Fundamental of Probability and Statistics

Master 1 : E3A, CSN, VIA
MAT7421 - MAT7001

December 31, 2024
Jean-Pierre Delmas
jean-pierre.delmas@it-sudparis.eu
http://www-public.imtbs-tsp.eu/∼delmas/

Contents

Preface

This handout is not a probability course but a collection of problems and exercises. It is conventionally divided into 7 chapters. In each chapter, there is a list of exercises and homeworks. They are preceded by a set of important results (formulas and theorems) which are necessary for their solving.

Recommended references

- Pierre Bremaud, *An introduction to probabilistic modeling*, Edition 1988, Springer Verlag.
- Athanasios Papoulis and S. Unnikrishna Pillai, *Probability, random variables and stochastic processes*, Fourth edition, MacGraw-Hill, 2002.
- Jean-Pierre Delmas, *Introduction aux Probabilités, Applications aux télécommunications avec exercices et problèmes commentés*, Ellipses, Paris, France, 2000.

Chapter 1

Random experiment, axioms of probability, general theorems

1.1 Events and axioms of probability and main theorems

- Random experiment $\Leftrightarrow (\Omega, \mathcal{A}, P)$ where
  - Ω denotes the sample space whose members ω are called outcomes
  - $\mathcal{A}$ denotes a collection of subset (called events) of Ω (structured as a σ field),
  - P is a mapping: $A \in \mathcal{A} \mapsto$ that satisfies three conditions: (a) $P(A) \in [0,1]$, (b) $P(\Omega) = 1$ and (c) the σ additive axiom for a countable family of events

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n), \quad \text{for} \quad A_i \cap A_j = \emptyset \quad \forall i \neq j$$

- A σ field $\mathcal{A}$ is a collection of subsets of Ω that satisfies three conditions: (a) $\Omega \in \mathcal{A}$, (b) if $A \in \mathcal{A}$ then $\bar{A} \in \mathcal{A}$ and (c) if $(A_n)_{n1,2,\ldots} \in \mathcal{A}$ then $\bigcup_{n=1}^{\infty} \in \mathcal{A}$.

$$\lim\ \sup_n A_n \stackrel{\text{def}}{=} \bigcap_{n=1}^{\infty} \left(\bigcup_{m=n}^{\infty} A_m \right) = \{\omega;\ \text{such that } \omega \text{ are in infinitely many of } A_n\}$$

$$\lim\ \inf_n A_n \stackrel{\text{def}}{=} \bigcup_{n=1}^{\infty} \left(\bigcap_{m=n}^{\infty} A_m \right) = \{\omega;\ \text{such that } \omega \text{ are in all } A_n \text{ except in a finite number of } A_n\}$$

$\lim_{n\to\infty} A_n$ exists by definition if $\lim\ \sup_n A_n = \lim\ \inf_n A_n$.

- The probability satisfies many properties, e.g.,

$$P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

$$A \subset B \quad \Rightarrow \quad P(A) \leq P(B)$$

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

$$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n), \quad \text{(sub } \sigma \text{ additive relation)}$$

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) \text{ with } \lim_{n\to\infty} A_n \stackrel{\text{def}}{=} \bigcup_{n=1}^{\infty} A_n, \quad \text{if } A_1 \subset ... \subset A_n \subset A_{n+1}$$

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) \text{ with } \lim_{n\to\infty} A_n \stackrel{\text{def}}{=} \bigcap_{n=1}^{\infty} A_n, \quad \text{if } A_{n+1} \subset A_n ... \subset A_1$$

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n), \quad \text{if } \lim_{n\to\infty} A_n \text{ exists (continuity theorem of probability)}$$

- Poincaré formula

$$\begin{aligned} P(\cup_{k=1}^{n} A_k) &= \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2})... \\ &+ (-1)^{k+1} \sum_{1\leq i_1 < i_2 < ... < i_k \leq n} P(A_{i_1} \cap A_{i_2} .. \cap A_{i_k}) ... + (-1)^{n+1} P(\cap_{k=1}^{n} A_k) \end{aligned}$$

1.2 Conditional probability and independence

- The conditional probability P_B is a mapping from $\mathcal{A}$ to $[0,1]$ which is a probability set function

$$P(A/B) \stackrel{\text{def}}{=} P_B(A) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) \neq 0$$

- The conditional probability P_B satisfies all the properties summarized above, e.g.,

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

- Multiplication rule

$$P(\cap_{k=1}^{n} A_k) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)......P(A_n/A_1 \cap ... \cap A_{n-1}),$$

- Law of total probability, if $(C_k)_{k=1,..,n}$, or $(C_k)_{k=1,..}$ forms a partition of Ω

$$P(A) = \sum_{k=1}^{n} P(C_k)P(A/C_k), \qquad P(A) = \sum_{k=1}^{\infty} P(C_k)P(A/C_k),$$

- Bayes relation

$$P(C/A) = \frac{P(C)}{P(A)} P(A/C)$$

- Bayes theorem = Bayes relation + Law of total probability (if $(C_k)_{k=1,..,n}$, or $(C_k)_{k=1,..}$ forms a partition of Ω)

$$P(C_i/A) = \frac{P(C_i)P(A \cap C_i)}{\sum_{k=1}^{n} P(C_k)P(A/C_k)}, \qquad P(C_i/A) = \frac{P(C_i)P(A \cap C_i)}{\sum_{k=1}^{\infty} P(C_k)P(A/C_k)},$$

- Two events A and B are independent $\Leftrightarrow P(A \cap B) = P(A)P(B)$.

  If $P(A) \neq 0$, A and B are independent $\Leftrightarrow P_A(B) = P(B)$. If $P(B) \neq 0$, A and B are independent $\Leftrightarrow P_B(A) = P(A)$.

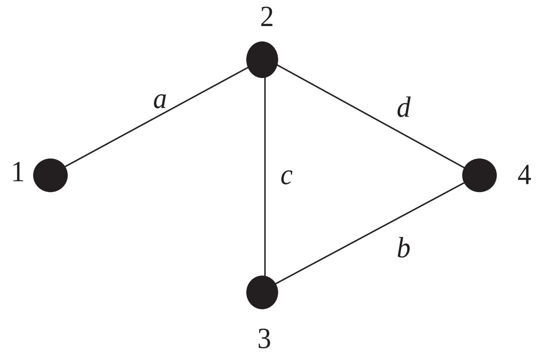
- The events of a family $(A_i)_{i \in I}$ are mutually independent $\Leftrightarrow P(\cap_{k=1}^{n} A_{i_k}) = \prod_{k=1}^{n} P(A_{i_k})$ for all finite subset $(i_1, i_2, ...i_n)$ of I.

  In this case, all combinations of these events and their complement are mutually independent, e.g., $(A_i)_{i=1,2,3,4,5}$ are mutually independent $\Rightarrow A_1 \cup \bar{A}_2$, $\bar{A}_3$ and $A_4 \cap \bar{A}_5$ are mutually independent.

1.3 Exercises

Exercise 1.1 Probability space, disjoint events, independent events

A communication network with four terminals 1, 2, 3, 4 are connected with four links a, b, c, d as shown in the figure. Not all links, however, are necessarily available. Let p denote the probability that any particular link is available and assume that the availability of each link is *independent* of the state of all others links. Two terminals can communicate if and only if they are connected by at least one chain of available links.



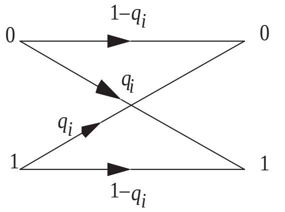
1] Construct an appropriate probability model with 16 sample points, on each state of the system. Specify Ω, the family of events $\mathcal{A}$ and the probability $P(.)$.
2] Let $A \stackrel{\text{def}}{=} \{\omega :$ 1 and 4 can communicate$\}$ and let $B \stackrel{\text{def}}{=} \{\omega :$ 2 and 3 can communicate$\}$. After carefully describing the events A and B, calculate $P(A)$, $P(B)$ and $P(A \cap B)$. Are the events A and B independent?
3] Show that $P(A) = pP(A/c$ available$) + (1-p)P(A/c$ not available$)$. Using this formula, re-evaluate $P(A)$ by inspection.
4] Prove that $P(A)$ would be increased if link c were connected between 1 and 3 rather than between 2 and 3.

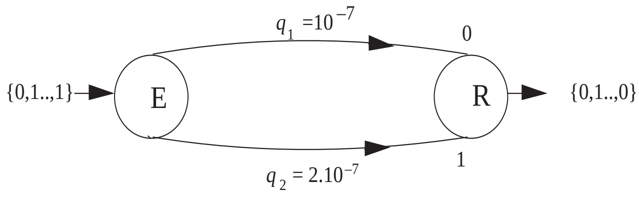
Exercise 1.2 Bayes' theorem, introduction to detection

Consider a binary channels linking a transmitter X to a receiver Y with $\pi_0 = P(X = 0)$ and $\pi_1 = 1 - \pi_0 = P(X = 1)$. This channel can make errors and we donote by $p_0 = P(Y = 1/X = 0)$ and $p_1 = P(Y = 0/X = 1)$.

1] Deduce the probability of error $P(X \neq Y)$ as a function of π_0, π_1, p_0, and p_1.

2] Consider now two independent binary symmetric (i.e., with $p_0 = p_1 \stackrel{\text{def}}{=} q$) channels with respective error probability q_1 and q_2 linking a transmitter and a receiver. The a priori probability at the transmitter are $P(0) = \pi_0 = 0.3$ and $P(1) = \pi_1 = 0.7$. We suppose that $q_1 = 10^{-7}$ and $q_2 = 2.10^{-7}$.





For each transmitted bit, the receiver receives two bits, one of the channel 1 and the other from the channel 2. Therefore, it receives one of the four couples (00), (01), (10) and (11) and must deliver a detected bit to the user of the transmission. Suppose that the receiver makes its decision along the following rule. For each received couple (Y_1, Y_2), it compares the probabilities $P(0 \text{ transmitted}/(Y_1, Y_2) \text{ received})$ and $P(1 \text{ transmitted}/(Y_1, Y_2) \text{ received})$ and decides to choose the most likely transmitted bit. Specify the decision rule. Begin by analyzing the case (01) received.

3] Calculate directly or by the law of total probability the probability of errors of such a receiver.

Exercise 1.3 Repetition code, elementary feedback scheme

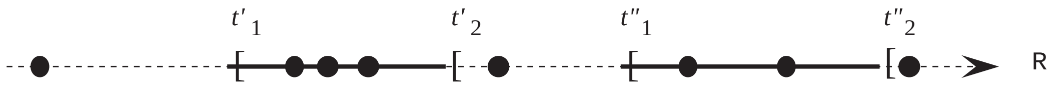
Equilikely symbols 0 and 1 are transmitted along the following scheme. Each symbol is transmitted twice. These symbols are transmitted along an independent binary symmetric channel with error probability p. If the received symbols are equal, the receiver makes the decision and decides in favor of the common symbol. But if the received bits are different, the receiver uses a free-error feedback channel to inform the transmitter that no decision was possible on the previous symbol and to retransmit the symbol twice. The process keeps on until two identical symbols are received.

1] Calculate the probability of error of such a feedback scheme. Comment on the result when $p \ll 1$.

2] Let N be the random variable that gives the number of symbols transmitted by useful symbol, $N \in \{2, 4, \ldots, 2k, \ldots.\}$. Compute $P(N = 2k)$ for $k \in \mathbb{N}^*$. Deduce the expectation $E(N)$ and the variance $\text{Var}(N)$. Comment on the result when $p \ll 1$.

Exercise 1.4 Introduction to Poisson process

A point process is called a homogenous Poisson process of intensity λ iff the following conditions are verified:



(a) If the intervals $(t'_1, t'_2]$ and $(t"_1, t"_2]$ are mutually exclusive, the events $\{N_{t'_1,t'_2} = k'\}$ and $\{N_{t"_1,t"_2} = k"\}$ are independent (where $\{N_{t_1,t_2} = k\}$ denotes the event $\{$there are k points in the interval $(t_1, t_2]$ $\}$),

(b) $P(N_{t_0,t_0+t} = 1) = \lambda t + o(t)$ where $o(t)$ is a function of t that satisfies $\lim_{t\to 0} \frac{o(t)}{t} = 0$,

(c) $P(N_{t_0,t_0+t} > 1) = o(t)$.

1] After commenting on hypothesis (a), (b) and (c), prove the following:

$$\begin{aligned}
P(N_{t_0,t_0+t}=0) &= 1-\lambda t+o(t), \qquad \text{for } t>0 &(1.1)\\
P(N_{t_0,t_0+t+\Delta t}=0) &= (1-\lambda\Delta t+o(\Delta t))P(N_{t_0,t_0+t}=0), \qquad \text{for } t>0,\ \Delta t>0 &(1.2)\\
P(N_{t_0,t_0+t+\Delta t}=k) &= (1-\lambda\Delta t+o(\Delta t))P(N_{t_0,t_0+t}=k)\\
&+ (\lambda\Delta t+o(\Delta t))P(N_{t_0,t_0+t}=k-1)+o(\Delta t),\ \text{for } t>0,\ \Delta t>0, k\in\mathbb{N}^*. &(1.3)
\end{aligned}$$

2] Deduce from (1.2) an ordinary differential equation for the function $P(N_{t_0,t_0+t}=0)$ of t. Solve this equation and give the expression of $P(N_{t_0,t_0+t}=0)$. In the same way, deduce from (1.3) an ordinary differential equation for the function $P(N_{t_0,t_0+t}=k)$ of t for $k>0$. Solve this equation and give the expression of $P(N_{t_0,t_0+t}=k)$ for $k>0$ by induction. Prove that

$$P(N_{t_0,t_0+t}=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t} \quad \text{for } k\in\mathbb{N},$$

i.e., the random variable N_{t_0,t_0+t} that represents the number of points in the interval $[t_0,t_0+t)$ is Poisson distributed with parameter λt.

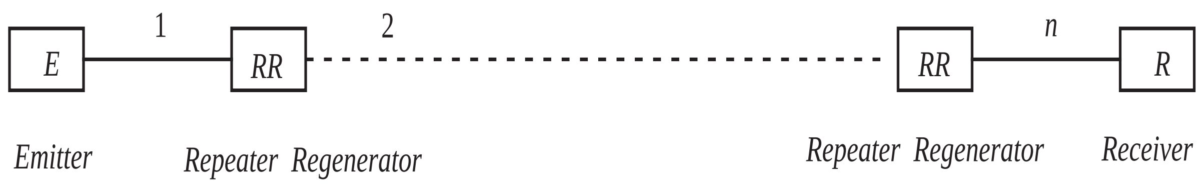
Exercise 1.5 The Monty Hall problem: goats and cars

In a game show, you have to choose one of three doors. One conceals a brand new car, two conceal old goats. You choose a door, say No.1, but your chosen door is not opened immediately. Instead the presenter (who knows which is behind the doors) opens another door, say No.3 which reveals a goat. He offers you the opportunity to change your choice to the third door (unopened and so far unchosen), say No.2. Using conditional probabilities, Is it your advantage to switch your choice?

1.4 Homeworks

Homework 1.1 Disjoint events, independent events

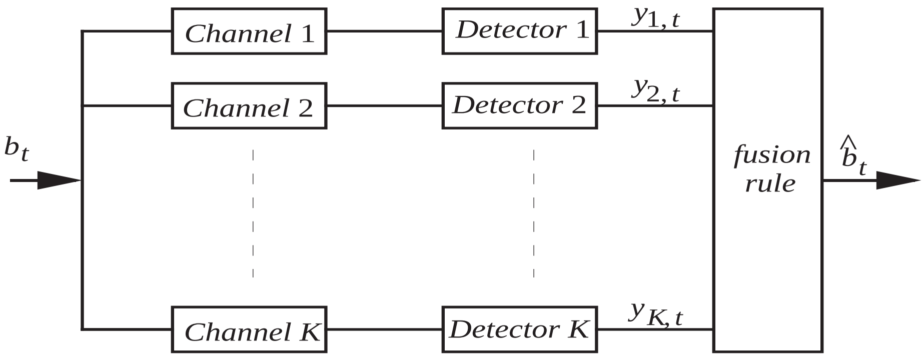
Consider a digital submarine channel. This channel is composed of n independent binary symmetric channels with error probability p, separated by $n-1$ repeaters.



1] Give the error probability $P_n(\text{error})$ of the digital submarine channel as a function of n and p under a sum expression.
2] Express $P_n(\text{error})$ in terms of $P_{n-1}(\text{error})$. Deduce a simple closed-form expression of $P_n(\text{error})$.

Homework 1.2 Fusion of bit estimates

In wireless cellular communications, fusing information from the various base stations can improve link performance. Consider a bank of K detectors, each operating on the same input data observed through individual independent channels. We assume the following: The K detectors make independent decisions, each channel is binary symmetric with respective error probabilities $(p_1, p_1 \dots, p_K)$ and the transmitted bits have equal a priori probability.



1] The receiver makes decisions along to the majority rule, i.e. it counts among the K received bits the respective number of 0 and 1 and decides in favor of the majority bit. Give the probability of error of this receiver $P(\text{error})$ as a function of $(p_1, p_2 \dots, p_K)$. It is useful to distinguish K even and K odd. Finally, consider the particular case $(p_1, p_2 \dots, p_K) = p$.

2] The optimum receiver (the best receiver with respect to the probability of error) is the receiver that compares the a posteriori probabilities $P(0 \text{ transmitted}/y_{1,t}, y_{2,t}, \dots, y_{K,t})$ and $P(1 \text{ transmitted}/y_{1,t}, y_{2,t}, \dots, y_{K,t})$ and decides in favor of that bit whose a posteriori probability is greatest. Prove that criteria is equivalent to compare the probabilities $P(y_{1,t}, y_{2,t}, \dots, y_{K,t}/0 \text{ transmitted})$ and $P(y_{1,t}, y_{2,t}, \dots, y_{K,t}/1 \text{ transmitted})$ (denoted Likelihoods of $(y_{1,t}, y_{2,t}, \dots, y_{K,t})$. Deduce the following optimum fusion rule:

$$\sum_{k=1}^{K} (2y_{k,t} - 1) \ln \frac{1 - p_k}{p_k} \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

3] Comment on this receiver. Prove that this optimum receiver comes down to the majority receiver when the probabilities of error $(p_k)_{k=1,\dots,K}$ are equal.

4] Suppose now that the probabilities of error $(p_1, p_2, ...p_K)$ are unknown. Propose a method to estimate $(p_1, p_2, ...p_K)$ with $K \geq 3$, from the observation of the receiver's symbols only $(y_k, t)_{k=1,2,..,K,t=1,2,..,T}$.

Chapter 2

Random variable

2.1 Definition, probability law

- A real-valued random variable is a mapping from Ω to $\mathbb{R}$ (univariate or scalar) or to $\mathbb{R}^n$ (multivariate or multidimensional): $\omega \in \Omega \mapsto x = X(\omega) \in \mathbb{R}$ or $\omega \in \Omega \mapsto (x_1, ..., x_n) = \mathbf{X}(\omega) = (X_1(\omega), .., X_n(\omega)) \in \mathbb{R}^n$ such that $X^{-1}(B) \in \mathcal{A}$ for any Borel set $\mathbb{R}$, respectively $\mathbb{R}^n$
- Notation $(X = x) \stackrel{\text{def}}{=} \{\omega; X(\omega) = x\}$ and $(X \in B) \stackrel{\text{def}}{=} \{\omega; X(\omega) \in B\}$
- Probability law (denoted also by probability distribution) (denoted P_X) : all mean characterizing $P(X \in B)$ for any Borel set of $\mathbb{R}$, respectively $\mathbb{R}^n$

2.2 Cumulative distribution function

- Univariate (scalar) $F_X(x) \stackrel{\text{def}}{=} P(X \leq x) = P(X \in (-\infty, x])$

  Properties

  - $F_X(x)$ is a nondecreasing function, $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to+\infty} F_X(x) = 1$
  - $F_X(x_0^+) = F_X(x_0)$ (continuous on the right) and $F_X(x_0) = F_X(x_0^-) + P(X = x_0)$ (finite or countable number of discontinuity)
  - $P(X \in (a, b]) = F_X(b) - F_X(a)$

- Multivariate (vector) $F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, .., x_n) \stackrel{\text{def}}{=} P[(X_1 \leq x_1) \cap ... \cap (X_n \leq x_n)]$

  Properties

  - $F_{\mathbf{X}}(x_1, .., x_n)$ is a nondecreasing function with respect to any component x_k
  - $F_{X_k}(x_k) = F_{\mathbf{X}}(+\infty, .. + \infty, x_k, +\infty, .. + \infty)$

- The cumulative distribution function characterizes the probability law for arbitrary random variable.

2.3 Discrete random variable

- $X(\omega) \in \{x_1, .., x_n\}$ (finite case) or $X(\omega) \in \{x_1, .., x_k, ...\}$ (countable case)
- The probability law is characterized by the probability mass function $P(X = x_k)$ $\forall k \in \{1, ..., n\}$ or $\forall k \in \{1, ..., n, ...\}$. The cumulative distribution function is a staircase function and is given by $F_X(x) = \sum_{x_k \leq x} P(X = x_k)$.
- Property
$$P(X = x_i) = \sum_j P[(X, Y) = (x_i, y_j)], \quad \text{(bivariate case)}.$$

2.4 Continuous random variable

- The cumulative distribution function is continuous and is derivable almost everywhere and its derivative is called *probability density function*
$$f_X(x) \stackrel{\text{def}}{=} \frac{dF_X(x)}{dx}, \quad \text{(univariate case)},$$
$$f_{\mathbf{X}}(x_1, ..x_n) \stackrel{\text{def}}{=} \frac{\partial^n F_{\mathbf{X}}(x_1, ..., x_n)}{\partial x_1, .., \partial x_n}, \quad \text{(multivariate case)}$$
- Properties
$$P(X = x_0) = 0, \ \forall x_0, \quad \text{(scalar case)},$$
$$P(X \in B) = \int_B f_X(x) dx, \quad P(X \in (-\infty, x]) = \int_{-\infty}^{x} f_X(u) du, \text{(univariate case)},$$
$$P(\mathbf{X} \in B) = \int_. . \int_B f_{\mathbf{X}}(x_1, ..x_n) dx_1..dx_n, \quad \text{(multivariate case)}$$
$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2) dx_2, \quad f_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2, x_3) dx_3, ...$$

2.5 Generalized law of total probability

- For arbitrary event A and continuous random variable X with probability density function $f_X(x)$, we have:
$$P(A) = \int_{-\infty}^{+\infty} P(A/X = x) f_X(x) dx.$$
In the particular case where A is the event $(Y = y_k)$, where Y is a discrete random variable with $Y \in \{y_k; k \ \in \{1, ..., n\}$ or $k \in \{1, ..., n, ...\}\}$, we get:
$$P(Y = y_k) = \int_{-\infty}^{+\infty} P(Y = y_k/X = x) f_X(x) dx.$$

2.6 Independence of random variables

- X_1 and X_2 are independent $\Leftrightarrow$ the events $(X_1 \in B_1)$ and $(X_2 \in B_2)$ are independent for all Borel sets B_1 and $B_2 \Leftrightarrow P[(X_1 \in B_1) \cap (X_2 \in B_2)] = P(X_1 \in B_1)P(X_2 \in B_2)$ for all Borel sets B_1 and B_2

- Proving that X_1 and X_2 are not independent $\Leftrightarrow$ Finding a specific couple (B_1, B_2) such that $P[(X_1 \in B_1) \cap (X_2 \in B_2)] \neq P(X_1 \in B_1)P(X_2 \in B_2)$

- The random variables of the family $(X_i)_{i\in I}$ are mutually independent $\Leftrightarrow$ the events of the family $(X_i \in B_i)_{i\in I}$ are mutually independent, $\forall B_i$

- The random variables of the family $(X_i)_{i\in I}$ are mutually independent $\Rightarrow$ the random variables of the family $g_i(X_i)_{i\in I}$ are mutually independent for arbitrary (measurable) function g_i, e.g., (X_1, X_2, X_3, X_4) mutually independent $\Rightarrow$ $\sin(X_1)$, $X_2^3 + \cos 2X_3$ and $1/X_4$ mutually independent

- Characterization of independence of X_1 and X_2

  - Discrete case : $P[(X_1 = x_{1,k} \cap (X_2 = x_{2,l})] = P(X_1 = x_{1,k})P(X_2 = x_l)$, $\forall (k,l) \in \mathbb{N}^2$
  - Continuous case : $f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ $\forall(x_1, x_2) \Leftrightarrow F_{\mathbf{X}}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ $\forall(x_1, x_2) \Leftrightarrow f_{\mathbf{X}}(x_1, x_2) = g(x_1)h(x_2)$ $\forall(x_1, x_2)$ for arbitrary functions g and $h \Leftrightarrow F_{\mathbf{X}}(x_1, x_2) = G(x_1)H(x_2)$ $\forall(x_1, x_2)$ for arbitrary functions G and H.

2.7 Change of variables

- Definition : Giving the distribution of a probability on a space $\mathbb{R}^n$ specified by the law of probability of $\mathbf{X}$ and a mapping g: $\mathbf{x} \in \mathbb{R}^n \mapsto \mathbf{y} = g(\mathbf{x}) \in \mathbb{R}^q$, we want to know the probability law of $\mathbf{Y} = g(\mathbf{X})$. The general approach to find this new probability law is to first derive the cumulative distribution function of $\mathbf{Y}$

- If X is discrete $\Rightarrow$ Y is discrete, but If X is continuous $\Rightarrow$ Y may be discrete.

- If g is one to one differentiable and X is continuous, the *smooth change of variable formula* applies:

  - monovariate case:

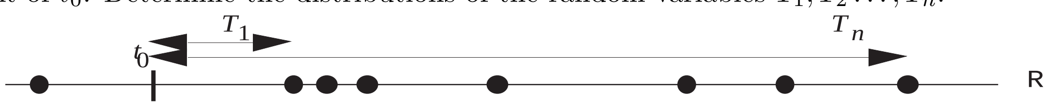
$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x)_{|x=g^{-1}(y)} = \frac{1}{\left|\frac{dy}{dx}\right|} f_X(x)_{|x=g^{-1}(y)}$$

  - bivariate case:

$$\begin{aligned} f_{\mathbf{Y}}(y_1, y_2) &= \left|\det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix}\right| f_{\mathbf{X}}(x_1, x_2)_{|(x_1,x_2)=g^{-1}(y_1,y_2)} \\ &= \frac{1}{\left|\det \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix}\right|} f_{\mathbf{X}}(x_1, x_2)_{|(x_1,x_2)=g^{-1}(y_1,y_2)} \end{aligned}$$

2.8 Exercises

Exercise 2.1 Introduction to Poisson process (continued)
2] From the preceding point process N, we build another point process N' defined as the following. Each point of the process N can by erased with a probability $1-\mu$ (and consequently not erased with probability μ). The erasing are mutually independent. Prove that the new point process is a homogenous Poisson process of intensity $\lambda\mu$. (This problem happens with Geiger counter in the process of emission of particles from a radioactive source in which only a fraction μ of the particles are counted).
3] Prove that if we know that the Poisson process has n points in the interval $(t_0, t_n]$, the instants of the n points can be obtained by independently choosing at random n points in $(t_0, t_n]$. To prove this result, compute the probability of the following events by two different means: the first point to the right of t_0 is located in $(t_0, t_1]$, the second point in the interval $(t_0, t_n]$ is located in $(t_1, t_2]$... and the n-th is located in $(t_{n-1}, t_n]$, conditional on n points in $(t_0, t_n]$. Comment on this result.
4] Prove that for $t_0 < t_1 < t_2$, the random variable N_{t_0,t_1} conditional on $N_{t_0,t_2} = n$ is a random variable of Binomial distribution with parameters n and $\frac{t_1-t_0}{t_2-t_0}$.
5] Let $T_1, T_2 \ldots, T_n$ be the duration from the base time t_0 of the first, second... and n-th point located to the right of t_0. Determine the distributions of the random variables $T_1, T_2 \ldots, T_n$.



6] Let $D_k \stackrel{\text{def}}{=} T_{k+1} - T_k$ be the random variables representing the time between two consecutive points of the Poisson process. We want to derive the probability density function of $(T_1, D_1, ..., D_n)$. For that a solution consists to derive the cumulative distribution function $F_{T_1,T_2,...,T_n,T_{n+1}}(t_1, t_2, ...t_n, t_{n+1})$ of the random variable $(T_1, T_2, ..., T_n, T_{n+1})$. For that fix points $(s_1, s_2, ...s_n)$ such that

$$0 < s_1 < t_1 < s_2 < ... < s_k < t_k < s_{k+1} < ... < s_n < t_n < t_{n+1}.$$

After giving the probability density function of the random variable $(T_1, T_2 \ldots, T_{n+1})$ prove that the random variables $T_1, D_1, D_2, \ldots, D_n$ are independent and exponentially distributed with parameter λ, i.e. with probability density function $f_T(t) = \lambda e^{-\lambda t} 1_{(0,+\infty)}(t)$.
7] Prove that if the random variable X is exponentially distributed, it satisfies the following property:

$$P(X > a + b/X > a) \qquad \text{do not depend on } a \text{ for all } a \geq 0,\ b \geq 0.$$

Conversely, prove that if a positive continuous random variable satisfies the preceding property, it is exponentially distributed. Comment on this curious property of no memory with reference to questions 6 and 7.

Exercise 2.2 Change of variables, continuous random variables

Consider n mutually independent continuous random variables $(X_k)_{k=1,...,n}$ identically distributed of probability density and cumulative distribution functions $f_X(x)$ and $F_X(x)$. Let I_n and S_n be respectively the minimum and maximum values among $(X_k)_{k=1,...,n}$.
1] Give the expression of the probability density functions $f_{I_n}(x)$ and $f_{S_n}(x)$.
2] Give the cumulative distribution function $F_{I_n,S_n}(x, y)$ of the random variable (I_n, S_n). For that, you

can use the derivation of the probability $P(I_n > x \cap S_n \leq y)$. Deduce the probability density function $f_{I_n,S_n}(x,y)$.
3] Let us consider the random variable range $R_n \stackrel{\text{def}}{=} S_n - I_n$. Give the cumulative distributive function $F_{R_n}(x)$. Deduce the probability density function $f_{R_n}(x)$.
4] Derive in the particular cases a) $(X_k)_{k=1,\ldots,n}$ are uniformly distributed in $(0,1)$, b) $(X_k)_{k=1,\ldots,n}$ are exponentially distributed with parameter λ, the expressions of $f_{S_n}(x)$, $f_{I_n}(x)$, $f_{R_n}(x)$, $F_{S_n}(x)$, $F_{I_n}(x)$ and $F_{5_n}(x)$.
5] Derive in the particular cases, the limit of the distribution of the sequence of random variables $W_n = n(S_n - 1)$ for a) and $G_n = \lambda S_n - \ln n$ for b) Comment on.

Exercise 2.3 Distribution of a sum of independent random variables,
1] Consider two independent continuous random variables X_1 and X_2 with probability density functions $f_{X_1}(x)$ and $f_{X_2}(x)$. Prove that the sum $X = X_1 + X_2$ is a continuous random variable and give its probability density function $f_X(x)$.
2] Consider two independent integer-valued discrete random variables X_1 and X_2 that take values $n \in N$ with respective probabilities $P(X_1 = n)$ and $P(X_2 = n)$ What are the probabilities taken by the sum $X = X_1 + X_2$.
3] Consider two independent random variables X_1 and X_2 either continuous uniform in $[0,1]$ or discrete uniform in $\{1,2,3,4,5,6\}$ (two dices). What are the distribution of the sum $X = X_1 + X_2$?
4] Let X_1 and X_2 be two independent Poisson distributed random variables with parameter λ_1 and λ_2. What is the distribution of the random variable $X = X_1 + X_2$?
5] Let X_1 and X_2 be two independent exponential distributed random variables (i.e, of probability density function $f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty[}$) with parameter λ_1 and λ_2. What is the distribution of the random variable $X = X_1 + X_2$?

Exercise 2.4 Distribution of a sum of independent random variables, mixture of distributions
1] Consider two independent random variables X_1 and X_2. X_1 is continuous with probability density function $f_{X_1}(x)$ and X_2 is discrete and takes the values $(a_n)_{n\in\mathbb{N}}$ ($\mathbb{N}$ is finite or countable) with probabilities $(p_n)_{n\in\mathbb{N}}$. Prove that the sum $X = X_1 + X_2$ is a continuous random variable and give its probability density function $f_X(x)$.
2] Let $(f_{X_n}(x))_{n\in\mathbb{N}}$ be a family of probability density functions and $(C_n)_{n\in\mathbb{N}}$ a set of causes of probabilities $(p_n)_{n\in\mathbb{N}}$. Consider a random variable X that conditional on C_n has the probability density function $f_{X_n}(x)$. Give the probability density function of the random variable X. Give some examples, comment on and compare the probability density functions obtained in questions 1 and 2.

Exercise 2.5 Mixed random variable, definitions

From the continuity property of the probability, i.e.,

$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$ with $\lim_{n\to\infty} A_n \stackrel{\text{def}}{=} \cup_{n=1}^{\infty} A_n$ if $A_1 \subset \ldots \subset A_n \subset A_{n+1}$

$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$ with $\lim_{n\to\infty} A_n \stackrel{\text{def}}{=} \cap_{n=1}^{\infty} A_n$ if $A_{n+1} \subset A_n \ldots \subset A_1$

prove the following properties of the cumulative distribution function $F_X(x) \stackrel{\text{def}}{=} P(X \leq x)$ of a real-valued random variable X for any x_0:
1] $F_X(x_0^+) = F_X(x_0)$

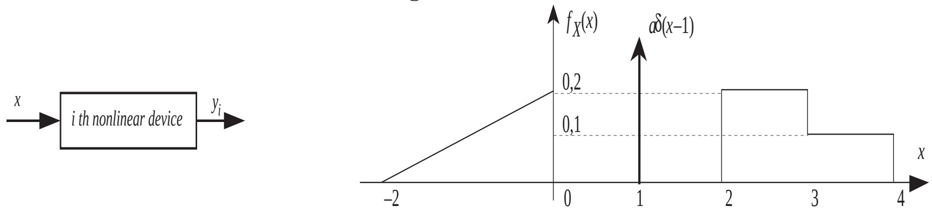
2] $F_X(x_0) = F_X(x_0^-) + P(X = x_0)$[1].
If the random variable is neither continuous, nor discrete, il is called mixed. In practice, the cumulative distribution function can be decomposed into:

$$F_X(x) = F_{X_c}(x) + F_{X_d}(x)$$

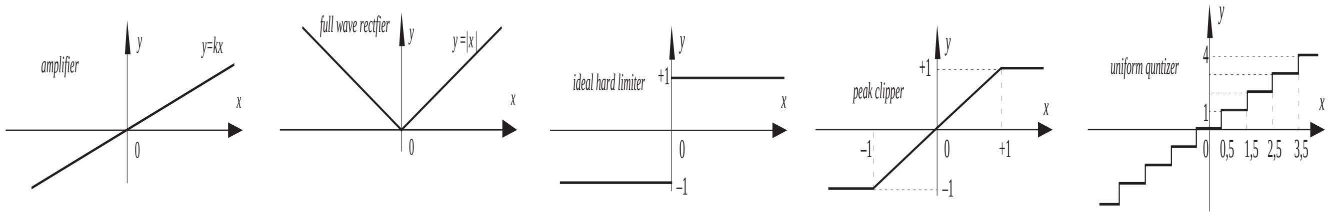
where $F_{X_c}(x)$ is continuous and $F_{X_d}(x) = \sum_{x_k \le x} P(X = x_k)$ with x_k are the discontinuity points of $F_X(x)$. $F_{X_c}(x)$ is derivable almost everywhere with derivative $f_{X_c}(x) \stackrel{\text{def}}{=} \frac{dF_{X_c}(x)}{dx}$. In the engineering literature, a derivative $f_{X_d}(x)$ of $F_{X_d}(x)$ is sometimes defined by $f_{X_d}(x) \stackrel{\text{def}}{=} \sum_{x_k} P(X = x_k)\delta(x - x_k)$. In this case a derivative of $F_X(x)$ is defined by $f_X(x) \stackrel{\text{def}}{=} f_{X_c}(x) + f_{X_d}(x)$.
3] Make a figure of $F_X(x)$ and $f_X(x)$ in the particular case of a single point x_0 of discontinuity of $F_X(x)$. Consider an elementary queueing process in which the probability of waiting is p and the distribution of waiting time is exponential with parameter λ. Give the probability density $f_X(x)$ and the cumulative distribution function $F_X(x)$ of the random variable X giving the waiting time. For $p = 0.1$ and $\lambda = 1s^{-1}$, specify the numerical value of $P(X \le 1\text{s})$.

Exercise 2.6 Change of variables, mixed random variables

A random variable X with probability density function $f_X(x)$ is applied at the input of each of the five nonlinear devices illustrated in the following.



1] Calculate a and plot the cumulative distribution function $F_X(x)$.



2] Calculate and plot the resulting cumulative distribution function $F_{Y_i}(y)$ and probability density functions $f_{Y_i}(y)$ for $i = 1, \ldots, 5$.

2.9 Homeworks

Homework 2.1 Two overlapping Poisson processes (Exercise 2.1 continued)

Let N_1 and N_2 be two independent Poisson processes with parameter λ_1 and λ_2 respectively, and N' be the number of points of N_2 between any two successive points of N_1. What is the distribution of the integer valued random variable N' (it is useful to use the generalized law of total probability)? Comment on this distribution of probability (e.g., if N_1 and N_2 represent the arrival and departure Poisson processes

[1] $F_X(x_0^-)$ and $F_X(x_0^+)$ denote the limit to the left and to the right, respectively.

at a counter, N' represents the number of departures between two successive arrivals).

Homework 2.2 Geometrical probability

Pick a point M at random inside the unit radius circle C of origine O. Let (X, Y) and (R, Θ) (with $R > 0$ and $\Theta \in [0, 2\pi)$) be the cartesian and polar coordinates of M, respectively.

1] Find the cumulative distribution function of the random variables Θ, R, X and Y.

2] Then, deduce the probability density function of these four random variables.

3] Are Θ and R independent? Are X and Y independent?

Chapter 3

Expectation

3.1 Definition, fundamental theorem of expectation

- For arbitrary random variable $\mathbf{X} = (X_1, .., X_n)$, if the integral on Ω: $\int_\Omega |\mathbf{X}(\omega)| dP < \infty$, the expectation of $\mathbf{X}$ is given by $\mathrm{E}(\mathbf{X}) \stackrel{\text{def}}{=} \int_\Omega \mathbf{X}(\omega) dP$.
- Specific definitions of the expectation

$$\mathrm{E}(X) = \sum_{k=1}^{n} x_k P(X = x_k), \text{ (finite discrete case)}, \mathrm{E}(X) = \sum_{k=1}^{+\infty} x_k P(X = x_k), \quad \text{(countable discrete case)},$$

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx, \quad \text{(continuous case)}$$

$$\mathrm{E}(X) = \sum_k x_k P(X = x_k) + \int_{-\infty}^{+\infty} x f_{X_c}(x) dx, \quad \text{(mixed case)}$$

- Fundamental theorem of expectation for $Y = g(\mathbf{X}) = g(X_1, .., X_n)$

$$\mathrm{E}(Y) = \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} g(x_1, .., x_n) f_{\mathbf{X}}((x_1, .., x_n) dx_1..dx_n, \text{ (for } \mathbf{X} \text{ continuous)}$$

$$\mathrm{E}(Y) = \sum_k g(x_{1,k}, .., x_{n,k}) P(\mathbf{X} = (x_{1,k}, .., x_{n,k})), \text{ (for } \mathbf{X} \text{ discrete)}$$

3.2 Variance, covariance matrix

- Variance

$$\mathrm{var}(X) \stackrel{\text{def}}{=} \mathrm{E}[(X - \mathrm{E}(X))^2] = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

- Covariance

$$\mathrm{cov}(X_1, X_2) \stackrel{\text{def}}{=} \mathrm{E}[(X_1 - \mathrm{E}(X_1))(X_2 - \mathrm{E}(X_2))] = \mathrm{E}(X_1 X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2)$$

- Correlation coefficient of X_1 and X_2 = covariance of the reduced random variables $X_{r_1} \stackrel{\text{def}}{=} \frac{X_1 - m_1}{\sigma_1}$

and $X_{r_2} \stackrel{\text{def}}{=} \frac{X_2 - m_2}{\sigma_2}$

$$\rho_{X_1,X_2} \stackrel{\text{def}}{=} \text{cov}(X_{r_1}, X_{r_2}) = \frac{\text{E}[(X_1 - m_1)(X_2 - m_2)]}{\sigma_1 \sigma_2}$$

$$\rho_{X_1,X_2} \in [-1, +1], \ \rho_{X_1,X_2} = +1 \Leftrightarrow \frac{X_2 - m_2}{\sigma_2} = \frac{X_1 - m_1}{\sigma_1}, \ \rho_{X_1,X_2} = -1 \Leftrightarrow \frac{X_2 - m_2}{\sigma_2} = -\frac{X_1 - m_1}{\sigma_1}$$

- Covariance matrix of $\mathbf{X} = (X_1, ..X_n)$

$$[\mathbf{C}]_{k,l} \stackrel{\text{def}}{=} \text{cov}(X_k, X_l) \ \Rightarrow \mathbf{C} = \text{E}[(\mathbf{X} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T] = \text{E}(\mathbf{X}\mathbf{X}^T) - \mathbf{m}\,\mathbf{m}^T$$

$\mathbf{C}$ is $n \times n$ symmetric, positive ($\mathbf{a}^T\mathbf{C}\mathbf{a} \geq 0, \forall \mathbf{a} \neq \mathbf{0}$) or nonnegative matrix ($\mathbf{a}^T\mathbf{C}\mathbf{a} > 0, \forall \mathbf{a} \neq \mathbf{0}$)

$$\text{var}(X_k) = \mathbf{u}_k^T\mathbf{C}\mathbf{u}_k, \text{with, } \mathbf{u}_k \stackrel{\text{def}}{=} (0, , 0, 1, 0..0)^T, 1 \text{ in position } k$$

$\text{var}(X_u) = \mathbf{u}^T\mathbf{C}\mathbf{u}$, where $X_u \stackrel{\text{def}}{=} \mathbf{u}^T\mathbf{X}$ orthogonal projection of $\mathbf{X}$ on $\mathbf{u}$ with $\mathbf{u}$ arbitrary unit vector

- Properties deduced from the fundamental theorem of expectation

$$\text{E}\left(\left(\sum_{k=1}^{n} a_k X_k\right) + b\right) = \left(\sum_{k=1}^{n} a_k \text{E}(X_k)\right) + b$$

$$\begin{aligned} \text{var}\left(\left(\sum_{k=1}^{n} a_k X_k\right) + b\right) &= \sum_{k=1}^{n} a_k^2 \text{var}(X_k) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \text{cov}(X_i, X_j) \\ &= \mathbf{a}^T\mathbf{C}\mathbf{a} \ \text{ with } \mathbf{a} \stackrel{\text{def}}{=} (a_1, .., a_n)^T \end{aligned}$$

- If the random variables X_1, X_2,.. and X_n are mutually independent
  $\Rightarrow \text{E}[\prod_{k=1}^{n} g_k(X_k)] = \prod_{k=1}^{n} \text{E}(g_k(X_k))$ for arbitrary measurable functions g_k
  $\Rightarrow X_1$, X_2,.. and X_n are uncorrelated, i.e., $\text{cov}(X_i, X_j) = 0, \forall i \neq j$

3.3 Markov and Chebyshev's inequalities

- Markov's inequality

$$P(|X| \geq c) \leq \frac{\text{E}(|X|^r)}{c^r}, \ \ \forall c > 0 \text{ and } \forall r > 0$$

- Chebyshev's inequality = Markov's inequality with $r = 2$ and X replaced by $X - m$

$$P(|X - m| \geq c) \leq \frac{\text{var}(X)}{c^2}, \ \ \forall c > 0 \quad \Longleftrightarrow \quad P(|X - m| < c) \geq 1 - \frac{\text{var}(X)}{c^2}, \ \ \forall c > 0$$

3.4 Characteristic function

- The characteristic function $\phi_{\mathbf{X}}(\mathbf{u}) \stackrel{\text{def}}{=} \int_\Omega e^{\mathbf{u}^T\mathbf{X}(\omega)}dP$ exists for arbitrary random variable $\mathbf{X}$

$$\phi_X(u) = \sum_k e^{iux_k} P(X = x_k), \quad \text{(univariate discrete case)},$$

$$\phi_X(u) = \int_{-\infty}^{+\infty} e^{iux} f_X(x)dx, \quad \text{(univariate continuous case)}$$

$$\phi_{\mathbf{X}}(\mathbf{u}) = \int_{-\infty}^{+\infty} .. \int_{-\infty}^{+\infty} e^{i\mathbf{u}^T\mathbf{x}} f_{\mathbf{X}}(x_1, ..x_n)dx_1...dx_n, \quad \text{(multivariate continuous case)}$$

- Properties

$$X_1, ..X_n \text{ mutually independent } \Rightarrow \phi_Y(u) = \prod_{k=1}^n \phi_{X_k}(u), \text{with } Y \stackrel{\text{def}}{=} \sum_{k=1}^n X_k$$

Expansion of $\phi_{\mathbf{X}}(\mathbf{u})$ around zero $\Rightarrow$ calculus of moments $E(X^k)$:

$$\phi_X(u) = \sum_{k=0}^n \frac{i^k u^k}{k!} E(X^k) + o(u^n), \quad \text{if } E(X^n) \text{ exists}$$

3.5 Exercises

Exercise 3.1 Lognormal distribution

Consider the change of random variable $Y \stackrel{\text{def}}{=} e^X$ where X is normally distributed with mean m and variance σ^2.

1] Find the probability density function of the random variable Y. The random variable is said to have a lognormal distribution (since $\ln Y = X$ has a normal distribution).

2] Specify the mode, the expectation, the median and the variance of the random variable Y.

3] What is the probability density function of the random variable $Z \stackrel{\text{def}}{=} z_0 + Y$? This random variable is said to have a generalized lognormal distribution. Suppose that this distribution models the monthly salaries of the employees of a company. Specify the parameters m, σ and z_0 of this distribution if out of 10000 observed employees, half the employees earn more than \$2000, 1587 earn more than \$3000 and 1587 earn less \$1500.

Exercise 3.2 Mixture of Gaussian random variables

Consider a image satellite whose gray level of each pixel can be modeled by the realization of a random variable X. Each pixel is supposed to be issued from one among two classes: vegetation and water (image of nature in Finland). The proportion of vegetation and water are p_1 and p_2 respectively ($p_1 + p_2 = 1$). Each pixel of the two classes are modeled as Gaussian random variables with mean m_1 and m_2, and variance σ_1^2 and σ_2^2 respectively.

1] Give the expression of the probability density function of an arbitrary pixel X.

2] Give closed form expressions for $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

Exercise 3.3 Continuous bivariate random variable

Let be

$$f_{X,Y}(x,y) = (x+y)\mathbb{1}_{[0,1]^2}(x,y)$$

a probability density function of the random variable (X, Y).

1] Check that $f_{X,Y}(x,y)$ is a probability density function. Derive $f_X(x)$ and $f_Y(y)$.

2] Give the values of $\mathrm{E}(X)$, $\mathrm{Var}(X)$ and $\mathrm{Cov}(X,Y)$.

Exercise 3.4 Monte Carlo method

It is required to estimate $J = \int_0^1 g(x)dx$, where $0 \le g(x) \le 1$ for all $x \in [0,1]$. Let X and Y be independent random variables uniformly distributed in $[0,1]$. Let[1] $U \stackrel{\text{def}}{=} \mathbb{1}_{Y \le g(X)}$, $V \stackrel{\text{def}}{=} g(X)$ and $W \stackrel{\text{def}}{=} \frac{1}{2}[g(X) + g(1-X)]$.

1] Prove that $\mathrm{E}(U) = \mathrm{E}(V) = \mathrm{E}(W) = J$.

2] Prove that $\mathrm{Var}(U) \ge \mathrm{Var}(V) \ge \mathrm{Var}(W)$.

3] Derive a scheme to estimate efficiently J from a sequence of n independent random variables $(X_k)_{k=1,\dots,n}$ uniformly distributed in $[0,1]$.

3.6 Homeworks

Homework 3.1 Fundamental theorem of expectation, optimization

In the mass production of bars, the length of the bars must in fact be considered as a random variable X because of the dispersion of the length around the mean value m. Let $f_X(x)$ be the probability density function of the length of the bars produced. Suppose we want the bars to have exactly the length a. Consequently if the produced length is larger than a, the bar are cut at the length a and the rest $X - a$ is discarded, otherwise if the produced length is smaller than a all the bar is X discarded.

1] Give the expression of the cumulative distribution function $F_Y(y)$ of the length lost by produced bar represented by the random variable Y and derive the associated probability density function $f_Y(y)$. We assume in this question that $f_X(x) = 0$ outside $[0, 2a]$.

2] Give the expression of $\mathrm{E}(Y)$ from $f_Y(y)$ and from the theorem of expectation.

3] Suppose that the distribution of the random variable X is Gaussian with mean m and variance σ^2. What does m and σ mean? Explain why $\frac{2\sigma}{m}$ may mean the relative accuracy (or uncertainty) of this mass production. Comment on this model, in particular prove that $P(X < 0) \ll 1$. Use for this purpose the following upper bound that will be proved:

$$\int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt < \frac{1}{2} e^{-\frac{x^2}{2}}, \qquad \text{for } x > 0$$

4] Give the optimum value of m that minimizes the expectation of the lost length Y. Consider the particular case $a = 2$ meters and $\sigma = 0.02$ meter.

[1] $\mathbb{1}_A \stackrel{\text{def}}{=} 1$ if the event A is satisfied and 0 elsewhere (indicator function of the event A).

Chapter 4

Gaussian distribution

4.1 Univariate Gaussian distribution

- The random variable X is Gaussian (normal) distributed $\mathcal{N}(m,\sigma^2)$ $\Leftrightarrow$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad E(X) = m \text{ and } \mathrm{var}(X) = \sigma^2$$

- X is Gaussian distributed $\mathcal{N}(m,\sigma^2)$ $\Rightarrow$ $Y = aX + b$ is Gaussian distributed $\mathcal{N}(am+b, a^2\sigma^2)$

- Associated normalized (reduced) random variable

$$X_r \overset{\text{def}}{=} \frac{X-m}{\sigma} \quad \Rightarrow \quad f_{X_r}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad E(X_r) = 0 \text{ and } \mathrm{var}(X_r) = 1$$

- Order of magnitude

$$P(X \in [m-2\sigma, m+2\sigma]) \approx 0.95 \quad P(X \in [m-3\sigma, m+3\sigma]) \approx 0.997$$

$$P(X \in [m-6\sigma, m+6\sigma]) \approx 1 - 1.2\ 10^{-8}$$

- Error function $Q(x) \overset{\text{def}}{=} \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$

$$\frac{1}{x\sqrt{2\pi}}\left(1 - \frac{1}{x^2}\right) e^{-\frac{x^2}{2}} \leq Q(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{for } x > 0$$

- Characteristic function

$$\phi_X(u) = e^{imu} e^{-\frac{\sigma^2 u^2}{2}}, \quad \phi_{X_r}(u) = e^{-\frac{u^2}{2}} \quad \Rightarrow \quad \mathrm{E}(X_r^k) = \left(\begin{array}{ll} \frac{(2n)!}{2^n n!} & \text{for } k = 2n \\ 0 & \text{for } k \text{ odd} \end{array} \right.$$

- Derivation of $\mathrm{E}(X^k)$

$$\mathrm{E}(X^k) = \mathrm{E}[(\sigma X_r + m)^k], \text{ with e.g., } \mathrm{E}(X_r^2) = 1, \mathrm{E}(X_r^4) = 3$$

4.2 Multivariate Gaussian distribution

- The random variable $\mathbf{X} = (X_1, ..., X_n)$ is Gaussian distributed $\Leftrightarrow Y = \sum_{k=1}^{n} u_k X_k$ is Gaussian distributed $\forall \mathbf{u} = (u_1, .., u_n) \in \mathbb{R}^n$
- Its distribution is characterized by the expectation $\mathrm{E}(\mathbf{X}) = \mathbf{m}$ and the covariance matrix $\mathrm{cov}(\mathbf{X}) = \mathbf{C}$ and denoted by $\mathcal{N}(\mathbf{m}, \mathbf{C})$
- Properties
  - If $\mathbf{X}$ is $\mathcal{N}(\mathbf{m}, \mathbf{C}) \;\Rightarrow\; \mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is $\mathcal{N}(\mathbf{Am} + \mathbf{b}, \mathbf{ACA}^T)$ where $\mathbf{A}$ and $\mathbf{b}$ are deterministic. In particular the marginal random variables X_k are Gaussian distributed $\mathcal{N}(m_k, \sigma_k^2)$ with m_k is the k-th component of $\mathbf{m}$ and $\sigma_k^2 = [\mathbf{C}]_{k,k}$
  - If X_1, X_2.. X_n are Gaussian distributed, $\mathbf{X} = (X_1, ..., X_n)$ is not necessarily Gaussian distributed

    Ex: Let X be Gaussian distributed $\mathcal{N}(0, 1)$ and $X_1 \stackrel{\text{def}}{=} X$ and $X_2 \stackrel{\text{def}}{=} \left(\begin{array}{ll} X & \text{if } |X| \leq 2 \\ -X & \text{if } |X| > 2 \end{array} \right.$

    $\Rightarrow$ X_1 and X_2 are Gaussian distributed $\mathcal{N}(0, 1)$, but $\mathbf{X} = (X_1, X_2)$ is not Gaussian distributed because $Y \stackrel{\text{def}}{=} X_1 + X_2$ is a mixed random variable ($P(Y = 0) = P(|X| > 2) \approx 0.05 \neq 0$)
  - But if X_1, X_2.. and X_n are Gaussian distributed and mutually independent $\Rightarrow$ $\mathbf{X} = (X_1, ..., X_n)$ is Gaussian distributed
  - If $\mathbf{X} = (X_1, ..., X_n)$ is Gaussian distributed and X_1, X_2.. and X_n mutually uncorrelated $\Rightarrow$ X_1, X_2.. and X_n are mutually independent
  - If the covariance matrix $\mathbf{C}$ is not singular, the probability density function exists

$$f_{\mathbf{X}}(x_1, .., x_n) = \frac{1}{(2\pi)^{n/2} \det^{1/2}(\mathbf{C})} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})}$$

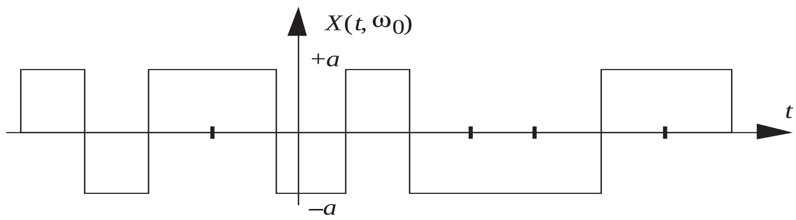
for $n = 2$, $\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ and $\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} \\ \sigma_1\sigma_2\rho_{1,2} & \sigma_2^2 \end{pmatrix} \Rightarrow$

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{1,2}^2}} e^{-\frac{1}{2(1-\rho_{1,2}^2)}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - 2\rho_{1,2}\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)}$$

4.3 Exercises

Exercise 4.1 Baseband data transmission, introduction to probability of error calculus

Consider a sequence of bits 0 and 1 transmitted along a channel with respective probabilities $1 - p$ and p. Suppose that the transmitter maps bits into a voltage $X(t)$, say $0 \rightarrow -a$ and $1 \rightarrow +a$. Suppose that the channel corrupts the transmitted signal by the addition of statistically independent noise voltage which is zero-mean Gaussian distributed with variance σ^2. Consequently, the transmitted and received signals are random and denoted $X(t)$ and $Y(t) = X(t) + N(t)$ respectively, where $N(t)$ denotes the additive noise.



1] Considering the detection of one bit, suppose that the receiver observes the voltage at only on time t_0. It samples $Y(t)$ at time t_0 and compares the value of this random variable $Y(t_0)$ to a threshold s. Its decision rule is the following:

$$\begin{cases} \text{if } Y(t_0) > s & \Rightarrow \text{bit 1 detected} \\ \text{if } Y(t_0) < s & \Rightarrow \text{bit 0 detected} \end{cases}$$

Give the probability of error of such a detector as a function of p, a, σ and s. What is the optimal value of this threshold s (value that minimizes the probability of error)? Examine the particular case $p = \frac{1}{2}$. Show that in this case the probability of error can be expressed as a function of $\frac{a}{\sigma}$ thanks to the error function:

$$Q(x) \stackrel{\text{def}}{=} \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

2] Considering the detection of one bit, suppose that the receiver now observes the voltage at several times $t_1, t_2, \dots t_K$. It samples $Y(t)$ at times $t_1, t_2, \dots t_K$ during the reception of this bit and compares the mean value $Y \stackrel{\text{def}}{=} \frac{1}{K} \sum_{k=1}^{K} Y(t_k)$ to a threshold s and the receiver decision rule is now the following:

$$\begin{cases} \text{if } Y > s & \Rightarrow \text{bit 1 detected} \\ \text{if } Y < s & \Rightarrow \text{bit 0 detected} \end{cases}$$

The random variables $N(t_1), N(t_2), \dots, N(t_K)$ are supposed independent. Prove that the results of question 1] remain valid if σ is replaced by $\frac{\sigma}{\sqrt{K}}$. Examine the particular case $p = \frac{1}{2}$. Compare the expression of the probability of error to those of question 1. What strange result happens when K increases indefinitely.

3] Explain why the hypothesis "the random variables $N(t_1), N(t_2), \dots, N(t_K)$ are supposed independent" if $(t_k)_{k\in(t_0,t_0+T)}$ (i.e. during the observed bit, where $\frac{1}{T}$ bits/s is the bit rate) when K increases without bound is unrealistic. The correct model for random signal $N(t)$ is to consider it as a zero-mean Gaussian random process with spectra (power spectral density) $S_n(f)$ constant, i.e. $S_n(f) = \frac{N_0}{2}$ on a large domain of frequency (typically large with respect to $\frac{1}{T}$). We consider the particular case $p = \frac{1}{2}$. Using

$$\lim_{k\to\infty} \frac{1}{k} \sum_{k=1}^{K} Y(t_k) = \frac{1}{T} \int_{t_0}^{T+t_0} Y(t) dt$$

where $(t_k)_{k=1,\dots,K}$ are regularly spaced in (t_0, t_0+T), give an expression of the probability of error of the receiver that uses all the values of $Y(t)$ during the observed bit.

Exercise 4.2 Correlation coefficient

Consider a two dimensional Gaussian random variable (X_1, X_2) with zero mean, variance σ^2 and correlation coefficient ρ. Let $W_1 \stackrel{\text{def}}{=} \text{sign}(X_1)$ and $W_2 \stackrel{\text{def}}{=} \text{sign}(X_2)$. Give a closed form expression of $\text{E}(W_1 W_2)$.

4.4 Homeworks

Homework 4.1 Independent and uncorrelated random variables.

Consider a two independent zero mean Gaussian random variable X_1 and X_2 with variance σ^2. Let $Y_1 \stackrel{\text{def}}{=} X_1 + X_2$ and $Y_2 \stackrel{\text{def}}{=} X_1 + aX_2$. What are the values of a for which Y_1 and Y_2 are independent?

Chapter 5

Convergence of sequences of random variables

5.1 Convergence of sequences of random variables

- There exists two family of convergence associated with two meanings : closeness of X_n to X and closeness of the distribution P_{X_n} of X_n to the distribution P_X of X when $n \to \infty$.
- Convergence in probability: $X_n \to_{\mathcal{P}} X$

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \iff \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

- Convergence in mean: $X_n \to_{\mathcal{M}} X$

$$\lim_{n\to\infty} \mathrm{E}|X_n - X| = 0$$

- Quadratique convergence: $X_n \to_{\mathcal{Q}} X$

$$\lim_{n\to\infty} \mathrm{E}[(X_n - X)^2] = 0$$

- Almost surely convergence: $X_n \to_{\mathcal{AS}} X$

$$P[\omega; \lim_{n\to\infty} X_n(\omega) = X(\omega)] = 1$$

- Convergence in distribution (in law): $X_n \to_{\mathcal{L}} X$ or $X_n \to_{\mathcal{L}} P_X$

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),\ \forall x \text{ such that } F_X(x) \text{ is continuous}$$

$$\Leftrightarrow \lim_{n\to\infty} P(X_n \in [a,b]) = P(X \in [a,b])\ \forall (a,b) \text{ if } F_X(x) \text{ is continuous}$$

  - Characterization:

    $X_n \to_{\mathcal{L}} X \Leftrightarrow \lim_{n\to\infty} \phi_{X_n}(u) = \phi_X(u)$, $\forall u \in \mathbb{R}$

    $X_n \to_{\mathcal{L}} X \Leftrightarrow \lim_{n\to\infty} \mathrm{E}[g(X_n)] = \mathrm{E}[g(X)]$, $\forall g$ bounded and continuous

- Property: When $F_X(x)$ is continuous, $\lim_{n\to\infty} \text{Sup}_x\{|F_{X_n}(x) - F_{X_n}(x)|\} = 0$.

- There exist the following relations between these convergences:

$$\boxed{\text{Almost surely convergence}} \Rightarrow \boxed{\text{Convergence in probability}} \Rightarrow \boxed{\text{Convergence in distribution}}$$

$$\Uparrow$$

$$\boxed{\text{Quadratique convergence}} \Rightarrow \boxed{\text{Convergence in mean}}$$

5.2 Central limit theorem

There are many central limit theorems (CLT) depending on the assumptions made on the sequence $(X_n)_{n=1,\ldots k..}$. The simplest and most known is the following:

- Let $(X_n)_{n=1,\ldots k..}$ be a sequence of mutually independent identically distributed random variables where $m = \text{E}(X_k)$ and $\sigma^2 = \text{var}(X_k)$ exist, if U_n is the reduced sum which is equal to the reduced mean:

$$U_n = \frac{(\sum_{k=1}^n X_k) - nm}{\sigma\sqrt{n}} = \frac{(\frac{1}{n}\sum_{k=1}^n X_k) - m}{\frac{\sigma}{\sqrt{n}}} \to_{\mathcal{L}} \mathcal{N}(0,1)$$

- Extension to the multivariate case: if $\mathbf{m} = \text{E}(\mathbf{X}_k)$ and $\mathbf{C} = \text{Cov}(\mathbf{X}_k)$ exist

$$\frac{(\sum_{k=1}^n \mathbf{X}_k) - n\mathbf{m}}{\sqrt{n}} = \frac{(\frac{1}{n}\sum_{k=1}^n \mathbf{X}_k) - \mathbf{m}}{\frac{1}{\sqrt{n}}} \to_{\mathcal{L}} \mathcal{N}(\mathbf{0},\mathbf{C})$$

In the particular case where $\mathbf{C}$ is not singular, there exist square roots $\mathbf{\Sigma}$ of $\mathbf{C}$ (i.e., $\mathbf{C} = \mathbf{\Sigma}\mathbf{\Sigma}^T$)

$$\mathbf{\Sigma}^{-1}\left(\frac{(\sum_{k=1}^n \mathbf{X}_k) - n\mathbf{m}}{\sqrt{n}}\right) = \mathbf{\Sigma}^{-1}\left(\frac{(\frac{1}{n}\sum_{k=1}^n \mathbf{X}_k) - \mathbf{m}}{\frac{1}{\sqrt{n}}}\right) \to_{\mathcal{L}} \mathcal{N}(\mathbf{0},\mathbf{I})$$

5.3 Useful theorems

- Theorem of continuity: for all continuous function $x \mapsto g(x)$ such that $P\{x; g(x) \text{ is continuous}\}) = 1$
  - $X_n \to_{\mathcal{P}} X \Rightarrow g(X_n) \to_{\mathcal{P}} g(X)$
  - $X_n \to_{\mathcal{AS}} X \Rightarrow g(X_n) \to_{\mathcal{AS}} g(X)$
  - $X_n \to_{\mathcal{L}} X \Rightarrow g(X_n) \to_{\mathcal{L}} g(X)$

- Slutsky theorem: Let U_n be a sequence which converges in law to U and V_n be a sequence which converges in probability to a constant c, then this implies that for all continuous function $(u,v) \mapsto g(u,v)$, the sequence $g(U_n,V_n)$ converges in law to $g(U,c)$.

5.4 Δ method

- Let $(X_n)_{n=1,\ldots k..}$ be a sequence of random variables such that

$$c_n(X_n - \theta) \to_{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

where $(c_n)_{n=1,\ldots k..}$ is a deterministic sequence such that $\lim_{n\to\infty} c_n = \infty$ (ex $c_n = \sqrt{n}$). Suppose the function g differentiable at θ with $g'(x)|_{x=\theta} \neq 0$, then:

$$c_n[g(X_n) - g(\theta)] \to_{\mathcal{L}} \mathcal{N}(0, \sigma^2(g'(\theta))^2)$$

5.5 Laws of large numbers

There are many laws of large numbers depending on the assumptions made on the sequence $(X_n)_{n=1,\ldots k..}$. The most known are the following:

- Let $(X_n)_{n=1,\ldots k..}$ be a sequence of mutually uncorrelated random variables identically distributed (but non necessary independent) with same $m = \mathrm{E}(X_k)$ and $\sigma^2 = \mathrm{var}(X_k)$ that exist

$$M_n = \frac{1}{n}\sum_{k=1}^{n} X_k \to_{\mathcal{P}} m, \quad \text{(weak law of large numbers)}$$

- Let $(X_n)_{n=1,\ldots k..}$ be a sequence of mutually independent identically distributed random variables where $m = \mathrm{E}(X_k)$ exists

$$M_n = \frac{1}{n}\sum_{k=1}^{n} X_k \to_{\mathcal{AS}} m, \quad \text{(strong law of large numbers)}$$

- In the case where X_n is Bernoulli (p) distributed: $P(X_n = 1) = p$, $P(X_n = 0) = 1 - p$,

$$m = p, \text{ and } \sigma^2 = p(1-p).$$

  $X_n = 1$ can be associated with the realization of a specific event A in a repetition of independent random experiments: $X_n = \mathbb{1}_{A_n}$.

  $\Rightarrow \frac{1}{n}\sum_{k=1}^{n} X_k$ and p represent the relative frequency of realization of A in the first n experiments, and the probability of A, respectively.

  The relative frequency of A converges in probability and almost surely to the probability of A.

5.6 Approximation of probability distribution

- For each convergence in distribution : $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, we deduce that for $n \gg 1$ the approximations:

$$F_{X_n}(x) \approx F_X(x) \text{ and } P(X_n \in B) \approx P(X \in B) \text{ for any Borel set of } \mathbb{R}.$$

  More precisely, is it required to specify from what value of n, we have

$$\frac{\mathrm{Sup}_{x; F_{X_n}(x)\neq 0, F_{X_n}(x)\neq 1} |F_{X_n}(x) - F_X(x)|}{F_{X_n}(x)} \ll 1 \text{ (in practice } 10^{-1} \text{ or } 10^{-2}).$$

- For $X_n = \sum_{k=1}^{n} 2^{-k} V_k \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, .., 1 - \frac{1}{2^n}\}$ where $(V_k)_{k=1,...}$ is a sequence of mutually independent uniform distributed in $\{0;1\}$ random variables,

  $\Rightarrow F_{X_n}(x) = \frac{1}{2^n} \sum_{k=1}^{2^n} \mathbb{1}_{[\frac{k-1}{2^n},1]}(x) + \mathbb{1}_{(1,+\infty)}(x) \Rightarrow \lim_{n\to\infty} F_{X_n}(x) = x\mathbb{1}_{(0,1]}(x) + \mathbb{1}_{(1,+\infty)}(x) \iff X_n \to_{\mathcal{L}} X$, where X is uniformly distributed in $[0,1]$

  $\Rightarrow$ for $n \gg 1$ $F_{X_n}(x) \approx F_X(x)$, i.e., X_n is approximately uniformly distributed in $[0,1]$.

- From the central limit theorem, let $(X_n)_{n=1,...k..}$ be a sequence of mutually independent identically distributed random variables, we get for $n \gg 1$, $U_n = \frac{\sum_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}}$ is approximately Gaussian $\mathcal{N}(0,1)$ distributed. Consequently for $n \gg 1$:

$$S_n \overset{\text{def}}{=} \sum_{k=1}^{n} X_k \text{ is approximately Gaussian } \mathcal{N}(nm, n\sigma^2) \text{ distributed}$$

$$M_n \overset{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n} X_k \text{ is approximately Gaussian } \mathcal{N}(m, \frac{\sigma^2}{n}) \text{ distributed}$$

  - Ex: If X_n are uniformly distributed in $\{0;1\}$, $m = \frac{1}{2}$ and $\sigma^2 = \frac{1}{4}$, $U_n \in [-\sqrt{n}, +\sqrt{n}]$
    $\Rightarrow$ for $n \gg 1$ $(n \geq 30)$, S_n is approximately Gaussian $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$ distributed.
  - Ex: If X_n are uniformly distributed in $[0,1]$, $m = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$, $U_n \in [-\sqrt{3n}, +\sqrt{3n}]$
    $\Rightarrow$ for $n \gg 1$ $(n \geq 12)$, S_n is approximately Gaussian $\mathcal{N}(\frac{n}{2}, \frac{n}{12})$ distributed.

  The speed of convergence specifies the order of magnitude of n to have a good approximation. Roughly speaking, the more symmetric the distribution of X_n, the better the approximation.

  This is justified by the Berry Essen inequality:

$$\text{Sup}_x |F_{U_n}(x) - F_U(x)| \leq 0.8 \frac{\text{E}(|X - m|^3)}{\sigma^3 \sqrt{n}}.$$

  which also implies that $F_{U_n}(x) = F_U(x) + O(\frac{1}{\sqrt{n}})$ with the big O notation.

  Ex: If X_n are Bernoulli (p) distributed, $m = p$ and $\sigma^2 = p(1-p)$, $U_n \in [-\sqrt{\frac{np}{1-p}}, +\sqrt{\frac{n(1-p)}{p}}]$ and $\text{Sup}_x |F_{U_n}(x) - F_U(x)| \leq \frac{0.8}{\sqrt{n}} \frac{(p^2+(1-p)^2)}{\sqrt{p(1-p)}}$

  $\Rightarrow$ for $n \gg 1$ $(n \geq 30$, $np > 5$ and $n(1-p) > 5)$, S_n is approximately Gaussian $\mathcal{N}(np, np(1-p))$ distributed, i.e., the Binomial (n,p) distribution is approximated by a Gaussian $\mathcal{N}(np, np(1-p))$ distribution.

  Ex: X_n is Poisson (λ) distributed, $m = \lambda$ and $\sigma^2 = \lambda \Rightarrow$ for $n \gg 1, S_n$ that is Poisson $(n\lambda)$ distributed is approximated by a Gaussian $\mathcal{N}(n\lambda, n\lambda)$ distribution.

  $\Rightarrow$ the Poisson (λ) distribution is approximated by a Gaussian (λ, λ) distribution if $\lambda \geq 10$.

- The convergence X_n Binomial $(n, p_n = \frac{\lambda}{n})$ distributed $\to_{\mathcal{L}} X$ Poisson (λ) distributed, because for k fixed $\lim_{n\to\infty} \binom{n}{k} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$

  $\Rightarrow$ the Binomial (n,p) distribution is approximated by a Poisson (np) distribution if $n \geq 30$ and $np < 5$.

5.7 Exercises

Exercise 5.1 Notion of confidence interval for the estimation of a probability of error

We wish to estimate the probability error $p = P(\text{error})$ of a digital communication. To do so, we observe the communication and we count the number k of errors on n observed symbols. Consequently the rate of error which is defined as $\frac{k}{n}$ is an estimate of the unknown parameter p.

1] Considering the zero-one random variable associated with each observed symbol i, i.e.

$$X_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if error on symbol } i \\ 0 & \text{otherwise} \end{cases},$$

the rate of error can be considered as the sample mean $\bar{X}_n$ of the sequence of random variables $(X_i)_{i=1,\ldots,n}$, i.e.,

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}.$$

Prove the convergence in distribution

$$\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow_{\mathcal{L}} \mathcal{N}(0,1) \tag{5.1}$$

Deduce by application of the weak law of large number, the continuity theorem and the Slutsky theorem, that (5.1) implies:

$$\frac{\bar{X}_n - p}{\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}} \rightarrow_{\mathcal{L}} \mathcal{N}(0,1) \tag{5.2}$$

2] An approximate *confident interval* is a random interval $[p_1(X_1,\ldots,X_n), p_2(X_1,...,X_n)]$ such that

$$P[(p_1(X_1,\ldots,X_n), p_2(X_1,...,X_n)) \ni p] \approx \alpha$$

for $n \gg 1$, where α is the so-called *confidence coefficient*. In practice it takes the values 0.90, 0.95 or 0.99 according to the applications.

Prove from (5.1), then from (5.2) that $p_1(X_1,\ldots,X_n) = \bar{X}_n - h(\alpha)\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}$ and $p_2(X_1,\ldots,X_n) = \bar{X}_n + h(\alpha)\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}$ is such an interval where $h(\alpha)$ is defined by:

$$\int_0^{h(\alpha)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \alpha.$$

Comment on the relative precision $\frac{p_2(X_1,\ldots,X_n) - p_1(X_1,\ldots,X_n))}{2p}$ with respect to n, p and α.

3] We consider now the practical case where $p \ll 1$. How many symbols n must be observed to estimate p with a relative precision of 1% with a confidence coefficient of 95%. In applications, p is unknown. Prove that in practice the problem reduces to count the number $k = \sum_{i=1}^n X_i$ of errors. Give for a confidence coefficient of 95%, the number of errors that must be observed to have a relative precision of 10%, then 1%. Comment on these results.

5.8 Homeworks

Homework 5.1 Generation of a sequence of independent Gaussian random variables by the central limit theorem

1] Suppose, we have a generator of a sequence U_n of independent uniformly distributed in (0,1) random variables (such variates are often simply called random numbers in (0,1)). Deduce from an approximation of the central limit theorem, a generator of sequence X_n of independent Gaussian distributed random variables with mean m and variance σ^2 from the sequence U_n. Comment on this approach. 12 successive values of U_n are frequently used. Explain why.

2] Same question, where the sequence U_n is replaced by sequence V_n of independent uniformly distributed in $\{0, 1\}$ random variables, where 30 successive values of V_n are frequently used.

Chapter 6

Conditional expectation, conditional distribution

6.1 Conditional expectation, solution of the least mean square approximation

- Approximation of Y by a constant

$$\min_a \mathrm{E}[(Y-a)^2] \Rightarrow \quad a = E(Y) \;\text{ and }\; \min_a \mathrm{E}[(Y-a)^2] = \mathrm{var}(Y)$$

$$\min_a \mathrm{E}|Y-a| \Rightarrow \quad a = \mathrm{mediane}(Y)$$

- Approximation of Y by an affine function $\hat{Y} = aX + b$

$$\min_{a,b} \mathrm{E}[(Y - (\underbrace{aX+b}_{\hat{Y}}))^2] \Rightarrow \quad \hat{Y} = m_y + \rho_{x,y}\frac{\sigma_y}{\sigma_x}(X - m_x),\; \min_{a,b} \mathrm{E}[(Y-(aX+b))^2] = \sigma_y^2(1-\rho_{x,y}^2)$$

with $m_x \stackrel{\text{def}}{=} \mathrm{E}(X)$, $m_y \stackrel{\text{def}}{=} \mathrm{E}(Y)$, $\sigma_x^2 \stackrel{\text{def}}{=} \mathrm{Var}(X)$, $\sigma_y^2 \stackrel{\text{def}}{=} \mathrm{Var}(Y)$ and $\rho_{x,y} \stackrel{\text{def}}{=} \frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}$.
Approximation of X by an affine function $\hat{X} = cY + d$

$$\min_{c,d} \mathrm{E}[(X - (\underbrace{cY+d}_{\hat{X}}))^2] \Rightarrow \quad \hat{X} = m_x + \rho_{x,y}\frac{\sigma_x}{\sigma_y}(Y - m_y),\; \min_{a,b} \mathrm{E}[(X-(cY+d))^2] = \sigma_x^2(1-\rho_{x,y}^2)$$

- Approximation of Y by a function $\hat{Y} = g(X)$

$$\min_g \mathrm{E}[(Y-g(X))^2] \Rightarrow \quad \hat{Y} = \mathrm{E}(Y/X = x).$$

By the fundamental theorem of expectation

$$\begin{aligned}\mathrm{E}[(Y-g(X))^2] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(y-g(x))^2 f_{X,Y}(x,y)dxdy \quad \text{(continuous case)}\\ &= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}(y-g(x))^2 f_{Y/X=x}(y)dy\right] f_X(x)dx\end{aligned}$$

where $f_{Y/X=x}(y) \stackrel{\text{def}}{=} \frac{f_{X,Y}(x,y)}{f_X(x)}$ is the probability density function of Y given $X = x$

$\Rightarrow \hat{Y} = E(Y/X = x) = \int_{-\infty}^{+\infty} y f_{Y/X=x}(y)dy$

$$\begin{aligned}\mathrm{E}[(Y-g(X))^2] &= \sum_i\sum_j (y_j - g(x_i))^2 P((X,Y) = (x_i, y_j)) \quad \text{(discret case)}\\ &= \sum_i\left[\sum_j (y_j - g(x_i))^2 P(Y = y_j/X = x_i)\right] P(X = x_i)\end{aligned}$$

where $P(Y = y_j/X = x_i)) = \frac{P((X,Y)=(x_i,y_j))}{P(X=x_i)} = \frac{P(X=x_i)\cap(Y=y_j))}{P(X=x_i)}$ is the conditional probability mass function of Y given that $X = x_i$

$\Rightarrow \hat{Y} = E(Y/X = x_i) = \sum_j y_j P(Y = y_j/X = x_i)$.

- $\min_g \mathrm{E}|Y - g(X)| \Rightarrow \quad g(x) = \text{mediane}(Y)$ given $X = x$
- X and Y are independent (continuous case) $\Leftrightarrow\ f_{Y/X=x}(y) = f_Y(y) \ \Leftrightarrow\ f_{X/Y=y}(x) = f_X(x), \forall(x,y)$
- X and Y are independent (discrete case) $\Leftrightarrow\ P(Y = y_j/X = x_i) = P(Y = y_j) \ \Leftrightarrow\ P(X = x_i/Y = y_j) = P(X = x_i), \forall(x_i, y_j)$
- Two meanings of the conditional expectation: random variable $\mathrm{E}(Y/X) = g(X)$ and deterministic function $x \mapsto \mathrm{E}(Y/X = x) = g(x)$

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y/X)] = \int_{-\infty}^{+\infty}\mathrm{E}(Y/X = x)f_X(x)dx$$

known as law of total expectation.

- Generalized law of total probability

$$P(A) = \int_{-\infty}^{+\infty} P(A/X = x)f_X(x)dx$$

6.2 Conditional cumulative distribution function

- Distribution of Y given $X = x$ (continuous case)

$$\begin{aligned} P(Y \leq y/X \in (x_0, x]) &= \frac{P((Y \leq y) \cap (X \in (x_0, x]))}{P(X \in (x_0, x])} = \frac{P((Y \leq y) \cap (X \in (x_0, x]))}{F_X(x) - F_X(x_0)} \\ &= \frac{1}{\frac{F_X(x)-F_X(x_0)}{x-x_0}} \frac{1}{(x-x_0)} \int_{x_0}^{x} \left[\int_{-\infty}^{y} f_{X,Y}(u,v)dv \right] du \\ \Rightarrow \lim_{x \to x_0} P(Y \leq y/X \in (x_0, x]) &= \frac{1}{f_X(x_0)} \int_{-\infty}^{y} f_{X,Y}(x_0, v)dv = \int_{-\infty}^{y} f_{Y/X=x_0}(v)dv \end{aligned}$$

$$P(Y \leq y/X = x_0) \stackrel{\text{def}}{=} \lim_{x \to x_0} P(Y \leq y/X \in (x_0, x]) = F_{Y/X=x_0}(y)$$

$$f_{Y/X=x_0}(y) = \frac{dF_{Y/X=x_0}(y)}{dy}, \quad \text{(conditional probability density function)}$$

- Conditional variance

$$\text{var}(Y/X = x_0) \stackrel{\text{def}}{=} \text{E}[(Y - \text{E}(Y/X = x_0))^2] = \int_{-\infty}^{+\infty} (y - \text{E}(Y/X = x_0))^2 f_{Y/X=x_0}(y)dy.$$

6.3 Specific case of the Gaussian distribution

- (X, Y) is Gaussian distributed $\mathcal{N}\left(\begin{pmatrix} m_x \\ m_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x\sigma_2\rho_{x,y} \\ \sigma_x\sigma_y\rho_{x,y} & \sigma_y^2 \end{pmatrix}\right)$

$$f_{Y/X=x}(y) = \frac{1}{\sigma_y\sqrt{1-\rho_{x,y}^2}\sqrt{2\pi}} e^{-\frac{[y-(m_y+\rho_{x,y}\frac{\sigma_x}{\sigma_y}(x-m_x))]^2}{2\sigma_y^2(1-\rho_{x,y}^2)}}$$

$$\text{E}(Y/X = x) = m_y + \rho_{x,y}\frac{\sigma_x}{\sigma_y}(x - m_x)$$

$$\text{var}(Y/X = x) = \sigma_y^2(1 - \rho_{x,y}^2)$$

6.4 Exercises

Exercise 6.1 Consider the estimation of a parameter θ through the noisy observation x.

$$x = \theta + n$$

n is a realization of the random variable N of Gaussian distribution $\mathcal{N}(0, \sigma_n^2)$. The prior knowledge about θ is such that θ is a realization of a random variable Θ which is Gaussian distributed with mean m_θ and variance σ_θ^2. The random variables N and Θ are assumed independent.

We are going to choose as estimate $\widehat{\theta}(x)$ of the unknown parameter θ, the more likely value, once x is known. This estimator is called the maximum a posteriori probability (MAP) estimate, i.e., the mode of the posterior distribution of Θ.

1] Give an example of such case.
2] Derive the posterior probability density function $f_{\Theta/X=x}(\theta)$ using the probability density function $f_{\Theta,X}(\theta, x)$. Comment on the posterior probability density function $f_{\Theta/X=x}(\theta)$.
3] Deduce the MAP estimate $\widehat{\theta}(x)$. Prove that $\widehat{\theta} = \mathrm{E}(\Theta/X = x)$
4] Derive the posterior probability density function $f_{\Theta/X=x}(\theta)$ in another way using the Bayes' relation

$$f_{\Theta/X=x}(\theta) = f_{X/\Theta=\theta}(x)\frac{f_\Theta(\theta)}{f_X(x)}$$

and the interpretation of $f_{X/\Theta=\theta}(x)$.
5] Derive the expression of the mean square error $\mathrm{E}[(\widehat{\theta}(X) - \Theta)^2]$. Comment on.
6] Explain why, the posterior probability density function $f_{\Theta/X=x}(\theta)$ cannot be directly derived from the relation $\theta = x - n$ by the interpretation of $f_{\Theta/X=x}(\theta)$?

Exercise 6.2 Two meanings of the conditional expectation: Random variable $\mathrm{E}(Y/X)$ and deterministic function $\mathrm{E}(Y/X = x)$. Geometric interpretation of random variables.

Let (X, Y) be a continuous bivariate random variable such that $\mathrm{E}(Y^2)$ exists.
1] Prove $\mathrm{E}[\mathrm{E}(Y/X)] = \mathrm{E}(Y)$. Comment on.
2] Prove by the geometric interpretation of random variables as vectors in a vector space that: $\mathrm{E}(Y^2) = \mathrm{E}[(Y - \mathrm{E}(Y/X))^2] + \mathrm{E}[(\mathrm{E}(Y/X))^2]$. Deduce that $\mathrm{Var}[\mathrm{E}(Y/X)] \leq \mathrm{Var}(Y)$. Comment on.
3] Derive $\mathrm{E}(Y)$, $\mathrm{Var}(Y)$, $\mathrm{E}[\mathrm{E}(Y/X)]$ and $\mathrm{Var}[\mathrm{E}(Y/X)]$ in the particular case $f_{X,Y}(x, y) = 2\mathbb{1}_{0\leq y\leq x\leq 1}(x, y)$.

6.5 Homeworks

Homework 6.1 Some properties of Markov chain.

Consider a Markov chain, i.e. a sequence of discrete random variables $(X_n)_{n\in\mathbb{N}}$ that take values in a finite or countable set E that satisfies the condition:

$$P(X_n = x_n/X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = P(X_n = x_n/X_{n-1} = x_{n-1})$$

for all $n \geq 1$ and all $x_0, x_1, \ldots, x_n$ in E. We note this property by the simplest notation $P(x_n/x_0, x_1, \ldots, x_{n-1}) = P(x_n/x_{n-1})$ for short. Using this notation, the multiplicative rule

$$P(\cap_{k=1}^n A_k) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)......P(A_n/A_1 \cap ... \cap A_{n-1}),$$

and the partition rule

$$P(A) = \sum_{i\in I} P(A \cap C_i),\ (C_i)_{i\in I} \text{ partition of } \Omega,\ \ I \text{ finite or countable}$$

prove and comment on, at least some of the following properties for all $x_0, \ldots, x_n$ in E.

$$\begin{aligned}
P(x_0, x_1, \ldots, x_n) &= P(x_0)P(x_1/x_0)\ldots P(x_n/x_{n-1}), \quad \text{for } n \geq 0 && (6.1)\\
P(x_n/x_k, x_{k+1}, \ldots, x_{n-1}) &= P(x_n/x_{n-1}), \quad \text{for } 0 \leq k \leq n-1 && (6.2)\\
P(x_n/x_{n-k}, x_{n-k+1}, \ldots, x_{n-l}) &= P(x_n/x_{n-l}), \quad \text{for } 0 \leq l \leq k \leq n && (6.3)\\
P(x_k, x_{k+1}, \ldots, x_n) &= P(x_k)P(x_{k+1}/x_k)\ldots P(x_n/x_{n-1}), \quad \text{for } k < n && (6.4)\\
P(x_k/x_{k+1}, \ldots, x_n) &= P(x_k/x_{k+1}), \quad \text{for } k < n && (6.5)\\
P(x_0, x_1, \ldots, x_n) &= P(x_n)P(x_{n-1}/x_n)\ldots P(x_0/x_1) && (6.6)\\
P(x_n/x_{n+l}, \ldots, x_{n+k}) &= P(x_n/x_{n+l}), \quad \text{for all } 0 \leq l \leq k && (6.7)\\
P(x_n/x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_{n+k}) &= P(x_n/x_{n-1}, x_{n+1}) \quad \text{for } k \geq 1 && (6.8)
\end{aligned}$$

and if x_a^b denotes the set $\{x_a, x_{a+1}, \ldots, x_b\}$,

$$\begin{aligned}
P(x_n/x_{n-l}^{n-k}, x_{n+l'}^{n+k'}) &= P(x_n/x_{n-k}, x_{n+l'}) \quad \text{for } l' \leq k' \text{ and } k \leq l && (6.9)\\
P(x_{n-l}^{n-k}, x_n, x_{n+l'}^{n+k'}) &= P(x_{n-l}^{n-k}/x_n)P(x_n)P(x_{n+l'}^{n+k'}/x_n), \quad \text{for } l' \leq k' \text{ and } k \leq l && (6.10)\\
P(x_{n-l}^{n-k}, x_{n+l'}^{n+k'}/x_n) &= P(x_{n-l}^{n-k}/x_n)P(x_{n+l'}^{n+k'}/x_n), \quad \text{for } l' \leq k' \text{ and } k \leq l && (6.11)\\
P(x_{n-k}, \ldots, x_{n-1}, x_n, \ldots, x_{n+l}) &= P(x_{n-k}/x_{n-k+1})\ldots P(x_{n-1}/x_n)P(x_n)P(x_{n+1}/x_n)\ldots P(x_{n+l}/x_{n+l-1})\\
&\qquad \text{for } k \leq 0 \text{ and } l \leq 0. && (6.12)
\end{aligned}$$

Chapter 7

Random variables simulation

7.1 Meaning of random number sequence

- A periodic deterministic sequence of numbers of very large period such that it passes various statistical tests of hypothesis that the numbers are uniformly distributed on [0,1] and are independent.

  Property:

  From a sequence $(U_n)_{n=1,2,...}$ of mutually independent random variables U_n uniformly distributed in $[0,1]$ or from a sequence $(V_n)_{n=1,2,...}$ of mutually independent random variables V_n uniformly distributed in $\{0;1\}$ (i.e., with $P(V_n = 0) = P(V_n = 1) = 1/2$), it is possible to generate sequences X_n of mutually independent or not independent and arbitrary distributed univariate or multivariate random variables.

7.2 Pseudo-random numbers generators

Generation of sequences of random numbers mutually independent and uniformly distributed in $\{0;1\}$ or $[0,1]$

- Generation by *shift register* of $(V_n)_{n=1,2,...}$: $v_n \equiv \sum_{k=1}^{m} \alpha_k v_{n-k} \bmod 2$, where $\alpha_k \in \{0,1\}$ and $(v_1, v_2, ..., v_m) \neq (0, 0, .., 0)$ called root. Ex: $v_n \equiv v_{n-152} + v_{n-401} \in \{0;1\}$ with period $2^{401} - 1 \approx 5.1\ 10^{120}$

- Generation of $(U_n)_{n=1,2,...}$ from $(V_n)_{n=1,2,...}$: $u_n = \sum_{k=1}^{q} 2^{-k} v_{(n-1)q+k} \in \{0, \frac{1}{2^q}, \frac{2}{2^q}, ..., 1 - \frac{1}{2^q}\} \in [0, 1)$

- Generation of $(U_n)_{n=1,2,...}$ by *linear congruential generator*: $u_n = \frac{w_n}{m}$ where $w_n \equiv a w_{n-1} + c \bmod m$, with $w_0 \in \mathbb{N}$, $w_0 \neq 0 \bmod m$ and $a \in \mathbb{N}^*$, $c \in \mathbb{N}$. Ex: $a = 13^{13}$, $c = 0$ and $m = 2^{59} \approx 5.8\ 10^{17}$ with period $m - 1$

7.3 Generation of random number sequences with arbitrary distributions

- Generic methods : Inverse method and acceptance reject method

– Inverse method for univariate case:

$$x_n = \inf\{\ x\ ;\ u_n \leq F_X(x)\}$$

which reduces for continuous strictly increasing cumulative distribution functions $F_X(x)$, to

$$x_n = F_X^{-1}(u_n), \quad \text{(see exercise 7.1 for discrete random variables).}$$

Ex: $F_X(x) = \frac{x-a}{b-a}\mathbb{1}_{[a,b]}(x) + \mathbb{1}_{(b;+\infty)}(x)$ for uniform distribution in $[a,b] \Rightarrow\ x_n = a + (b-a)u_n$
Ex: $F_X(x) = (1 - e^{-\lambda x})\mathbb{1}_{(0;+\infty)}(x)$ for exponential distribution $(\lambda) \Rightarrow\ x_n = -\frac{1}{\lambda}\ln(1-u_n)$ or more simply $x_n = -\frac{1}{\lambda}\ln(u_n)$

– Inverse method for bivariate case (for inversible cumulative distribution functions $F_{X_1}(x_1)$ and $F_{X_2/X_1=x_1}(x_2)$

$$(x_{1,n}, x_{2,n}) \text{ solution of } F_{X_1}(x_{1,n}) = u_{2n-1} \text{ and } F_{X_2/X_1=x_{1,n}}(x_{2,n}) = u_{2n}$$

– Acceptance reject method for univariate case : see exercise 7.3

– Acceptance reject method for bivariate case: (suppose $f_{X_1,X_2}(x_1,x_2) < c$, $X_1 \in [a_1,b_1]$ and $X_2 \in [a_2,b_2]$)
Draw u_1' uniform on $[a_1,b_1]$ $(u_1' = a_1 + (b_1-a_1)u_1)$,
Draw u_2' uniform on $[a_2,b_2]$ $(u_2' = a_2 + (b_2-a_2)u_2)$,
Draw w uniform on $[0,c]$ $(W = cu_3)$,
If $W < f_{X_1,X_2}(u_1',u_2')$, then $(x_1,x_2) = (u_1',u_2')$, otherwise draws again (u_1',u_2',w)

- Specific methods for the Gaussian (see homework 5.1 and 7.1) and Poisson (see exercise 7.2) distributions for example.

7.4 Exercises

Exercise 7.1 Generation of a sequence of independent identically distributed random variables, by the inverse transformation method. Let U_n be a sequence of independent uniform in [0,1] random variables.
1] For any strictly increasing continuous distribution function $F(x)$, if we define the random variable X_n by $X_n \overset{\text{def}}{=} F^{-1}(U_n)$, prove that the random variable X_n has the distribution function $F(x)$.

Apply this result to the generation of a sequence of uniformly distributed random variables in $[a,b]$, then to exponential distributed distributed random variables from a sequence of independent uniform [0,1] random variables U_n.
2] Let $(p_i)_{i\in I}$ be strictly positive real numbers that satisfy $\sum_{i\in I} p_i = 1$. Prove that the sequence of random variables defined in the following is a sequence of independent discrete random variables with

probabilities $P(X_n = x_i) = p_i$, $i \in I$.

$$X_n = \begin{cases} x_1 & \text{if } U_n < p_1 \\ x_2 & \text{if } p_1 < U_n < p_1 + p_2 \\ \vdots & \\ x_i & \text{if } \sum_{j=1}^{i-1} p_j < U_n < \sum_{j=1}^{i} p_j \\ \vdots & \end{cases}$$

Apply this result to the generation of a sequence of Poisson distributed random variables. Specify the drawback of this approach.

Exercise 7.2 Generation of a sequence of independent Poisson distributed random variables deduced from the Poisson processes (Exercise 2.1 continued)

1] Deduce from the relation $T_n \geq t \Leftrightarrow N_{t_0,t_0+t} \leq n-1$ for $t > 0$ and $n \in \mathbb{N}^*$ used in the Poisson point process between N_{t_0,t_0+t} (number of points between t_0 and $t_0 + t$) and T_n (duration between t_0 to the n-th points to the right of t_0) that has been used in exercise 2.1, the relation for $t > 0$:

$$\begin{aligned} t \leq T_1 \quad &\Leftrightarrow \quad N_{t_0,t_0+t} = 0 \\ T_l < t \leq T_{k+1} \quad &\Leftrightarrow \quad N_{t_0,t_0+t} = k, k \in \mathbb{N}^* \end{aligned}$$

and then a mean to generate a sequence of independent Poisson distributed random variables from a sequence of exponentially distributed random variables of parameter λ.

2] Prove, independently of exercise 2.1, that if $(Y_i)_{i=1,\ldots}$ is a sequence of independent exponentially distributed random variables of parameter λ, the following discrete random variable X is Poisson distributed with parameter λ

$$X = \begin{cases} 0 & \text{if } Y_1 \geq 1 \\ 1 & \text{if } Y_1 < 1 \leq Y_1 + Y_2 \\ \vdots & \\ k & \text{if } \sum_{j=1}^{k} Y_j < 1 \leq \sum_{j=1}^{k+1} Y_j \\ \vdots & \end{cases}$$

(You may used the relation $\int .. \int_{\mathcal{R}^{*k} \cap y_1+\ldots,y_k<1} dy_1 \ldots dy_k = \frac{1}{k!}$).

3] Let $(U_k)_{k=1,\ldots}$ be a sequence of independent uniform (0,1) random variables. Using exercise 7.1], prove that the procedure

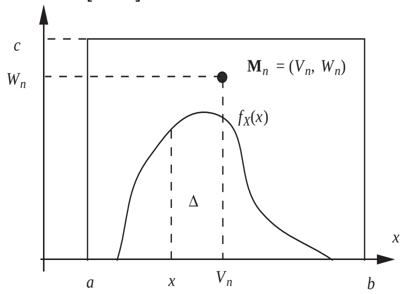
$$X = (\arg\min_n \prod_{k=1}^{n} U_k \leq e^{-\lambda}) - 1$$

gives a random variable X that is Poisson distributed with parameter λ.

4] Let K denote the random number of samples of U_k to give a sample of X. Specify its distribution and its mean $\mathrm{E}(K)$. Comment on?

Exercise 7.3 Acceptance rejection method.

Consider a continuous random variable X with a bounded probability density function $f_X(x)$ with a bounded support included in $[a,b]$ with c such that $f_X(x) \leq c$ and a sequence $(U_n)_{n=1,2,...}$ of independent uniformly distributed random variables in $[0,1]$.



1] Prove that it is possible to generate a sequence of independent points $\mathbf{M}_n = (V_n, W_n)$ uniformly distributed in the rectangle $[a,b] \times [0,c]$ from the sequence $(U_n)_{n=1,2,...}$.
2] Consider the following algorithm: If the point $\mathbf{M}_n = (V_n, W_n)$ is located in the domain Δ below the probability density function $f_X(x)$, we define X by V_n. If the point $\mathbf{M}_n = (V_n, W_n)$ is not located in the domain Δ, we generate a new point $\mathbf{M}_{n+1} = (V_{n+1}, W_{n+1})$ and so on, until a point $\mathbf{M}_n$ is located in the domain Δ. Decompose the event $\{X \leq x\}$ in a countable union of events easier using the events $(M_n \in \Delta)$ and $(M_n \in \Delta_x)$ with $\Delta_x \stackrel{\text{def}}{=} \Delta \cap \{(v,w); v \leq x\}$.
3] Deduce the cumulative distribution function $F_X(x)$, then the probability density function $f_X(x)$ of the random variable Xobtained by this algorithm.
4] If N denotes the number of generated terms of the sequence $(U_n)_{n=1,2,..}$ by generated X, give the expression of $\mathrm{E}(N)$ (the identity $\sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p}$ for $p \in (0,1]$ can be used). Comment on this algorithm.
5] This method can be extended to arbitrary probability density functions $f_X(x)$ if there exists another probability density function $f_Y(y)$ such that the random variable Y can be easily generated, and a constant $c > 0$ such that $f_X(x) \leq c f_Y(x)$. Prove that the following algorithm generates the distribution of X: Generate a sequence Y_n of independent random variables of probability density function $f_Y(y)$ and a sequence U_n independent of Y_n. If $cU_n f_Y(Y_n) \leq f_X(Y_n)$, we define X by Y_n. If $cU_n f_Y(Y_n) > f_X(Y_n)$, we consider the couple (Y_{n+1}, U_{n+1}) and so on, until $cU_n f_Y(Y_n) \leq f_X(Y_n)$.

7.5 Homeworks

Homework 7.1 Generation of a sequence of independent Gaussian random variables by the polar method

Let X and Y be two independent zero-mean unit Gaussian random variables ($\mathrm{E}(X) = \mathrm{E}(Y) = 0$ and $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$).
1] What is the probability density of the two-dimensional random variable (X, Y)? Consider the following change of variables where R and Θ are defined by

$$R = \sqrt{X^2 + Y^2} \qquad \text{and} \qquad \Theta = \mathrm{Arg}(X + iY), \qquad \Theta \in [0, 2\pi)$$

Using the smooth change of variable formula, derive the joint probability density function $f_{R,\Theta}(r,\theta)$ of the random variable (R, Θ). Deduce the marginal probability density functions of R and Θ. Are the random variables R and Θ independent?

2] Deduce from R and Θ, two one to one changes of variable $R = g(U)$ and $\Theta = h(V)$ such that the random variables U and V are independent and uniformly distributed in (0,1).
3] Deduce from $X = R\cos(\Theta) = g(U)\cos(h(V))$ and $Y = R\sin(\Theta) = g(U)\sin(h(V))$, a generation of a sequence X_n of independent Gaussian random variables with mean m and variance σ^2 from a sequence U_n of independent random variables uniformly distributed in the interval (0,1).
4] Compare the pros and cons of this method versus this method deduced from the central limit theorem that is reminded.