}[k] = p(p - 1) \tag{13}$$

Definition 1.6. (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

$$p(k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad \forall k \in S \tag{14}$$

Its expected value and variance are

$$\mu = \mathrm{E}[k] = np \qquad \sigma^2 = \mathrm{Var}[k] = npq \tag{15}$$

Definition 1.7. (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

$$p(x_i) = \frac{1}{n} \qquad \forall x_i \in S \tag{16}$$

In this case we say $X \sim Unif(\{x_1, \dots, x_n\})$.
Its expected value, variance and bias are

$$\mu = \mathrm{E}[X] = \frac{n+1}{2} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{n^2-1}{12} \tag{17}$$

Definition 1.8. (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

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \forall k \in S \tag{18}$$

In this case $X \sim Poiss(\lambda)$.
Its expected value, variance and bias are

$$\mu = \mathrm{E}[k] = \lambda \qquad \sigma^2 = \mathrm{Var}[k] = \lambda \qquad \mu c_3 = \sqrt{\lambda} \tag{19}$$

Definition 1.9. (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

$$p(k) = P\{X = k\} = \frac{\binom{N}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad \forall k \in S \tag{20}$$

Its expected value and variance are

$$\mu = \mathrm{E}[X] = n\frac{K}{N} \tag{21}$$

$$\sigma^2 = \mathrm{Var}[X] = \frac{nK}{N}\left(1 - \frac{K}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{22}$$

Definition 1.10. (Absolutely continuous random variable). We say X is an absolutely continuous random variable if there exists a probability density function (PDF) f such that

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(y)\,\mathrm{d}y \qquad \forall x \in \mathbb{R} \tag{23}$$

Definition 1.11. (Probability density function). Analogous to distribution function, we say a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a PDF if

$$f(x) \geq 0 \qquad \forall x \in \mathbb{R} \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \tag{24}$$

Definition 1.12. (Cumulative probability function). Analogous to probability function, we say a function $F : \mathbb{R} \longrightarrow \mathbb{R}$ is a CDF if

$$F(x) = P\{X \leq x\} \qquad \text{and} \qquad f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} \tag{25}$$

Definition 1.13. (Normal or Gaussian distribution). Let X be an absolutely continuous random variable, we say it follows a *normal distribution* if its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{26}$$

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

$$\mu = \mathrm{E}[k] \qquad \sigma^2 = \mathrm{Var}[k] \qquad \Gamma_{FWHM} = 2\sigma\sqrt{2\ln 2} \tag{27}$$

Definition 1.14. (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

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x) \tag{28}$$

In this case $X \sim Unif([a, b])$.
Its expected value and variance are

$$\mu = \mathrm{E}[X] = \frac{b+a}{2} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{(b-a)^2}{12} \tag{29}$$

Definition 1.15. (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

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \qquad \forall x \in (0, \infty) \tag{30}$$

In this case $X \sim \mathscr{E}xp(\lambda)$.
Its expected value and variance are

$$\mu = \mathrm{E}[X] = \lambda \qquad \sigma^2 = \mathrm{Var}[X] = \lambda^2 \tag{31}$$

Definition 1.16. (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

$$f(x) = \frac{\mu^k}{\Gamma(k)} x^{k-1} e^{-\mu x} \tag{32}$$

In this case $X \sim Gamma(\mu, k)$.
Its expected value and variance are

$$\mu = \mathrm{E}[X] = \frac{k}{\mu} \qquad \sigma^2 = \mathrm{Var}[X] = \frac{k}{\mu^2} \tag{33}$$

Definition 1.17. (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

$$f(x; n) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \qquad x > 0 \tag{34}$$

In this case $X \sim \chi^2(n)$.
Its expected value, variance and Γ_{FWHM} are

$$\mu = \mathrm{E}[X] = n \qquad \sigma^2 = \mathrm{Var}[X] = 2n \tag{35}$$

Definition 1.18. (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

$$f(t;n-1) \equiv f(t;r) = \frac{\Gamma\left(\frac{1}{2}(r+1)\right)}{\sqrt{r\pi}\Gamma\left(\frac{1}{2}r\right)}\left(1+\frac{t^2}{r}\right)^{-(r+1)/2} \tag{36}$$

In this case $X \sim$ t-Student(n).
Its expected value and variance are

$$\mu = \mathrm{E}[X] = 0 \qquad \sigma^2 = \mathrm{Var}[X] = \frac{r}{r-2} \tag{37}$$

Definition 1.19. (Cauchy [Lorentz] distribution). Let X be an absolutely continuous random variable, we say it follows a *Cauchy distribution* if its PDF is

$$f(x) = \frac{1}{\pi}\frac{\frac{1}{2}\Gamma}{(x-m)^2+(\frac{1}{2}\Gamma)^2} \tag{38}$$

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

$$F(x) = \frac{1}{2}+\frac{1}{\pi}\tan^-1\left(\frac{2(x-m)}{\Gamma}\right) \tag{39}$$

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

Definition 1.20. (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

$$f(x) = \frac{1}{\pi(1+x)^2} \tag{40}$$

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

Definition 1.21. (Landau distribution). Let X be an absolutely continuous random variable, we say it follows a *Landau distribution* if its PDF is

$$f(x) = \frac{1}{\pi}\int_0^\infty e^{-t(\ln t + xt)}\sin(\pi t)dt \approx \frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}(x+e^{-x})} \tag{41}$$

In this case we write $X \sim$ Landau.
Its expected value and variance aren't defined as integral diverges.

Theorem 1.10. *(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*

$$f_Y(y) = f_X(h^{-1}(y))\ |(h^{-1})'(y)| 1_{\mathcal{V}}(y) = \tag{42}$$

$$f_X(h^{-1}(y))\left|J\left(\frac{x}{y}\right)\right| \tag{43}$$

Definition 1.22. (Expected value). Let X be a discrete random variable such that

$$\sum_{k\in S}|k|P\{X=k\} < \infty \tag{44}$$

then we define the expected value of X as

$$\mathrm{E}[X] = \sum_{k\in S} kP\{X=k\} \tag{45}$$

Let X be an absolutely random variable such that

$$\int_{-\infty}^{\infty} |x|f(x) < \infty \tag{46}$$

then we define the expected value of X as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x) \tag{47}$$

Proposition 1.11. *(Main properties of expected value). The expected value of a variable X is linear, namely*

$$\mathrm{E}[X+Y] = \mathrm{E}[X]+\mathrm{E}[Y] \qquad \mathrm{E}[aX] = a\mathrm{E}[X] \tag{48}$$

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

$$E[a] = a \qquad \forall a \in \mathbb{R} \tag{49}$$

Proposition 1.12. *(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*

$$\mathrm{E}[z(X)] = \sum_{k\in S} z(k)P(X=k), \tag{50}$$

$$E[z(X)] = \int_{-\infty}^{\infty} z(x)f(x)\,\mathrm{d}x \tag{51}$$

Proposition 1.13. *(Transformation of random variables regarding expected value). Let $f(x)$ be a density function of X, let $z(x)$ be another function, then*

$$\mathrm{E}[z(x)] = \mathrm{E}[z(x(y))] \tag{52}$$

Definition 1.23. (Median). Let X be a discrete random variable, then we define the median x_m of X as the value with probability

$$P\{X \geq x_m\} = P\{X \leq x_m\} \geq \frac{1}{2} \tag{53}$$

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

$$F(x_m) = \int_{-\infty}^{x_m} f(x)\,\mathrm{d}x = \frac{1}{2} \tag{54}$$

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

Definition 1.24. (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

$$\left.\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{X=x_M} \tag{55}$$

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

Definition 1.25. (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

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X^2])] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 \tag{56}$$

Moreover, we define the standard deviation of X as

$$\mathrm{Sd}[X] = \sqrt{\mathrm{Var}[X]} \tag{57}$$

Proposition 1.14. *(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 1.15. *(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*

$$\mathrm{Var}\left[\sum_{i=1}^{N} a_i X_i\right] = \sum_{i=1}^{N} a_i^2 \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{CoV}[X_i, X_j] \tag{58}$$

Definition 1.26. (Reduced variable). Let X be a random variable, then the *reduced (normal) variable* of X is defined as

$$u = \frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}} \tag{59}$$

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

$$\mathrm{E}[u] = 0 \qquad \mathrm{Var}[u] = 1 \tag{60}$$

Definition 1.27. (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

$$\mathrm{CoV}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y]] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] \tag{61}$$

Proposition 1.16. *(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 1.28. (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

$$\rho_{XY} = \frac{\mathrm{CoV}[X, Y]}{\mathrm{Sd}[X]\mathrm{Sd}[Y]} \tag{62}$$

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

Proposition 1.17. *Let X and Y be independent random variables such that $E[|X|], E[|Y|], E[|XY|] < \infty$, then*

$$\mathrm{CoV}[X, Y] = 0 \tag{63}$$

and, consequently, they are non-correlated

Definition 1.29. (Probability density function). We say a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a PDF if

$$f(x, y) \geq 0 \qquad \forall x \in \mathbb{R} \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1 \tag{64}$$

Definition 1.30. (Cumulative probability function). We say a function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a CDF if

$$F(x, y) = P\{X \leq x, Y \leq y\} \qquad \text{and} \qquad f(x, y) = \frac{\mathrm{d}F(x)^2}{\mathrm{d}xy} \tag{65}$$

Definition 1.31. (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

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y \\ h(y) &= \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x \end{aligned}$$

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

Definition 1.32. (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

$$f(y|x) = \frac{f(x, y)}{g(x)} \tag{66}$$

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

Definition 1.33. (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

$$f(x, y) = g(x)h(y) \iff f(y|x) = h(y) \qquad \text{and} \qquad f(x|y) = g(x) \tag{67}$$

Proposition 1.18. *(Multivariable generalization). Let* $\mathbf{X} = (X_1, \dots, X_n)$ *absolutely continuous random variables following a distribution* $f(\mathbf{x})$*, then its marginal distributions and expected values are*

$$\begin{gathered}\mathbf{X} = (X_1, X_2, \dots, X_n) \\ g_i(x_i) \int f(\mathbf{x}) \prod_{j \neq i} \mathrm{d}x_j \\ \mathrm{E}[X_i] = \int_{-\infty}^{\infty} x_i f(\mathbf{x})\, \mathrm{d}\mathbf{x}\end{gathered}$$

The covariance can be represented with a matrix whose elements are

$$\mathrm{C}_{ij} = \mathrm{CoV}[X_i, X_j] = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j]] \quad (68)$$

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

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

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

$$\mathrm{Var}[z] = \sum_{i,j=1}^{n} \mathrm{C}_{ij} \left[\frac{\partial z}{\partial x_i} \frac{\partial z}{\partial x_j} \right]_{\mathbf{x}=\mathbf{E}[X]} \quad (69)$$

A non-linear transformation of covariance matrix is

$$\mathcal{C}_y = \mathcal{T} \mathcal{C}_x \mathcal{T}^{\mathrm{T}} \qquad \mathcal{T} = \left. \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \right|_{\mathbf{X}=\mathbf{E}[X]} \quad (70)$$

Theorem 1.19. *(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*

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{n \to \infty} N(0, \sigma^2) \quad (71)$$

Definition 1.34. 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

$$\underline{x} = \{x_1, x_2, \dots, x_n\} \quad (72)$$

Definition 1.35. (Sample mean). Let $\{x_1, x_2, \dots, x_n\}$ be a sample of an experiment X, then we define the mean of the sample as

$$\hat{\mu} = \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (73)$$

Definition 1.36. (Sample median). Let $\{x_1, x_2, \dots, x_n\}$ be an even sample of an experiment X, then we define the median of the sample as

$$x_m = x_{(n+1)}/2 \quad (74)$$

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

$$x_m = \frac{x_{n/2} + x_{(n/2)+1}}{2} \quad (75)$$

Definition 1.37. (Sample mode). Let $\{x_1, x_2, \dots, x_n\}$ be a sample of an experiment X, then we define the mode of the sample to the most repeated value

$$x_M = \text{most repeated } x_i \text{ value} \quad (76)$$

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

Proposition 1.20. *(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* $\{x_1, x_2, \dots, x_n\}$ *of* X*, then*

$$\mathrm{E}[\overline{x}] = \mu \qquad \mathrm{Var}[\overline{x}] = \frac{\sigma^2}{n} \quad (77)$$

Definition 1.38. (Sample variance). Let $\{x_1, x_2, \dots, x_n\}$ be a sample. Then we call

$$\widehat{\mathrm{Var}}[x] = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad (78)$$

$$\widehat{\mathrm{Var}}[x] = S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \quad (79)$$

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 1.21. *(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* $\{x_1, x_2, \dots, x_n\}$ *of* X*, then*

$$\mathrm{E}[S^2] = \mathrm{Var}[X], \qquad \mathrm{Var}[s^2] = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right), \quad (80)$$

$$Var[S^2] = \frac{2\sigma^4}{n} \ (Gaussian) \quad (81)$$

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

Definition 1.39. (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 Statistical inference

Definition 2.1. (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 2.2. (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

$$P\{a < X \leq b\} = 1 - \alpha \tag{82}$$

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

$$P\{X \leq b\} = 1 - \alpha \qquad \text{and} \qquad P\{X \geq a\} = 1 - \alpha \tag{83}$$

respectively

Definition 2.3. (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 $\{x_1, \dots, x_n\}$ as

$$T = T(x_1, \dots x_n) \tag{84}$$

3 Parameters estimation

Definition 3.1. (Statistical model). A statistical model is a family of probability distributions.

Definition 3.2. (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 $(\theta_1, \dots, \theta_d)$.
We say an statistical model is *regular* if it can be differentiated under the integral sign with respect to θ three times.

Definition 3.3. (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 3.4. (Inference problem). Given a sample $\underline{x} = \{x_1, x_2, \dots, x_n\}$ 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).

Definition 3.5. (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

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

Definition 3.6. (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 $ke = ku - 3$.

Definition 3.7. (Likelihood function). Given a sample $\underline{x} = \{x_1, x_2, \dots, x_n\}$ of a model $f(x;\theta)$ (discrete or continuous) with $\theta = (\theta_1, \dots, \theta_d)$, we define the likelihood function as

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

Definition 3.8. (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 3.9. (Maximum likelihood method). Given a sample $\underline{x} = \{x_1, x_2, \dots, x_n\}$ 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 3.10. (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 3.1. *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.*

4 Comparison and evaluation of estimators

Definition 4.1. (Bias of an estimator). Let T be an estimator of θ, then the bias of T is

$$\mathrm{b}_\theta(T) = \mathrm{E}_\theta[T] - \theta \qquad \forall \theta \tag{85}$$

Definition 4.2. (Non-biased estimator). Let T be an estimator of θ, then we will say it is a non-biased estimator if

$$\mathrm{b}_\theta(T) = 0 \iff \mathrm{E}_\theta[T] = \theta \qquad \forall \theta \tag{86}$$

Definition 4.3. (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

$$T_n \xrightarrow{p} \theta \iff \lim_{n \to \infty} \mathrm{Var}[T_n] = 0 \tag{87}$$

Definition 4.4. (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

$$\mathrm{E}_\theta[T_n] \xrightarrow{n\to\infty} \theta \tag{88}$$

Proposition 4.1. *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 4.5. (Mean squared error). Let T be an estimator of θ parameter, then the mean squared error is

$$\mathrm{MSE}(T) = \mathrm{E}_\theta[(T-\theta)^2] \tag{89}$$

Definition 4.6. (Effectiveness). Let T be an estimator of θ parameter, then the efficiency is

$$\mathrm{eff}(T) = \frac{1}{\mathrm{MSE}(T)} \tag{90}$$

Definition 4.7. (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

$$\mathrm{MSE}(T_1) < \mathrm{MSE}(T_2) \tag{91}$$

Proposition 4.2. *Let T be an estimator of θ, then*

$$\mathrm{MSE}(T) = b^2(T) + \mathrm{Var}_\theta[T] \tag{92}$$

Corollary 4.3. *Let T be a non-biased estimator of θ. Then*

$$\mathrm{MSE}(T) = \mathrm{Var}_\theta[T] \tag{93}$$

Definition 4.8. (Observed and expected Fisher's Information). Let $J(\theta;\underline{x})$ be Fisher's information for a n-dimensional sample...

$$I(\theta) = \mathrm{Var}_\theta[S] = \mathrm{E}_\theta[J(\theta,\underline{x})] = \tag{94}$$

$$E\left[\left(\frac{\partial}{\partial\theta}\sum_i \ln\left(f(x_i;\theta)\right)\right)^2\right] = \mathrm{E}[l'^2] \tag{95}$$

Theorem 4.4. *(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*

$$\mathrm{Var}_\theta[T] \geq \frac{(E'_\theta[T])^2}{I(\theta)} = \frac{\left(1+\frac{\partial b}{\partial\theta}\right)^2}{I(\theta)} \tag{96}$$

Corollary 4.5. *In case T is a non-biased estimator of θ, Cramér-Rao inequality is reduced to*

$$\mathrm{Var}_\theta[T] \geq \frac{1}{I(\theta)} \tag{97}$$

Definition 4.9. We will say an estimator T is efficient if it reaches Cramér-Rao bound.

Definition 4.10. (Efficiency). Let T be an estimator of θ parameter, then the efficiency is

$$\epsilon(T) = \frac{\mathrm{Var}_{CRF,\theta}[T]}{\mathrm{Var}_\theta[T]} \tag{98}$$

Proposition 4.6. *Let T an efficient and non-biased estimator of θ. Then T is the θ parameter estimator with least mean squared error.*