Short notes on Probability Theory

F. Borgonovo

Contents

Chapter 1

Probabilities

1.1 Definitions and axioms

Let $\mathcal{E}$ be an experiment which provides a finite number of outcomes $\alpha_1, \alpha_2, \ldots, \alpha_n$. We state the following definitions:

- **Trial** is the execution of $\mathcal{E}$ which leads to a outcome, or sample, α and only one.
- **Space or stochastic universe** associated with the experiment $\mathcal{E}$ is the set $S = \alpha_1, \alpha_2, \ldots . \alpha_n$ of all possible outcomes of $\mathcal{E}$.
- **Event** is any set A of outcomes and what is a any subset of S.
- An **Elementary Event** (or Sample Event) is a set $E \subset S$ with a single outcome.
- A **Certain Event** is the event corresponding to S.
- The **Impossible Event** is the empty set $\emptyset$.
- Events can be combined with operations in use in Set theory, obtaining events such as **union** (or sum) events, **conjunction** (or product or intersection) events, **complement** events, and **difference** events.
- **We say that in a trial event A occurs if the outcome of the trial belongs to A.**

From the above definitions the following properties hold:

- The certain event always occurs;
- The impossible event never occurs;
- A union event occurs if at least one of the component events occurs;
- A joint event occurs if all the components events occur simultaneously;

- Disjoint events can not occur simultaneously, and for this reasons they are called mutually exclusive (as for example sample events and complementary events).

Definition
The probability [1] $P(A)$ *of an event* $A \subset S$ *is a* **measure** *defined on* S *so as to satisfy the following axioms:*

Axiom I: $P(A)$ *is a nonnegative real number associated with the event.*

$$P(A) \geq 0 \tag{1.1}$$

Axiom II: *the probability of the certain event is one.*

$$P(S) = 1 \tag{1.2}$$

Axiom III: *if* A *e* B *are disjoint events*

$$P(A + B) = P(A) + P(B) \tag{1.3}$$

Directly from these axioms, some corollaries follows:

Corollary 1

$$P(A) = 1 - P(\bar{A}) \leq 1 \tag{1.4}$$

In fact we have $A + \bar{A} = S$ e $A\bar{A} = \phi$ and the thesis follows from (1.3).

Corollary 2

$$P(\phi) = 0 \tag{1.5}$$

In fact we have $\phi = \bar{S}$ and the thesis follows from (1.4).

Corollary 3: *If* $B \subset A$, *then*

$$P(B) \leq P(A) \tag{1.6}$$

In fact we have $A = B + (A\bar{B})$ where B and $A\bar{B}$ are disjoint events, and the thesis follows from (1.3) e (1.1).

Corollary 4: *If* $A_1, A_2, \ldots, A_n$ *are disjoint events, and* $A = A_1 + A_2 + \ldots + A_n$, *then we have*

$$P(A) = P(A_1) + P(A_2) + \ldots + P(A_n). \tag{1.7}$$

The thesis follows from repeated application of (1.3).

[1] We use the short term "probability of A" instead of the more correct "probability that the outcome of a trial is in $A \subset S$"

Definition
We say that an experiment $\mathcal{E}$ (or probability space S) is completely described from the probabilistic point of view when, for each elementary event E_i it is given, or it is possible to obtain the corresponding probability

$$p_i = P(E_i)$$

In this case Corollary 4 allows to derive the probability of any kind of event A as the sum of the probabilities of the elementary events that compose A.

Definition
If all the n_S elementary events E_i are equally probable, the space S is called **uniform**.

Corollary 5: *For a uniform space S of n_S elements, the probability of an event A composed of r_A elementary events is*:

$$P(A) = \frac{r_A}{n_S} \tag{1.8}$$

In fact, being by hypothesis $P(e_i) = p$, $(i = 1, 2, \dots)$ and being the elementary events disjoint by definition, applying the (1.7) yields

$$P(A) = \sum_i P(E_i) = p\ r_A \qquad \text{(for every } E_i \subset A)$$

and from (1.1) and (1.7)

$$1 = P(S) = \sum_i P(E_i) = p\ n_S \qquad \text{(for every } E_i \subset S)$$

Dividing the right and left sides of the above expressions we get (1.8).

Several of the cases we will consider in the following are related to uniform spaces and, therefore, probability calculations are performed by counting techniques (as in combinatorial calculus) in order to get r_A and n_S.

1.2 Uniform and finite spaces

The experiments that we consider in this section are modeled with the *urn model* which is an ideal experiment consisting in drawing k objects (elements) from an urn containing n objects (like e.g. numbered or colored balls). The model assumes that all possible outcomes consisting of all the groups that can be formed with k out of n objects are equally probable. Therefore, probabilities can be evaluated via (1.8).

Assuming that groups differ in at least one element or in the order they appear in the group, if objects are drawn together, or one by one with no replacement, the number of such groups is

$$n_S = (n)_k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

While, if objects are drawn one by one with re-insertion/replacement of the drawn element in the urn, the number of such groups is

$$n_S = n^k$$

If the order does not count in distinguishing different groups we have to divide by the number $k!$ of possible permutations of different objects, and we get:

$$n_S = \binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!},$$

in the case of drawn with no replacement, and

$$n_S = \frac{n^k}{k!},$$

in the case of with replacement.

Example (1)
In an urn there are ten objects representing the ten digits $0, 1, \dots, 9$. Evaluate the probability that, upon drawing of 3 elements, the three digits form the event
$A = \{$*number 567*$\}$
$B = \{$*number with three consecutive increasing digits.*$\}$.

We have $n_S = (10)_3 = 10 \cdot 9 \cdot 8 = 720$

$$r_A = 1 \qquad P(A) = \frac{1}{720}$$

$$r_B = 8 \qquad P(B) = \frac{8}{720}$$

Example (2)
Like in previous example but assuming that we have three consecutive drawings with the replacement of the element previously drawn. Evaluate the probability of events A and B of previous example and of event $C = \{$number with all equal digits. $\}$

$$n_S = 10^3$$

$$r_A = 1 \qquad P(A) = \frac{1}{1000}$$

$$r_B = 8 \qquad P(B) = \frac{8}{1000}$$

$$r_C = 10 \qquad P(C) = \frac{10}{1000}$$

Example (3)
Assuming that people have equal probability to be born any day of the year, evaluate the minimum number of people k you need to pick so that the probability of having at least two people born on the same day is greater or equal 0.5.

The experiment is equivalent to drawing k numbers out of an urn that contains 365 objects, each representing a different day of the year. Denoting with $D_k = \{$ *extraction k objects all different* $\}$ we have

$$P(D_k) = \frac{(365)_k}{(365)^k} \qquad k \leq 365$$

The probability that at least two people are born in the same day is:

$$P(\overline{D}_k) = 1 - P(D_k) = 1 - \frac{365!}{(365-k)!\ 365^k} > 0,5.$$

The non linear equation can be solved numerically. Calculating some samples we get:

$$P(\overline{D}_{10}) = 0,166 \ldots P(\overline{D}_{22}) = 0,47576 \ldots P(\overline{D}_{23}) = 0,5072 \ldots P(\overline{D}_{30}) = 0,7062 \ldots$$

The result is then $k = 23$.

1.3 Union of non-disjoint events

Theorem: (1.9)
Given events A and $B \subseteq S$ the following relation holds

$$P(A+B) = P(A) + P(B) - P(AB) \tag{1.10}$$

The proof is obtained by writing event $A + B$ as union of two disjoint events (see Figure 1.1):

$A - AB$ (horizontal dash lines); AB (grid); $B - AB$ (vertical dash lines)

From Corollary 4 we have

$$P(A+B) = P(A-AB) + P(AB) + P(B-AB),$$

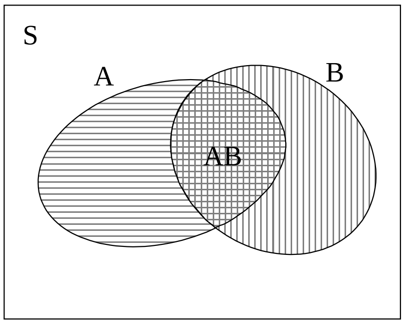

Figure 1.1:

observing that

$$P(A - AB) = P(A) - P(AB),$$

$$P(B - AB) = P(B) - P(AB),$$

and substituting, we get (1.10).♣

If A e B are disjoint events, (1.10) reduces to (1.3).

Example (4)
In a throw of dice, evaluate the probability that number is either even or less than 3.

Denote the event "even number" as A and the event "less than three" as B we have

$$P(D) = P(A + B) = P(A) + P(B) - P(AB),$$

We evaluate $P(A), P(B)$, and $P(AB)$ with the counting process and we find

$$P(A) = \frac{3}{6}, \qquad P(B) = \frac{2}{6}, \qquad P(AB) = \frac{1}{6}.$$

Substituting we get

$$P(D) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{4}{6}.$$

1.4 Conditional spaces, events and probabilities

Assume an experiment $\mathcal{E}$, a space S and a probability measure $P(\cdot)$.

We want to evaluate the probability that the outcome α of a trial of $\mathcal{E}$ verifies event A ($\alpha \in A$) knowing that event M, with $P(M) > 0$, also occurs ($\alpha \in M$).

Obviously, knowing that the outcome is in set M gives some additional information on the possible occurrence of a given α, since all α that do not belong to M are excluded. For this reason the probability of the occurrence of A is no longer the original one $P(A)$, usually referred to as "a priori" probability, but a different one, usually referred to as "a posteriori" (after knowing $\alpha \in M$).

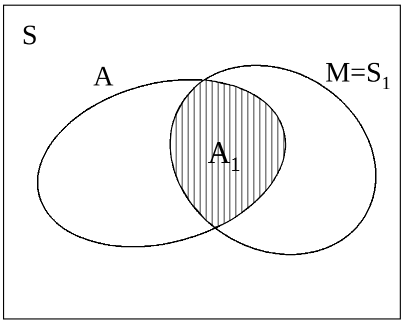

Figure 1.2:

We can formalize this concept by introducing an experiment $\mathcal{E}_1$ (conditional experiment) whose results, α_1, are only those of $\mathcal{E}$ that also belong to M. In this way the space (conditional) associated with $\mathcal{E}_1$ is $S_1 \equiv M$ and event A_1 is called "conditional event".

Obviously, the probabilistic description $P_1(\cdot)$ in S_1 "must" be linked to description $P(\cdot)$ in S, i.e. the probability of the conditional event A_1 must be a function of the probability of the corresponding event AM (unconditional) in S:

$$P_1(A_1) = g[p(AM)].$$

Given A_1 e B_1, disjoint events, Axiom III says that

$$P_1(A_1 + B_1) = P_1(A_1) + P_1(B_1),$$

that is,

$$g[P(AM + BM)] = g[P(AM)] + g[P(BM)],$$

which implies that $g(\dot{})$ is a linear function of the type

$$g[P] = \alpha P + \beta,$$

α and β being constant values.

Axiom IIIa implies $\beta = 0$, whereas Axiom II forces

$$g[P(M)] = 1,$$

i.e.

$$\alpha = \frac{1}{P(M)}.$$

Then we have

$$P_1(A_1) = \frac{P(AM)}{P(M)}$$

However, instead of making use of formalization $\mathcal{E}_1$ and A_1, it is preferable to write the conditional probability as $P(A/M)$, and then:

$$P(A/M) = \frac{P(AM)}{P(M)}. \tag{1.11}$$

Relation (1.11) is meaningful only if $P(M) \neq 0$ and not only excludes $M = \phi$, but also the case where M is a elementary event in a continuous space. Such a conditional probability is defined later.

Example (5)
Evaluate the probability that the outcome of a throw of the dice is 2 knowing that the result is even.

We have

$$S = \{1, 2, 3, 4, 5, 6\} \qquad S_1 = \{2, 4, 6\}$$

and from (1.11)

$$P(2/\text{ even}) = P_1(2) = \frac{1}{3}$$

1.5 Total Probability

It is often easy to determine the probability of an event A conditioned by other events M_i. In this case the probability of A can be determined as a function of the conditional probabilities by resorting to the Total Probability Theorem:

Theorem: Total Probability (1.12)
Given $M_1, M_2, \ldots M_n$ disjoint events such that $M_1 + M_2 + \ldots + M_n = S$ (or more in general $M_1 + M_2 \ldots M_N \supset A$), we have

$$P(A) = \sum_{i=1}^{n} P(A/M_i)P(M_i). \tag{1.13}$$

In fact, since events AM_i are disjoint, and their union provides A, we can write

$$P(A) = \sum_{i=1}^{n} P(AM_i), \tag{1.14}$$

and using the relation

$$P(AM_i) = P(A/M_i)P(M_i), \tag{1.15}$$

attained by reversing (1.11), we get (1.13). ♣

Example (6)
A box contains three types of objects, some of which are defective, in these proportions

type A - 2500 of which 10% defective
type B - 500 of which 40% defective
type C - 1000 of which 30% defective

If we draw an object at random, what is the probability $P(D)$ that, drawing an object, this is found to be defective?

Conditions are met for the validity of (1.13). The probability to draw an object of type A, B, C are respectively

$$P(A) = \frac{2500}{4000} = \frac{5}{8}; \qquad P(B) = \frac{500}{4000} = \frac{1}{8}; \qquad P(C) = \frac{1000}{4000} = \frac{2}{8}.$$

Then we have

$$P(D/A) = \frac{10}{100}; \qquad P(D/B) = \frac{40}{100}; \qquad P(D/C) = \frac{30}{100};$$

and, finally, from (1.13)

$$P(D) = P(D/A)P(A) + P(D/B)P(B) + P(D/C)P(C) = \frac{3}{16}$$

Obviously, in this case the probability can also be obtained as ratio between the number of defective items and the total number of objects:

$$P(D) = \frac{250 + 200 + 300}{4000} = \frac{3}{16}$$

However, in other cases this direct approach is not so easy.

Example (7)
A game is based on the following experiment. A box contains n tags, each one reporting a number arbitrarily determined. The player draws at first r tags and observes their maximum value m_r. Then, further drawings are performed until a value m is observed such as $m > M_r$. Player wins if $m = M$, where M is the maximum value among those reported on the n tags. We want to evaluate the probability $P(V)$ to win, and the value of $r = r_m$ for which this probability is maximum.

Since the positions of the maximum are equally likely, the probability that M is in position k is

$$P(M \text{ in } k) = \frac{1}{n} \tag{1.16}$$

The probability to win, with M in k, is zero if $k \leq r$. For $k > r$ player wins if the maximum m_{k-1} among the first $k-1$ tags is within the first r, and this happens with probability

$$P_r(V/M \text{ in } k) = \frac{r}{k-1}. \tag{1.17}$$

with $k > r$. By the Total Probability Theorem (1.13) we have:

$$P_r(V) = \sum_{k=1}^{n} P_r(V/M \text{ in } k)P(M \text{ in } k) =$$
$$= \sum_{k=1}^{r} P_r(V/M \text{ in } k)P(M \text{ in } k) + \sum_{k=r+1}^{n} P_r(V/M \text{ in } k)P(M \text{ in } k) =$$
$$= \sum_{k=r+1}^{n} \frac{r}{k-1}\,\frac{1}{n} = \frac{r}{n}\sum_{k=r}^{n-1}\frac{1}{k}.$$

Using this result we can write

$$P_r(V) - P_{r+1}(V) = \frac{1}{n}\left(1 - \frac{1}{n-1} - \frac{1}{n-2} - \ldots - \frac{1}{r+1}\right)$$

When r varies from $n-1$ to 1 the above expression changes from positive to negative and the value r_m that maximizes the probability to win is the smallest value of r for which the expression is positive. For n and r large enough you can derive an approximate result by substituting, into the expression of $P_r(V)$, the sum with an integral obtaining

$$P_r(V) \simeq \frac{r}{n}\int_r^n \frac{1}{x}dx = \frac{r}{n}\ln\,\frac{n}{r}.$$

The above expression presents a maximum at

$$\frac{r}{n} = e^{-1} = 0,3675\ldots$$

where it is

$$P(V) = e^{-1} = 0,3675\ldots$$

In Table below exact values of r_m e $P(V)$ are reported for some n.

n	r_m	$P(V)$
3	1	0,5
4	1	$0,458\overline{3}$
5	2	$0,4\overline{3}$
7	2	$0,4142\ldots$
10	3	$0,3986\ldots$
50	18	$0,3742\ldots$
100	37	$0,3710\ldots$
1000	368	$0,3681\ldots$

1.6 Bayes' Formula

If we use (1.11) two times in a direct and a reverse way we get:

$$P(M/A) = \frac{P(AM)}{P(A)} = P(A/M)\frac{P(M)}{P(A)}. \tag{1.18}$$

Using the total probability theorem (1.13) for the denominator, the above expression can be re-written as:

$$P(M_k/A) = \frac{P(A/M_k)P(M_k)}{\sum_{j=1}^{n} P(A/M_j)P(M_j)}. \tag{1.19}$$

Both formulas above are referred to as the Bayes' Theorem. In particular, the (1.19) is referred to as "Bayes' rule for the 'a posteriori' probability", that is after observing the occurrence of the event A. In this case, $P(M_i)$ are called "a priori probabilities" and $P(M_i/A)$ "a posteriori probabilities".

Example (8)
An object drawn at random from the box in Example (6) is found to be defective. Evaluate the probabilities that it is of type A, B and C respectively.

Using Bayes' formula and the preceding results we have

$$P(A/D) = \frac{P(D/A)P(A)}{P(D)} = \frac{10}{30}; \qquad P(B/D) = \frac{P(D/B)P(B)}{P(D)} = \frac{8}{30}.$$

Similarly we have

$$P(C/D) = \frac{12}{30}.$$

Note that while "a priori" the most likely type of object is A, after observing that the object is defective the most likely type of object is C.

Example (9)
With reference to Example (7), let compute the probability that the maximum M is in position $r = k$ knowing that those who played won.

We use (1.19) where A represents victory and M_k represents maximum in position k:

$$P(M \text{ in } k/V) = \frac{P(V/M \text{ in } k)P(M \text{ in } k)}{P(V)}$$

At denominator we have the probability to win already evaluated:

$$P(V) = \frac{r}{n}\sum_{k=r}^{n-1} \frac{1}{k},$$

while at numerator we have:

$$P(M \text{ in } k) = \frac{1}{n}$$

and

$$P(V/M \text{ in } k) = \frac{r}{k-1}$$

Then we have:

$$P(M_k/V) = \frac{C}{k-1} \qquad\qquad k > r$$

and zero elsewhere, being C the normalization constant:

$$C = \frac{1}{\sum_{k=r}^{n-1} \frac{1}{k}}$$

You can see how the knowledge of a preceding victory significantly modifies the distribution of the position of the maximum, which is very different from the uniform one we have "a priori". In particular, the "a posteriori" most probable position is in $k = r + 1$.

Example (10)
Assume that you are presented with three dices, two of them fair and the other a counterfeit that always gives 6. If you randomly pick one of the three dices, the probability that it's the counterfeit is 1/3. This is the a priori probability of the hypothesis that the dice is counterfeit. Now after throwing the dice, you get 6 for two consecutive times. Seeing this new evidence, you want to calculate the revised a posteriori probability that it is the counterfeit.

The 'a priori' probability of counterfeit dice is

$$P(D_c) = \frac{1}{3},$$

while that of a fair dice is

$$P(D_f) = \frac{2}{3}.$$

We have:

$$P(66/D_f) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

$$P(66/D_c) = 1$$

and then using Bayes' formula:

$$P(D_c/66) = \frac{P(66/D_c)P(D_c)}{P(66/D_c)P(D_c) + P(66/D_f)P(D_f)} = \frac{18}{19}$$

Example (11)
A binary communication channel with binary input alphabet, say X_0 and X_1, a binary output

alphabet, say Y_0 and Y_1, and the matrix of probabilities $P(Y_i/X_i)$. The communication problem is a decision problem, that is, to determine which letter among the input alphabet has been transmitted knowing the output letter.

Among the many decision criteria, the soundest one is the one called *Maximum A posteriori Probability* (MAP):

The Bayes' rule allows to write

$$P(X_i/Y_j) = P(Y_j/X_i)\frac{P(X_i)}{P(Y_j)}. \tag{1.20}$$

If we know that a given output has occurred, say Y_0, we say that it has been trasmitted the one among X_0 and X_1 that maximizes (1.20).

For example let us assume that

$$P(Y_0/X_0) = 0.8, \qquad P(Y_1/X_0) = 0.2, \qquad P(Y_0/X_1) = 0.2, \qquad P(Y_1/X_1) = 0.8,$$

(this is called the *Binary Symmetric channel*). Let also assume that input symbols are equally probable. Then also output symbols are equally probable and from (1.20), maximizing $P(X_i/Y_j)$ means maximizing $P(Y_j/X_i)$.

Assume we receive Y_1. Then, being $P(Y_1/X_1) > P(Y_1/X_0)$ we must decide that X_1 has been transmitted. Similarly, when we receive Y_0 we must decide that X_0 has been transmitted.

The following channel is called perfect channel or noiseless channel.

$$P(Y_0/X_0) = 1, \qquad P(Y_1/X_0) = 0,$$

$$P(Y_0/X_1) = 0, \qquad P(Y_1/X_1) = 1,$$

This means that when entering X_0 the output is always Y_0 and when entering X_1 the output is always Y_1, so that the decision process is straightforward.

Also consider the following channel

$$P(Y_0/X_0) = 0.5, \qquad P(Y_1/X_0) = 0.5,$$

$$P(Y_0/X_1) = 0.5, \qquad P(Y_1/X_1) = 0.5,$$

The best decision when receiving Y_0, is left to the reader.

Since error affects all physical measures, and repeating the measure often provides different values, the process toward the interpretation of a measure is exactly a decision process as the one shown above, and shows the importance of the Bayes' rule.

1.7 Statistical independence

Definition. *Two events A and $B \subset S$ are said statistically independent if and only if*

$$P(AB) = P(A)P(B) \tag{1.21}$$

The meaning of statistical independence and the validity of (1.21) is immediate if we observe that if the latter is true, the (1.11) results in:

$$P(A/B) = P(A), \quad P(B/A) = P(B).$$

Based on the definition of conditional probability, this means that the probability of A is not influenced by the occurrence of B and vice versa.

It is easy to see also that if two events A and B are statistically independent so are their complements $\overline{A}$ and $\overline{B}$. Sometimes the statistical independence can be predicted when probabilistic events correspond to physical events which do not influence each other "physically".

Example (12)
In a throw of the dice, check whether the following events

$$A = \{\textit{even number}\}$$
$$B = \{\textit{number one, or two or three}\}$$

are statistically independent.

We have

$$P(A) = 1/2; \qquad P(B) = 1/2; \qquad P(AB) = 1/6;$$

that is

$$P(AB) \neq P(A)P(B)$$

Hence, events A e B are not statistically independent.

Example (13)
In a throw of the dice, check whether the following events $A = \{\textit{an even number appears}\}$ $B = \{\textit{number one, or two, or three, or four appears}\}$ *are statistically independent.*

We have

$$P(A) = 1/2; \qquad P(B) = 2/3; \qquad P(AB) = 2/6$$

and

$$P(AB) = P(A)P(B)$$

Hence, events A e B are indeed statistically independent.

The definition of independence is extended to more than two events in the following way:

Definition

events $A_1, A_2, \dots A_n$ are said statistically independent if and only if

$$\begin{array}{rcll} P(A_iA_j) & = & P(A_i)P(A_j) & (i,j = 1,2,\dots n), (i \neq j) \\ P(A_iA_jA_k) & = & P(A_i)P(A_j)P(A_k) & (i,j,k = 1,2,\dots n), (i \neq j \neq k) \\ \dots\dots\dots & & & \\ \dots\dots\dots & & & \\ P(A_1A_2\dots A_n) & = & P(A_1)P(A_2)\dots P(A_n) & \end{array} \quad (1.22)$$

The number of relations (1.22) is $2^n n - 1$ and ensure the independence of any number of joint events formed by groups extracted from any $A_1, A_2, \dots A_n$.

1.8 Problems for solution

P.1.1 In a throw of three dices, evaluate the probability of having k equal faces, with $k \in [0; 2; 3]$.

P.1.2 Throwing a dice three times, evaluate the probability of having at least one 6.

P.1.3 Assuming women and men exist in equal number, and assuming that 5% of the men are colour blind and that 0,25% of the women are colour blind, evaluate the probability that a person picked at random is colour blind. Then evaluate the probability that, having picked a colour-blind person, this is a male.

P.1.4 In a draw of a card from a deck of 52 cards, verify whether the following events are statistically independent:

a) $A = \{$*drawing of a picture card*$\}$; $B = \{$*drawing of a hearth card*$\}$

b) What if the king of hearths is missing from the deck of cards?
c) What if a card, at random, is missing?

P.1.5 A dice A has four red faces and two white faces. A dice B, vice-versa, has two red faces and four white faces. You flip a coin once, if heads the game continues with dice$A$, otherwise it continues with dice B. a) throwing the dice, what is the probability to get a red face? b) and at the second throwing of the same dice? c) If the first two throws show a red face, what is the probability that also on the third rolling is red? d) If the first n throws show a red face, what is the probability that you are using dice A?

P.1.6 An urn contains two white balls and two black. A ball is drawn and replaced with a ball of a different color. Then a second ball is drawn. Calculate the probability p that the first extracted was white, when the second is white.

P.1.7 The probabilities that three different archers, A, B hit the mark, independently of one another, are respectively 1/6, 1/4 and 1/3. Everyone shoots an arrow. a) Find the probability that only one hits the mark. b) If only one hits the mark, what is the probability he is archer A?

P.1.8 A duel among three people A, B and C is carried out according to the Russian roulette. A six round revolver is loaded with two cartridges. The duelists pass cyclically the weapon, spinning the cylinder every time (so that each duelist has 1/3 probability of being on a loaded chamber) and shooting themselves as long as only one remains alive. Assuming that A is the first, what is the probability that each duelist is the first to die? b) and to win?

P.1.9 From a deck of 52 cards we draw two cards. Find the probabilities of the following events $A = \{$ first card is a King; the second figure$\} = \{K_1; F_2\}$ $\quad B = \{$ at least one figure$\}$

Chapter 2

Random Variables

2.1 Spaces with infinite outcomes: Random Variables

To deal with spaces S with infinite outcomes, we must add another axiom to the ones already introduced in Chapter 1 that extends the summation of the probability measure over infinite terms:

Axiom IIIa: *If $A_1, A_2, \ldots, A_n \ldots$ are disjoint events and $A = A_1 + A_2 + \ldots + A_n + \ldots$, then*

$$P(A) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots \tag{2.1}$$

An example of this type of space is the number of coin flips to get a head. This number is not limited, as the head could never appear in n trial, whichever n is. These spaces are said countable (number of elements of the same cardinality of natural numbers), and can be managed with the methods presented in the previous chapter.

However, we can consider also uncountable spaces, such as is the case when the outcomes is, for example, a point in an interval, or in any general geometrical space. The problem to assign a probability measure to such spaces is not as simple as it is in countable spaces. in fact, it is not straightforward how to extend the method to assign a probability to any subset of the space.

The approach used here is that of "transforming" the space of outcomes into another one more convenient for assigning probabilities. In particular, we map outcomes and events (subsets) of S into the space of real numbers (using integer numbers as a subset to include countable spaces as a special case), or into vectorial spaces for the multidimensional cases.

Let us consider a real function $X(\alpha)$ defined on the space S of the outcomes that binds the set S and the set of real numbers R in order to match every $\alpha \in S$ with one and only one value $X(\alpha) inR$.

With this function, each event $A \subset S$ corresponds a set $I \subset R$ such that for every $\alpha \in A$ we have $X(\alpha) \in I$. In this way the description of an experiment in terms of results α, A and probability events for $P_S(A)$ in S, can be replaced, or rather unified, by the description in terms of real numbers x, sets I and probabilities $P_R(I)$ in R.

A function $X(\alpha)$ which satisfies the above conditions is called "random variable". Typically, the notation is simplified omitting the relation with α and capital letters, such as X, Y, Z, are used to indicate random variables.

2.1.1 Describing a Random Variable

Let X be a Random Variable (RV) and x a real number. The probability of the event $\{X \leq x\}$ is a function of the real variable x. Such function is denoted by $F_X(x)$ (or $F(x)$ when there is no doubt on the RV to which it refers) and is called "Cumulative Probability Distribution Function" (CDF) of X. We have thus,

$$F_X(x) = P(X \leq x) \tag{2.2}$$

with the following properties

1. it has the following limits
$$F(-\infty) = 0 \qquad F(+\infty) = 1 \tag{2.3}$$
2. it is a monotonic non decreasing function of x:
$$F(x_1) \leq F(x_2) \text{ per } x_1 \leq x_2 \tag{2.4}$$
3. it is right continuous [1]:
$$F(x^+) = F(x) \tag{2.5}$$

$F_X(x)$ completely describes RV X; In fact we have, for any x_1, x_2 and x:

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) \tag{2.6}$$

$$P(X = x) = F(x) - F(x^-) \tag{2.7}$$

Proofs

- Properties 1 comes from the fact that $\{X \leq -\infty\}$ is the empty set, and $\{X \leq \infty\}$ is the whole space.
- Property (2.4) comes from (1.6) by observing that for $x_1 < x_2$, we have $\{X \leq x_1\} \subset \{X \leq x_2\}$.
- Furthermore, we have $P(X \leq x + \varepsilon) = P(X \leq x) + P(x < X \leq x + \varepsilon)$, and for $\varepsilon \to 0$ the second term tend to zero because the corresponding event becomes the empty set; hence (2.5) [2].
- Property (2.6) comes from the fact that
$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$$

[1] We denote $F(x^+) = \lim_{\varepsilon\to 0}$ and $F(x^-) = \lim_{\varepsilon\to 0} F(x - \varepsilon)$

[2] Notice that if we had set $F(x) = P(X < x)$, this would have been left continuous

- Property (2.7) comes from (2.6) by setting $x_1 = x - \varepsilon$, $x_2 = x$ and taking the limit $\varepsilon \to 0$.

We say that two RV X and Y are equal if for every α we have $X(\alpha) = Y(\alpha)$, while we say that they are *equally distributed* if they have the same CDF, i.e. if $F_X(z) = F_Y(z)$.

We emphasize that given any function $G(x)$ with properties (2.3) (2.4) and (2.5), we can always build an experiment and define a RV which has $G(x)$ as its CDF. This allows us to manages distributions without specifying to which experiment it refers.

2.2 Continuous Random Variables

A RV X is *continuous* if its CDF $F_X(x)$ is a continuous function in R, together with its first derivative, except at most a countable set of points where the derivative does not exist.

Since for a continuous RV $F_X(x)$ is left continuous, we have from (2.7)

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

For this reason, it is useful introducing the *probability density function* (pdf) of RV X" $f_X(x)$ defined as the derivative of the corresponding CDF:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{2.8}$$

The definition is then completed by assigning arbitrary positive values where the derivative does not exist.

From the definition and properties of $F(x)$ we have

$$f(x) \geq 0 \tag{2.9}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{2.10}$$

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(x)dx \tag{2.11}$$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx \tag{2.12}$$

Definition (2.8) shows that

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}. \tag{2.13}$$

This shows that the pdf can be interpreted as the normalized probability that the RV belongs to a small interval around x and, dimensionally, is a density, hence the name.

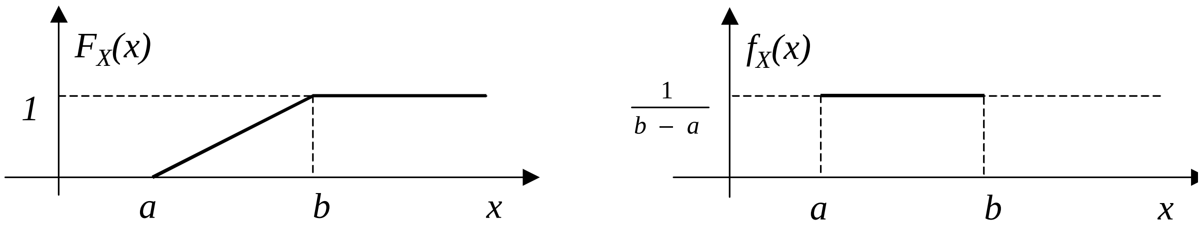

Figure 2.1:

Example (14)
We want to find the CDF and pdf of RV X, defined as the coordinate of a point randomly selected in interval $[a, b]$ of x axis.

As explained in section 2.2 and by (2.2), we immediately have

$$F_X(x) = \begin{cases} \dfrac{x-a}{b-a} & (a \leq x \leq b) \\ \\ 0 & (x < a) \\ 1 & (x > b) \end{cases} \tag{2.14}$$

and from (2.8) or straightly from (2.13)

$$f(x) = \lim_{\Delta x \to 0} \frac{\Delta x}{b-a} \frac{1}{\Delta x}$$

we get

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & (a \leq x \leq b) \\ \\ 0 & \text{elsewhere} \end{cases} \tag{2.15}$$

A RV that satisfies (2.14) and (2.15) is called "uniformly distributed" and the pdf is said "uniform".

Example (15)
A point P is drawn uniformly on a circumference of radius R and center in the origin of axes. Find the pdf of RV X, defined as the coordinate of orthogonal projection of P on the horizontal axis.

To find the pdf let us use the (2.13). With reference to Figure 2.2a, $P(x < X \leq x + \Delta x)$ is the probability that P lies in one of two small arcs shown in the figure, each having a length

$$d\ell = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + (\frac{dy}{dx})^2}$$

Being $y = \sqrt{R^2 - x^2}$, we get:

$$dy = -\frac{xdx}{\sqrt{R^2 - x^2}}$$

by replacing it in the expression above we get

$$d\ell = dx\sqrt{1 + \frac{x^2(dx)^2}{R^2 - x^2}\frac{1}{(dx)^2}} = \frac{dx}{\sqrt{1 - (\frac{x}{R})^2}}.$$

Then we have:

$$f_X(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x}\frac{2\Delta l}{2\pi R} = \lim_{\Delta x \to 0} \frac{1}{\Delta x}\frac{1}{2\pi R}\ \frac{2\Delta x}{\sqrt{1 - (x/R)^2}} = \frac{1}{\pi R}\ \frac{1}{\sqrt{1 - (x/R)^2}}$$

for ($|\ x\ | \leq R$) and zero elsewhere.

The graph of $f_X(x)$ is shown in Figure 2.2b, by which we see that it is more probable to pick the point closer to the extremes rather than to the center.

To obtain the pdf we can also derive the CDF (the task is left to the reader) and get:

$$P(X \leq x) = F_X(x) = \begin{cases} \dfrac{\pi - acos\dfrac{x}{R}}{\pi} & (|\ x\ | \leq R) \\ 0 & (x < -R) \\ 1 & (x > R) \end{cases}$$

Example (16)
A point P is drawn uniformly in a circle of radius R. Derive the pdf of RV Z, defined as the distance of P from the center O of the circle.

$P(z < Z \leq z + \Delta z)$ is the probability that P is taken in the annulus shown in figure 2.3a whose area is $2\pi z\Delta z$.

By (2.13) we get

$$f_Z(z) = \lim_{\Delta z \to 0} \frac{2\pi z\Delta z}{\pi R^2}\ \frac{1}{\Delta z} = \frac{2z}{R^2} \quad (0 \leq z \leq R)$$

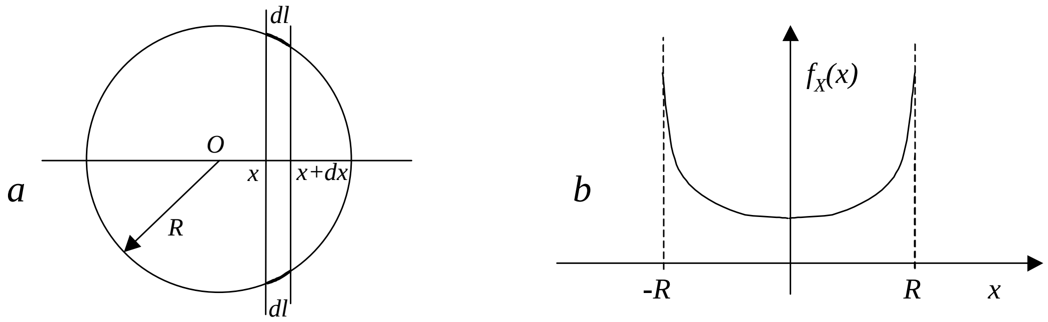

Figure 2.2:

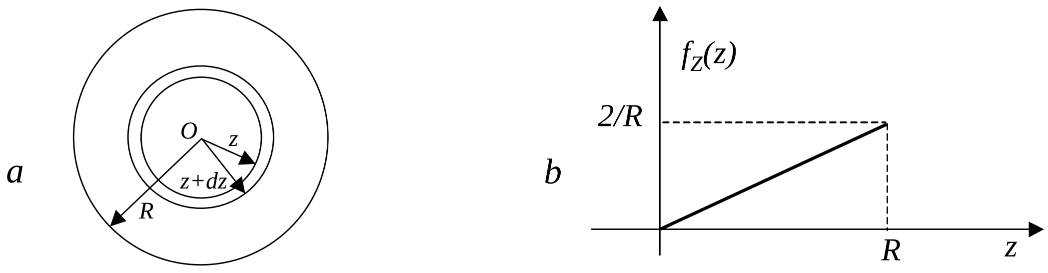

Figure 2.3:

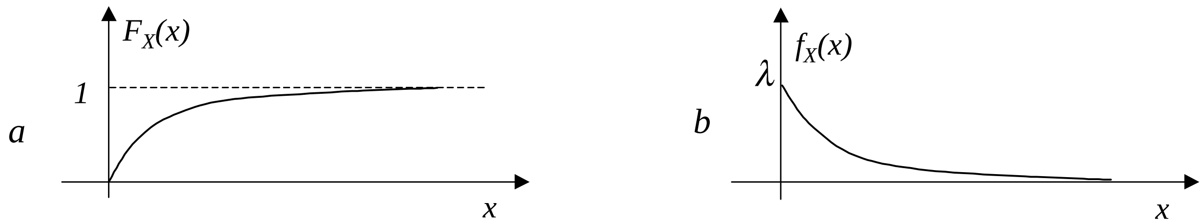

Figure 2.4: CDF and pdf of RV Negative Exponential.

The graph of $f_Z(z)$ is shown in Figure 2.3 b. As we can see P is more likely to be selected next to the circumference than at the center.

Example (17)
The "Negative Exponential" pdf (figure 2.4) is defined as :

$$F(x) = 1 - e^{-\lambda x} \quad (x \geq 0) \tag{2.16}$$

$$f(x) = \lambda e^{-\lambda x} \quad (x \geq 0) \tag{2.17}$$

2.3 Discrete and mixed RV's

A discrete RV X is characterized by a CDF $F_X(x)$ of a staircase type, with discontinuities in a countable set of points $x_i (i = 0, \pm 1, \pm 2 \ldots)$, where it presents steps of value p_i (figure 2.5). In this case from (2.7) we get

$$P(X = x) = \begin{cases} p_i & x = x_i \\ 0 & x \neq x_i \end{cases} \tag{2.18}$$

(2.18) is called *Probability Distribution* of X (not to be confused with the CDF). We have

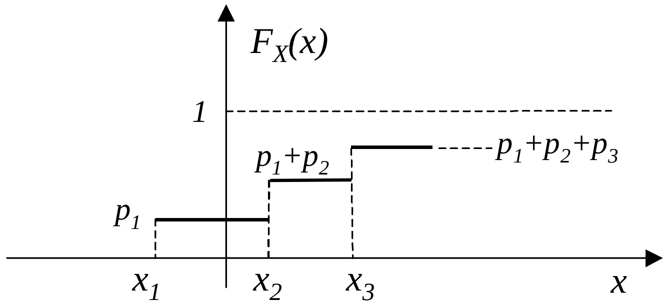

Figure 2.5:

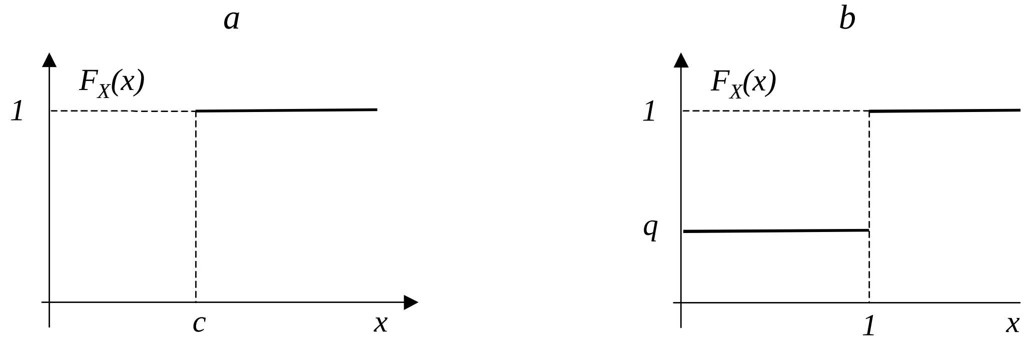

Figure 2.6:

$$p_i \geq 0 \tag{2.19}$$

$$\sum_{i=-\infty}^{\infty} p_i = 1 \tag{2.20}$$

$$F(x) = \sum_{i=-\infty}^{M} p_i \tag{2.21}$$

where M is the maximum i for which $x_i \leq x$.

If the values x_i are integers, then RV X is said an integer RV. It is easy to see that the experiments of previous chapter can be described with integer RVs.

Interesting special cases are:

- the distribution of a constant c (figura 2.6a) :

$$P(X = x) = \begin{cases} 1 & \text{for} \quad x = c \\ 0 & \text{elsewhere} \end{cases} \tag{2.22}$$

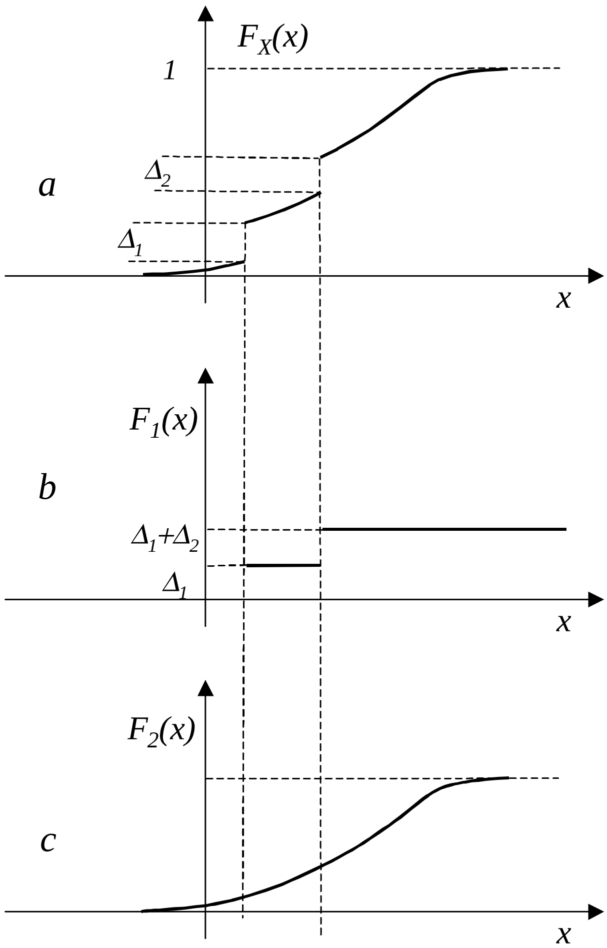

Figure 2.7:

- the Bernoulli (binary) distribution (figure 2.6b):

$$P(X=x)=\begin{cases} p & \text{for} \quad x=1 \\ 1-p=q & \text{for} \quad x=0 \\ 0 & \text{elsewhere} \end{cases} \tag{2.23}$$

- the uniform distribution

$$P(X=x)=\begin{cases} \dfrac{1}{n} & \text{for} \quad x=x_i \quad (i=1,\ldots,n) \\ \\ 0 & \text{elsewhere} \end{cases} \tag{2.24}$$

already encountered in examples with dices, draws from urns, etc.

A RV X is said of the "mixed" type if it is not integer and its CDF presents discontinuities (Figure 2.7a). Such a CDF can always be seen as the sum of a suitable staircase function $F_1(x)$ (figure 2.7b) and a continuous function $F_2(x)$ (figure 2.7c).

We can observer that for integer and mixed RV's we can not rigorously define the pdf, which is often much more convenient description of RVs than the CDF. This difficulty can be overcome by resorting to the theory of generalized functions.

2.4 Bernoulli trials and the Binomial distribution

Experiments of great relevance are those obtained from repeating a single experiment multiple times, always under the same conditions, i.e. *in such a way that the repeats can be assumed statistically independent.*

Repeated independent trials, each of which with only two possible outcomes, say "*success*" (S) and "*failure* (F) are called *Bernoulli* trials.

Denoted $P(S) = p$ and $P(F) = q = 1 - p$, independence ensures that the probability of a particular sequence of successes or failures is achieved by replacing the symbol S and F with p and q, respectively. example:

$$P(SSFSF \dots FFS) = ppqpq \dots qqp$$

Theorem: (2.25)
The probability $P(S_n = k)$ that in n Bernoulli trials k successes and $n - k$ failures occur is given by the following distribution

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k} \qquad (0 \leq k \leq n) \tag{2.26}$$

In fact, each sequence of k successes and $n-k$ failures has probability $p^k q^{n-k}$, whatever the position of the successes. The k successes can be in $\binom{n}{k}$ distinct positions, that represent disjoint events. The probability of k successes is then the sum of the probabilities of such sequences. ♣

Distribution (2.26) is called *Binomial of order n* and it represents the generic term of the power expansion of the binomial

$$1 = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k \, q^{n-k} \tag{2.27}$$

Note that the right hand side represents the probability of the certain event, that is obviously equal to one.

It is also easy to prove that when k ranges from 0 to n, $P(S_n = k)$ initially grows monotonically and then decreases monotonically, reaching its maximum value when $k = k_m$ Integer part of $(n+1)p$. If $(n+1)p$ is an integer we have $P(S_n = k_m - 1) = P(S_n = k_m)$.

Example (18)
A quality control process tests some components out of a factory and components are found defective with a probability $p = 10^{-2}$. Evaluate the probability that out of 10 components checked there are

$A = \{$ *only one defective*$\}$
$B = \{$*two defective*$\}$
$C = \{$*at least one defective*$\}$.

The 10 tests can be modeled as Bernoulli trials with success probability $p = \frac{1}{100}$. Then we have

$$P(A) = \binom{10}{1}(\frac{1}{100})^1(\frac{99}{100})^9 = 0,0913\ldots$$

$$P(B) = \binom{10}{2}(\frac{1}{100})^2(\frac{99}{100})^8 = 0,00415\ldots$$

$$P(C) = \sum_{k=1}^{10}\binom{10}{k}(\frac{1}{100})^k(\frac{99}{100})^{10-k} = 1 - \binom{10}{0}(\frac{1}{100})^0(\frac{99}{100})^{10} = 0,0956\ldots$$

2.5 Moments of a pdf

For the pdf we can define some parameters that resume some properties of the function. These are called *moments*, and the most used are:

1. $k-$th order moments ($k = 1, 2, \ldots$)

$$m_k = \int_{-\infty}^{+\infty} x^k\ f(x)dx \tag{2.28}$$

2. $k-$th order central moments

$$\mu_k = \int_{-\infty}^{+\infty} (x - m_1)^k f(x)dx \tag{2.29}$$

Note that, depending on the specific pdf, some moments may not exist. Parameters of the same meaning can be given also for discrete variables in the form:

$$m_k = \sum_{i=-\infty}^{\infty} x_i^k\ p_i \tag{2.30}$$

$$\mu_k \sum_{i=-\infty}^{\infty} (x_i - m_1)^k\ p_i \tag{2.31}$$

In particular, a parameter of paramount importance is the first order moment m_1. This can be considered as the coordinate of the center of mass interpreting the pdf as a mass distribution along the x with density $f(x)$. Similarly, m_2 is a further index of the dispersion of the distribution from the origin $x = 0$ of the axis, whereas μ_2 (we have $\mu_1 = 0$) provides an index of the dispersion of the distribution from its $x = m_1$ axis. The moments are obviously linked to the central moments by some expressions. For example, by the definition we have

$$\mu_2 = m_2 - m_1^2 \tag{2.32}$$

It can be shown that, if moments of any order exists, then the knowledge of these moments completely determines the pdf $f_X(x)$ (or the CDF in the discrete RV case).

Example (19)
Let us evaluate m_1 and μ_2 for the following RV's

a) Bernoulli RV

b) Binomial RV

a) For Bernoulli RV, from (2.30) and (2.31), we get:

$$m_1 = 0 \cdot q + 1 \cdot p = p$$

$$\mu_2 = (0-p)^2 q + (1-p)^2 p = pq$$

b) For Binomial RV we get

$$m_1 = \sum_{k=0}^{n} k\binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k q^{n-k} = np\sum_{h=0}^{n-1}\binom{n-1}{h} p^h q^{n-h-1} = np$$

By (2.40) we have

$$\begin{aligned} m_2 &= \sum_{k=0}^{n} k^2\binom{n}{k} p^k q^{n-k} = np\sum_{k=1}^{n} k\binom{n-1}{k-1} p^{k-1} q^{n-k} \\ &= np\sum_{h=0}^{n-1}(h+1)\binom{n-1}{h} p^h q^{n-h-1} = np(E[X]_{n-1\ prove}+1) = np[(n-1)p+1] \end{aligned}$$

$$\mu_2 = m_2 - m_1^2 = npq$$

Example (20)
Let us evaluate m_1 and μ_2 for the negative exponential RV. WE have

$$m_1 = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$\sigma^2 = m_2 - m_1^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$$

Theorem: Law of large numbers (2.33)
If X is a RV whose pdf has first order moment m_1, denoted $X_1, X_2, \ldots, x_n$ the outcomes of the RV in n independent repetitions of the experiment, and $\overline{X}_n$ the arithmetic mean of values of X_i, i.e.,

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

(note that $\overline{X}_n$ is itself a RV) we have:

$$P(\lim_{n\to\infty} \overline{X}_n = m_1) = 1 \tag{2.34}$$

Law (2.34) states that the average performed on a number n of outcomes of n independent trials, tends with probability 1 to m_1 when n tends to infinity. For this reason, m_1 is also called the *mean value* or *expected value* of RV X and in this sense, it is also denoted by $E[X]$.

This law is of great importance since it provides a relationship between a pure mathematical parameter, m_1, to another one $\overline{X}_n$ directly derived from an experiment. Similar laws can also be given for high order moments, if they exists.

The proof of the above law is quite involved and will not be given here. We however observe that it can be formulated in another way, tied to probability p_A of event A. To this purpose, define the binary RV X such that it is $X = 1$ if A occurs and $X = 0$ otherwise. Then, if we perform n trials we have

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

being n_A the number of times A occurs. we also observe that

$$m_1(X) = p_A$$

and that

$$\overline{X}_n = \frac{n_A}{n}.$$

Therefore, the law of large numbers can be written as

$$P(\lim_{n\to\infty} \frac{n_A}{n} = p_A) = 1 \quad \clubsuit \tag{2.35}$$

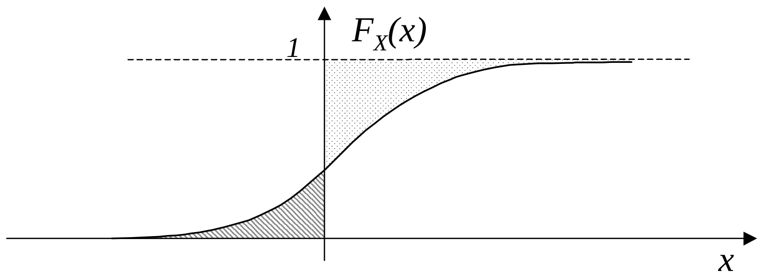

Figure 2.8:

The above formulation of the law provides the interpretation of probability $P(A)$ as the limit of frequencies n_A/n.

Other important properties of m_1 are:

1. If $f(x)$ is symmetric around a value of a and m_1 exists, then $m_1 = a$. In fact

$$m_1 = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} (y+a)f(y+a)dy = \int_{-\infty}^{+\infty} yf(y+a)dy + a = a$$

This come from (2.10) and from the observation that $f(y+a)$ is an even function ;

2. If m_1 exists, it can be expressed as

$$m_1 = \int_0^{\infty} (1-F(x))dx - \int_{-\infty}^{0} F(x)dx \tag{2.36}$$

that is, as the difference between the dashed and the dotted areas in Figure 2.8.

To prove (2.36) assume at first that $f(x)$ is greater than zero only within $a \le x \le b$. Then m_1 can be written as :

$$m_1 = \int_a^0 xf(x)dx + \int_0^b xf(x)dx$$

Taking the integration by parts and using $F(x)$ e $F(x)-1$ as primitive of $f(x)$ respectively in the first and second integrals we have:

$$m_1 = [xF(x)]_a^0 - \int_a^0 F(z)dx + [x(F(x)-1)]_0^b + \int_0^b (1-F(x))dx$$

Since we have assumed $F(a) = 0$ e $F(b) = 1$

$$m_1 = \int_0^b (1-F(x))dx - \int_a^o F(x)dx$$

and (2.36) comes out by setting $a = -\infty$ and $b = +\infty$;

3. If $F_X(x) = 0$ for $x < 0$, for $\alpha > 0$ the following inequality holds

$$P(X \ge \alpha) \le \frac{E[X]}{\alpha} \tag{2.37}$$

In fact,

$$E[X] = \int_0^{\infty} xf(x)dx \ \geq \int_{\alpha}^{\infty} xf(x)dx \ \geq \alpha \int_{\alpha}^{\infty} f(x)dx = \alpha P(X \geq a)$$

hence the thesis. Setting $v = \dfrac{\alpha}{E[X]}$ we get a different expression of (2.37)

$$P(X \geq vE[X]) \leq \frac{1}{v} \tag{2.38}$$

Inequality (2.38) shows how to establish a constraint upon the part of pdf that lies above the mean value ($v > 1$), based on the sole knowledge of the mean value.

$$\mu_1 = \int_{-\infty}^{+\infty} xf(x)dx - m_1 = 0 \tag{2.39}$$

$$\mu_2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - 2m_1 \int_{-\infty}^{+\infty} xf(x)dx + m^2 = m_2 - 2m_1^2 + m_1^2 = m_2 - m_1^2 \tag{2.40}$$

From the definition we see that μ_2 can not be negative; so it must be

$$m_2 \geq \ m_1^2 \tag{2.41}$$

Central moment μ_2, is also called *variance* of RV X and denoted by σ_X^2, whereas σ_X is called *standard deviation.* The variance represents a measure of the dispersion of $f(x)$ around its average value as shown in the following:

Tchebichev Inequality

when $\mu_2 = \sigma^2$ does exist, we have

$$P(\mid X - m_1 \mid \geq \ v\sigma) < \frac{1}{v^2} \tag{2.42}$$

In fact:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - m_1)^2 f(x)dx \ \geq \ \int_{|x-m_1| \ \geq v\sigma} (x - m_1)^2 f(x)dx \geq$$

$$\geq v^2\sigma^2 \int_{|x-m_1| \geq v\sigma} f(x)dx \geq v^2\sigma^2 P(\mid X - m_1 \mid \geq \ v\sigma)$$

hence the thesis. By setting $v\sigma = \varepsilon$ we get alternatively

$$P(m_1 - \varepsilon < X < m_1 + \varepsilon) \ \geq \ 1 - \frac{\sigma^2}{\varepsilon^2} \tag{2.43}$$

$$P(|X - m_1| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \tag{2.44}$$

from which we see that if σ^2 is small, there is a high probability that X belongs to a short interval around m_1. From (2.43) we also see that when $\sigma^2 = 0$, then

$$P(X = m_1) = 1$$

that is, X provides the same constant value for *almost all* the outcomes of the experiment.

Example (21)
Let us apply Tchebichev inequality to bound the probability that the frequency of HEADS in flipping a fair coin n times exceeds $0.5 \pm \varepsilon$.

The frequency of HEADS in n trials is H/n where H is the RV number of HEADS in n trials. This has a Binomial distribution with average $n/2$ and $\sigma^2(H) = n/4$. Therefore,

$$m_1(H/n) = \frac{1}{2}$$

$$\sigma^2(H/n) = \frac{1}{4n}$$

Tchebichev inequality says

$$P(|H/n - m_1| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

and substituting

$$P(|H/n - 0.5| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

we have

$$\begin{array}{llll} \varepsilon = 0.1, & n = 10, & P & \leq 2.5(???) \\ \varepsilon = 0.1, & n = 100, & P & \leq 0.25 \\ \varepsilon = 0.1, & n = 1000, & P & \leq 0.025 \\ \varepsilon = 0.1, & n = 10000, & P & \leq 0.0025 \end{array}$$

We also see that

$$\lim_{n\to\infty} P(|H/n - 0.5| \geq \varepsilon) = 0, \qquad \forall \varepsilon > 0$$

that provides a kind of demonstration of the law of large numbers.♣

Other interesting parameters are

- absolute moments

$$m'_k = \int_{-\infty}^{+\infty} \mid x^k \mid f(x)dx \tag{2.45}$$

- generalized moments

$$\mu_k^{(a)} = \int_{-\infty}^{+\infty} (x-a)^k f(x)dx \tag{2.46}$$

- the mode x_m, defined as the most probable value, i.e., the value at which the pdf is maximum;
- the median m, defined as the value exceeded (or not exceeded) with probability 0.5. In symmetrical pdf the median coincides with the mean m_1.

Example (Persistence of bad luck) (22)
RV X represents the measure of the misfortune experienced in a certain circumstance or trial (waiting time, financial loss, etc..). Denoted by X_0 the misfortune I experienced and $X_1, X_2, \ldots, X_n, \ldots$ the misfortune experienced by others after me in subsequent and independent trials. Denoted by N the number of the first trial in which one is more unfortunate than me (i.e., N is the smallest value of n for which $X_n > X_0$), we want to determine the distribution and average value of RV N.

$P(N = n)$ is the probability that among $n+1$ trials the most unfortunate trials lies in n and the second most unfortunate trials lies in 0 (me). The latter event can happen in $(n-1)!$ different ways among the $(n+1)!$ possible outcomes. Therefore we have

$$P(N = n) = \frac{(n-1)!}{(n+1)!} = \frac{1}{(n+1)n} \qquad (n = 1, 2, \ldots)$$

and

$$E[N] = \sum_{n=1}^{\infty} nP(N = n) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty$$

My bad luck is then without limit!

Example (23)
A gambling is said fair *if the average RV V gain (that is negative if one actually loses) is zero. Check whether the bets on a roulette number is a fair game (36 numbers on which you can bet plus a zero.*

Denoted by C the amount of the bet, the average gain is

$$V = \begin{cases} 35C & \text{with probability } \dfrac{1}{37} \\ \\ -C & \text{with probability } \dfrac{36}{37} \end{cases}$$

$$E[V] = 35C\frac{1}{37} - C\frac{36}{37} = -\frac{1}{37}C$$

The game is not fair, as the bank gets an average gain equal to $C/37$.

2.6 Conditional Distributions and Densities

Let M be an event of space S where RV X is defined. We define CDF of X conditional to M (provided that $P(M) \neq 0$) the function:

$$F_X(x/M) = P(X \leq x/M) \tag{2.47}$$

and similarly for the density a

$$f_X(x/M) = \frac{dF_X(x/M)}{dx} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x/M)}{\Delta x} \tag{2.48}$$

It is easy to check that the above defined functions have all the properties of the CDF and pdf.

In particular, we can define the conditional average

$$E[X/M] = \int_{-\infty}^{+\infty} x f_X(x/M) dx \tag{2.49}$$

and, in the same way, any other conditional moments.

Interesting cases are those where also event M is described in terms of RV X.

2.7 Events conditional to the values of a RV

We define probability of an event A conditional to the value x assumed by a RV X, assuming that $f_X(x) \neq 0$, as the limit

$$P(A/X = x) = \lim_{\Delta x \to 0} P(A/x < X \leq x + \Delta x) \tag{2.50}$$

From Bayes formula (1.18) we get:

$$P(A/X = x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x/A)P(A)}{P(x < X \leq x + \Delta x)}$$

and multiplying by Δx above and below, and taking the limit, from (2.13) and (2.48) we have finally

$$P(A/X = x) = \frac{f_X(x/A)P(A)}{f_X(x)} \tag{2.51}$$

In this way, definition (1.11) is extended also to the case where $P(M) = P(X = x) = 0$, provided that $M = \{X = x\} \neq \phi$. From (2.51) we have then

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

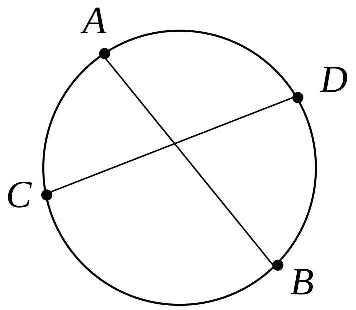

Figure 2.9:

and, by observing that $\int_{-\infty}^{+\infty} f_X(x/A)dx = 1$, we have

$$P(A) = \int_{-\infty}^{+\infty} P(A/X = x)f_X(x)dx \tag{2.52}$$

Furthermore from (2.51), and using (2.52), we obtain finally

$$f_X(x/A) = \frac{P(A/X = x)f_X(x)}{P(A)} = \frac{P(A/X = x)f_X(x)}{\int_{-\infty}^{+\infty} P(A/X = x)f_X(x)dx} \tag{2.53}$$

Relations (2.52) e (2.53) represent respectively the theorem of Total Probability (1.13) and the Bayes theorem (1.19) extended to the continuous case.

Example (24)
Four points A, B, C and D are chosen uniformly and independently on a circumference. Find the probability of event $I = \{$intersection of chords AB and $CD\}$.

Denoted by L the length of the circumference and by x the RV length of arc $\widehat{AB}$ (oriented, figure 2.9), assumed $X = x$, we have

$$P(I/X = x) = P(D \in \widehat{AB})P(C \in \widehat{BA}) + P(D \in \widehat{BA})P(C \in \widehat{AB}) = 2\frac{x(L-x)}{L^2}$$

From (2.52), and being $f_X(x) = \frac{1}{L}$, $(0 < x < L)$, we get

$$P(I) = \int_0^L P(I/X = x)f_X(x)dx = \int_0^L 2\frac{x(L-x)}{L^3}\, dx = \frac{1}{3}$$

The result can be found also observing that, once A is taken, the sequences derived from the permutations of the other 3 points are equally likely, and among these only two lead to a chord intersection.

Example (25)
Let $\widehat{AB}$ and $\widehat{BA}$ be the two arcs of a circumference of length L originated by two points A e B randomly chosen on the circumference itself. If we pick a third random point Q. Find:

a) the probability that Q belongs to $\widehat{AB}$;

b) the pdf and the average length of X, defined as the length of the arc that contains Q.

Denoted by Y e Z the RV that represent the length of $\widehat{AB}$ e $\widehat{BA}$ we have

$$f_Y(x) = f_Z(x) = \frac{1}{L} \qquad (0 \leq x < L)$$

$$E[Y] = E[Z] = \frac{L}{2}$$

a) Using (2.52)

$$P(Q \in \widehat{AB}) = \int_0^L P(Q \in \widehat{AB}/Y = y) f_Y(y) dy = \int_0^L \frac{y}{L} \frac{1}{L} dy = \frac{1}{2}$$

The result could be obtained also by considerations of symmetry.

Similarly $P(Q \in \widehat{AB}) = \frac{1}{2}$

b) From (2.53) we have

$$f_X(x) = f_Y(x/Q \in Y) = f_Z(x/Q \in Z) = \frac{\frac{x}{L}\frac{1}{L}}{\frac{1}{2}} = \frac{2x}{L^2} \qquad (0 \leq x < L)$$

$$E[X] = \frac{2}{3}L$$

2.8 Vectorial RVs

We extend here the concepts and definitions of a scalar RV to the case of two RV's. The extension to the case of more than two RVs is straightforward.

Consider two RV $X(\alpha)$ and $Y(\alpha)$ defined in the same result space S (e.g. the coordinates of a point in the plane, the height and the weight of a person, etc.). By means of this pair of functions a correspondence arises between each event $A \subset S$ and a set D_{xy} of the Cartesian plane, such that for every $\alpha \in A$ the point with coordinates $X(\alpha)$ and $Y(\alpha)$ belongs to D_{xy}. Thus, a joint event in S is represented by a domain D_{xy} in the Cartesian plane.

The probability of the joint events (Figure 2.10a)

$$\{X \leq x, Y \leq y\} = \{X \leq x\}\{Y \leq y\}$$

Is a function of the pair of real variables x and y. Such a function, denoted by $F_{X,Y}(x, y)$, is called *joint CDF* of RVs X and Y.

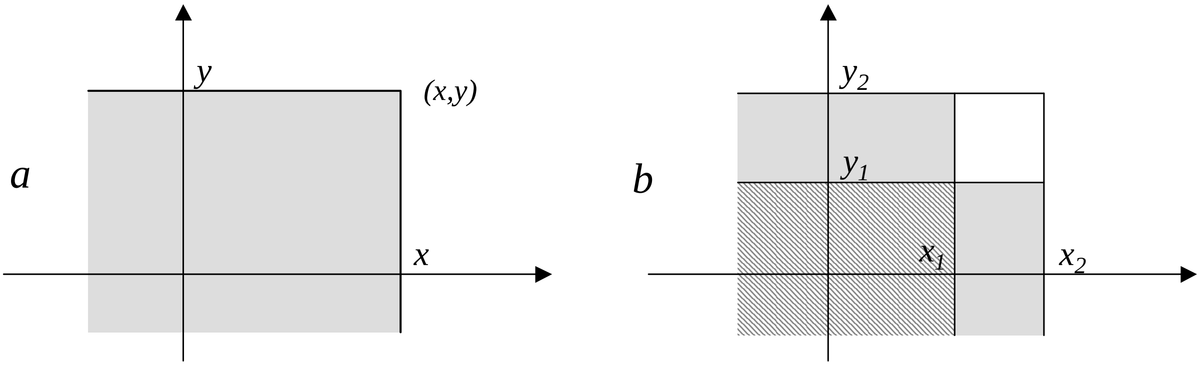

Figure 2.10:

We have then

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \tag{2.54}$$

From the definition we can easily verify the following relations:

$$F(x, \infty) = F_X(x); \quad F(\infty, y) = F_Y(y) \tag{2.55}$$

$$F(\infty, \infty) = 1 \tag{2.56}$$

$$F(x, -\infty) = 0; \quad F(-\infty, y) = 0 \tag{2.57}$$

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \tag{2.58}$$

The latter property can be easily deduced from (2.54) observing Figure 2.10b.

Relations (2.55), (2.56), (2.57) represent sufficient conditions to let a function of two variables represent a joint CDF.

Assuming now that $F_{XY}(x, y)$ has the derivatives that are needed, the joint pdf of RVs X and Y is

$$f_{XY}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{2.59}$$

From the properties previously described we also have

$$f(x, y) \geq 0 \tag{2.60}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \tag{2.61}$$

Furthermore, from the definition of the joint derivative we have

$$f(x, y) = \lim_{\Delta x, \Delta y \to 0} \frac{P(x < X < x + \Delta x, y < Y < y + \Delta y)}{\Delta x \Delta y} \tag{2.62}$$

Relation (2.62) can be often used as starting point to derive the pdf in several problems.

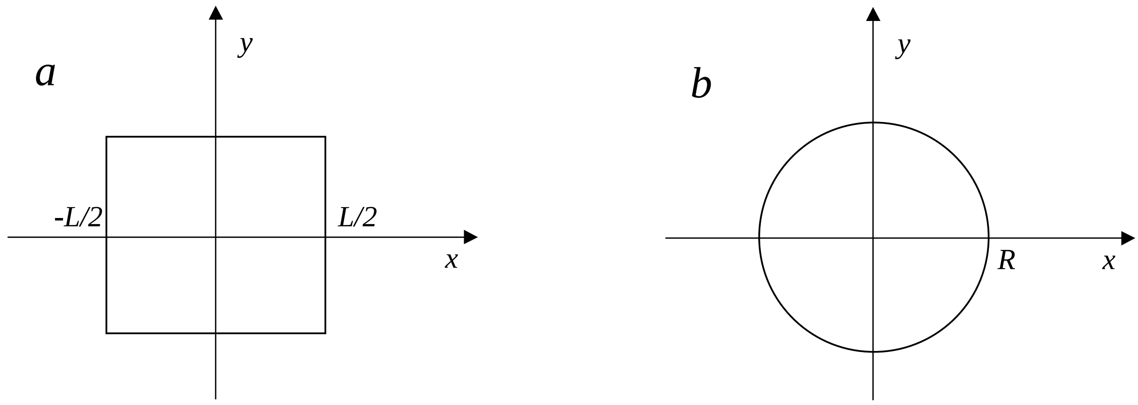

Figure 2.11:

Denoted by $\{(X,Y)\}$ the event of all results α where $X(\alpha)$ and $Y(\alpha)$ belong to domain D, it can be written as a union or intersection of elementary events of the type

$$\{x < X \leq x + \Delta x,\ y < Y \leq y + \Delta y\}$$

and, therefore, we have

$$P((X,Y) \in D) = \int\int_D f(x,y)dxdy \tag{2.63}$$

where the integral is extended over the domain D. It also follows

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy; \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \tag{2.64}$$

The concept of joint pdf can then be extended to the case of discrete and mixed RVs by proceeding as in Chapter 2.

When dealing with multiple RVs, the distributions and the densities of a single RV are called *marginal* to emphasize the difference with joint CDFs and pdf's.

Example (26)
Find the joint and marginal pdf of RV's X and Y Cartesian coordinates of a point Q chosen uniformly in a

a) square of side L and centered at the origin (Figure 2.11a)

b) circle of radius R and center at the origin (Figure 2.11b)

To find the joint density we use (2.62). In this expression the probability at the numerator the probability Q lies into the rectangle of coordinates x, $x + \Delta x$, y, $y + \Delta y$, but since Q is picked uniformly, this probability takes value $\frac{\Delta x \Delta y}{S}$, S being the area of the domain, regardless of the location of the small rectangle. Therefore, we obtain

$$f(x, y) = \begin{cases} \frac{1}{S} & \text{for}(x, y) \in S \\ 0 & \text{elsewhere} \end{cases} \tag{2.65}$$

Such a pdf is still called Uniform in S and the value of the constant $1/S$ depends only from the area of the domain and not by its shape.

About the marginal pdf we have
a)

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy = \int_{-\frac{L}{2}}^{\frac{L}{2}} \frac{1}{L^2} dy = \frac{1}{L}; \qquad \left(-\frac{L}{2} < x < \frac{L}{2}\right)$$

and similarly

$$f_Y(y) = \frac{1}{L}; \qquad \left(-\frac{L}{2} < y < \frac{L}{2}\right)$$

In this case, the marginal pdf are uniform.
b)

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2} dy = \frac{2}{\pi R^2}\sqrt{R^2 - x^2}; \qquad (| x |< R)$$

Here, the marginal pdf's are no longer uniform. In fact, the shape of the domain of point (X, Y) influences the result.

Example (27)
Find the joint and marginal pdf's of RV's Z and Θ, polar coordinates of the point A in the previous example.

Again from (2.62), noting that the area of the elementary surface whose corners have coordinates z, $z + dz$, θ, $\theta + d\theta$ (see Figure 2.12a) amounts to $dz \times z d\theta$, we get

$$f_{Z\Theta}(z, \theta) = \frac{z}{S} \qquad ((z, \theta) \in S)$$

We then have

b)

$$f(z, \theta) = \frac{z}{\pi R^2}; \qquad (0 \le z < R)(-\pi < \theta \le \pi)$$

$$f_Z(z) = \int_{-\pi}^{\pi} f(z, \theta)d\theta = \frac{2}{R^2}z \qquad (0 \le z < R)$$

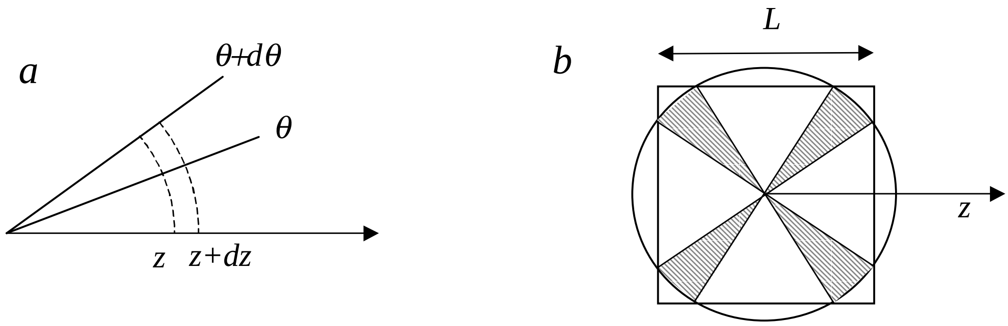

Figure 2.12:

$$f_\Theta(\theta) = \int_0^R f(z,\theta)dz = \frac{1}{2\pi} \qquad (-\pi < \theta \le \pi)$$

a)

$$f(z,\theta) = \frac{z}{L^2} \qquad (\mid z\cos\theta \mid < \frac{L}{2}, \mid z\sin\theta \mid < \frac{L}{2})$$

In evaluating $f_z(z)$ we must note that for $z < \frac{L}{2}$ the integration interval in θ is $-[\pi;\pi]$, which yields

$$f_Z(z) = \frac{2\pi}{L^2}z;$$

on the other side, for $\frac{L}{2} \le z < L\frac{\sqrt{2}}{2}$ the integration interval is shown by the dash sectors in Figure 2.12b; this yields

$$f_Z(z) = \frac{4}{L^2}(arcsen\frac{L}{2z} - arccos\frac{L}{2z})z$$

The complete graph is shown in Figure 2.13a.

Then, about RV θ we have

$$f_\Theta(\theta) = \int_0^{|L/2\cos\theta|} f(z,\theta)dz = \frac{1}{8\cos^2\theta} \qquad (-\pi < \theta \le -\frac{3}{4}\pi)(-\frac{\pi}{4} < \theta \le \frac{\pi}{4})(\frac{3}{4}\pi < \theta \le \pi)$$

$$f_\Theta(\theta) = \int_0^{|L/2\sin\theta|} f(z,\theta)dz) = \frac{1}{8\sin^2\theta} \qquad (-\frac{3}{4}\pi < \theta \le -\frac{\pi}{4})(\frac{\pi}{4} < \theta \le \frac{3}{4}\pi)$$

The graph is shown in Figure 2.13b.

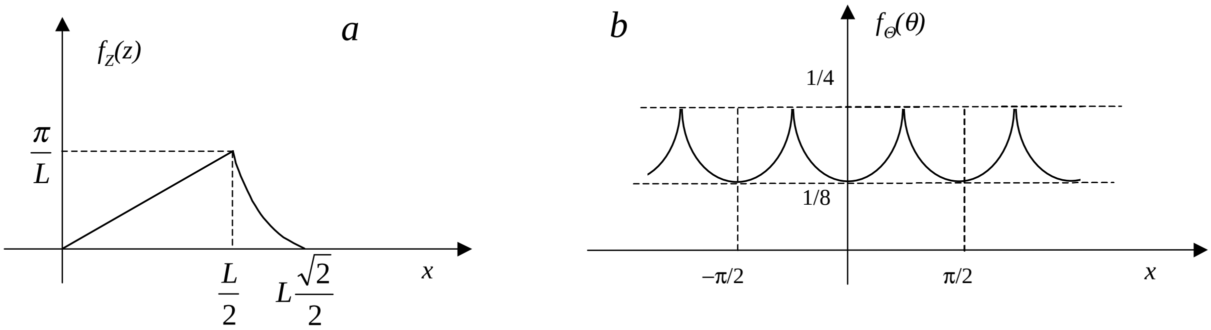

Figure 2.13:

2.9 Conditional pdf's

The extension to the case of two or more RV's of definitions and theorems of paragraph 2.7 are obtained immediately. Here, we limit ourselves to show the new cases of

- pdf of RV Y conditioned by the value assumed by another RV X

$$f_Y(y/X=x) = \frac{f_{XY}(x,y)}{f_X(x)} \tag{2.66}$$

- Total Probability Theorem

$$f_Y(y) = \int_{-\infty}^{+\infty} f_Y(y/X=x)f_X(x)dx \tag{2.67}$$

- Bayes' Theorem

$$f_Y(y/X=x) = \frac{f_X(x/Y=y)f_Y(y)}{f_X(x)} \tag{2.68}$$

- Conditional mean

$$E[Y/X=x] = \int_{-\infty}^{+\infty} yf_Y(y/X=x) \tag{2.69}$$

- Total Probability Theorem with respect to the mean

$$E[Y] = \int_{-\infty}^{+\infty} E[Y/x]f_X(x)dx \tag{2.70}$$

The demonstration of the relations shown above, their validity in the average values as well as the extension to the case of more RV's, is obtained in the usual way.

Example (28)
A point of coordinate X is uniformly selected within interval $[0;L]$ of x axis; Another point of

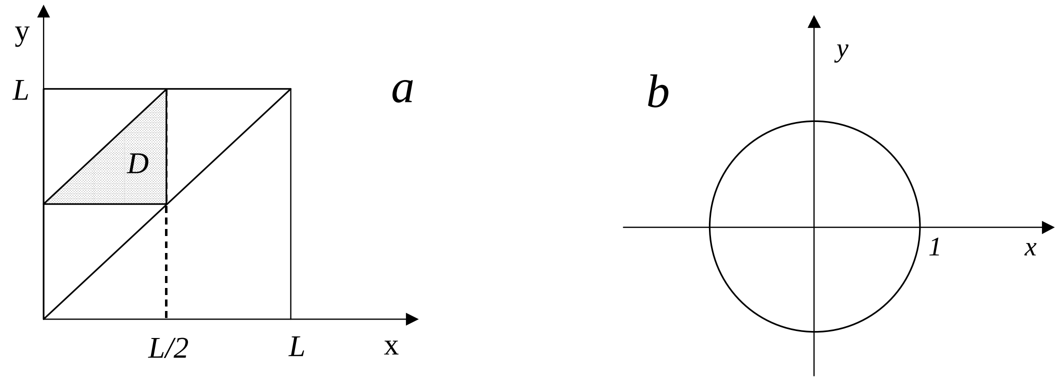

Figure 2.14:

coordinate Y is uniformly selected within interval $[X;L]$. Find the joint pdf of X,Y, the marginal pdf of Y and the probability P that the three segments of length X, Y, and $Y-X$ can form a triangle.

We have

$$f_X(x) = \frac{1}{L}; \qquad (0 < x < L)$$

$$f_Y(y/X=x) = \frac{1}{L-x}; \qquad (x < y < L)$$

Using (2.66) we get

$$f_{XY}(x,y) = \frac{1}{L(L-x)}; \qquad (0 < x < y < L)$$

and from (2.67), by observing that the expression under integration is zero for $x > y$, we have

$$f_Y(y) = \int_0^y \frac{1}{L(L-x)}dx = \frac{1}{L}ln\frac{L}{L-y}; \qquad (0 < y < L)$$

The domain D where X and Y are such as to allow the construction of the triangle, is shown in Figure 2.14a,) and we thus have

$$p = \int_0^{\frac{L}{2}} dx \int_{\frac{L}{2}}^{x+\frac{L}{2}} \frac{1}{L(L-x)}dy = ln2 - \frac{1}{2} = 0,1931\ldots$$

Example (29)
Find the joint and marginal pdf's of the coordinates X and Y of a point P uniformly chosen on a circumference of unit radius and center at the origin 0. (Figura 2.14b)

The marginal pdf of X has been already evaluated in Example 15 as

$$f_X(x) = \frac{1/\pi}{\sqrt{1-x^2}}; \qquad (| x | \leq 1)$$

and by symmetry $f_Y(y) = f_X(y)$.

The joint density is degenerate because is different from zero only on the circumference $x^2+y^2 = 1$. You can, however, make use (2.66) and the impulse function, to write

$$f_Y(y/X = x) = \frac{1}{2}\,\delta(y - \sqrt{1-x^2}) + \frac{1}{2}\,\delta(y + \sqrt{1-x^2})$$

In fact, for $X = x$, Y can take only the two values $\pm\sqrt{1-x^2}$ withe the same probability.

From (2.66) we obtain

$$f_{XY}(x, y) = \frac{1}{2}(\delta(y - \sqrt{1-x^2}) + \delta(y + \sqrt{1-x^2}))\frac{1/\pi}{\sqrt{1-x^2}}$$

or, by simmetry

$$f_{XY}(x, y) = \frac{1}{2}(\delta(x - \sqrt{1-y^2}) + \delta(x + \sqrt{1-y^2}))\frac{1/\pi}{\sqrt{1-y^2}}.$$

2.10 Statistically independent RV's

Two RV X and Y are said to be statistically independent if events $\{X \le x\}$ e $\{Y \le y\}$ are statistically independent for each x and y.

It follows then that two random RV are independent if one of the following relations holds

$$\begin{aligned} F_{XY}(x, y) &= F_X(x)F_Y(y) \\ f_{XY}(x, y) &= f_X(x)f_Y(y) \\ f_X(x/Y = y) &= f_X(x) \\ f_Y(y/X = x) &= f_Y(y) \end{aligned}$$

Similarly, from (1.22) when we have more than two RV's.

Example (30)
Check which pairs of RV's treated in previous examples are statistically independent.

In Example (26) X and Y are independent in case a) but not in the case b), whereas in Example (27) Z and θ are independent in b) but not in a). As we see, the statistical dependence may be linked to the choice of the variables used to describe a phenomenon.

Example (28) X and Y are not (obviously) independent.

Example (29) X and Y are tied together by a deterministic relation.

Example (31)
Given two RV's X and Y independent and exponentially distributed with the same average $1/\lambda$, find:

a) the probability of the event $\{Y > \alpha X\}$ with α real positive;

b) the pdf $f_Y(y/Y > \alpha X)$.

a) We could use the (2.63), being D the domain in which $y > \alpha x$, and given the independence we have

$$f_{XY}(x,y) = f_X(x)f_Y(y) = \lambda^2 e^{-\lambda(x+y)} \quad (x,y>0)$$

More immediately we can use the Total Probability Theorem

$$\begin{aligned} P(Y > \alpha X) &= \int_0^\infty P(Y > \alpha X/X = x) f_X(x)dx = \int_0^\infty e^{-\lambda\alpha x}\, \lambda e^{-\lambda x}dx = \\ &= \frac{1}{\alpha+1}\int_0^\infty \lambda(\alpha+1)e^{-\lambda(\alpha+1)x}dx = \frac{1}{(\alpha+1)} \end{aligned}$$

b) From the definition of conditional pdf, and from the result of point a) we get:

$$\begin{aligned} f_Y(y/Y > \alpha X) &= \lim_{\Delta y \to 0} \frac{1}{\Delta y}\frac{P(y < Y \le y + \Delta y, Y > \alpha X)}{P(Y > \alpha X)} = \\ &= \lim_{\Delta y \to 0} \frac{1}{\Delta y}\frac{P(y < Y \le y + \Delta y, X < y/\alpha)}{P(Y > \alpha X)} = \\ &= \frac{\displaystyle\int_0^{y/\alpha} f_{XY}(x,y)dx}{P(Y > \alpha X)} = \frac{\lambda e^{-\lambda y}\displaystyle\int_0^{y/\alpha} \lambda e^{-\lambda x}dx}{1/(\alpha+1)} = \\ &= (\alpha+1)\lambda e^{-\lambda y}(1 - e^{-\lambda y/\alpha}) \quad (y > 0) \end{aligned}$$

The same result could be easily found by using the Bayes' Theorem in the following way:

$$f_Y(y/Y > \alpha X) = P(Y > \alpha X/Y = y)\,\frac{f_Y(y)}{P(Y > \alpha X)} \tag{2.71}$$

Note that the limit expressions of the derived pdf for $\alpha \to 0$ and $\alpha \to \infty$ are respectively $\lambda e^{-\lambda y}$ e $\lambda^2 y e^{-\lambda y}$ (Erlang-2).

2.11 Joint Moments of two RV's

Given two RV's X and Y the joint moments of order h and k are defined as

$$m_{hk} = \int\int x^h y^k f_{xy}(x,y)dxdy$$

and the central moments of order h and k

$$\mu_{hk} = \int\int (x-m_x)^h (y-m_y)^k f_{xy}(x,y)dxdy.$$

Note that the marginal moments occur when one of the two indices h or k is zero; they coincide with the moments of the same order of the single RV that is, for example:

$$m_{0k} = \int\int y^k f_{xy}(x,y)dxdy = \int y^k f_y(y)dy$$

$$m_{h0} = \int x^h f_X(x)dx$$

and also

$$\begin{aligned} m_{10} &= m_X \\ m_{01} &= m_Y \\ \mu_{20} &= \sigma_X^2 \\ \mu_{02} &= \sigma_Y^2 \end{aligned}$$

The mixed second-order central moment μ_{11}, said also Covariance of V.C. X and Y, is of particular interest. It is linked to m_{11} by the following relation

$$\mu_{11} = m_{11} - m_{10}m_{01} \tag{2.72}$$

From the definitions we see that if X and Y are statistically independent the integral splits into the product of two separate integrals, and we have

$$m_{hk} = m_{h0} \cdot m_{0k}$$

$$\mu_{hk} = \mu_{h0} \cdot \mu_{0k}$$

Specifically, for independent RVs (2.72) shows that

$$\mu_{11} = 0 \tag{2.73}$$

Note, however, that zero covariance does not imply independence among the variables. We only sat that RVs with zero covariance are *uncorrelated*. Also, the following inequalities are valid.

$$m_{11}^2 \le m_{20}\ m_{02} \tag{2.74}$$

$$\mu_{11}^2 \le \mu_{20}\ \mu_{02} \tag{2.75}$$

2.12 Problems for solution

Chapter 2

P.2.1 Given the function $f(x) = \dfrac{C}{\alpha^2 + x^2}$, determine the relationship between C e α in order to make $f(x)$ a pdf. (Cauchy). (3.1)

P.2.2 A point P uniformly chosen in a square of Side L centered at the origin and the x-axis. Find the pdf of RV X, coordinate of the orthogonal projection of P on the horizontal axis.

P.2.3 A point P uniformly chosen in a circle of radius R centered at the origin and the x-axis. Find the pdf of RV X, coordinate of the orthogonal projection of P on the horizontal axis.(3.2)

P.2.4 Find the first order moment of pdf $f(x) = \lambda^2 x e^{-\lambda x}$, $x \geq 0$, and 0 elsewhere.

P.2.5 Find the first order moment of the integer distributions

1. $P(X = k) = (1-p)^{k-1}p, \quad k \geq 1;$
2. $P(X = k) = (1-p)^k p, k \geq \quad 0.$

P.2.6 2 points are chosen uniformly and independently in a segment of length L. Find the pdf of RV X distance to the origin of the point closest to the origin. Find the joint pdf of (X, Y) where Y is distance to the origin of the point farthest to the origin. Extend the result to the case of n points.(3.6)

P.2.7 2 points are chosen uniformly and independently in a circle of radius R. Find the pdf of RV X distance to the center of the point closest to the center.(3.8)

P.2.8 Take a number X from one to six, throw three dices. You win C if X appears once, $2C$ if X appears twice, $3C$ if it appears three times, and you lose C if X does not appear. Check whether this is a fair game. (3.12)

P.2.9 Assume the RV X, lifespan of a component, is uniform in $[0; L]$. We know that the component age is z; find the pdf of its lifespan. Find the pdf of Y, remaining lifespan.

P.2.10 Repeat the previous exercise assuming that the pdf of X is negative exponential. Find the fair amount a a customer of age z must pay to get a capital C if he dies before the year.

P.2.11 Check whether functions of x and y below can represent joint pdfs' and if so check whether X and Y are statistically independent. (5.1)

1. $f(x, y) = 4xy \qquad (0 \leq x \leq 1;\ 0 \leq y \leq 1),$
2. $f(x, y) = 8xy \qquad (0 \leq x \leq y;\ 0 \leq y \leq 1),$
3. $f(x, y) = 4x^2 y \qquad (0 \leq x \leq 1;\ 0 \leq y \leq 1)$

P.2.12 A person in phone booth makes a phone call whose duration is represented by RV X, with negative exponential pdf with mean value $1/\mu$. A second person comes after a time Y. RV negative exponentially with average $1/\lambda$, independent of X. Find the pdf of RV W, the time the latter has to wait to the end of the call. (5.6)

P.2.13 Given two independent RVs' X, and Y, find the probability of the event $\{Y \leq X\}$ when

1. f_X, f_Y are uniform within intervals respectively $[-1; 3], [0; 4]$;
2. f_X, f_Y with the same pdf (you do not need to know the pdf).

What about event $\{Y \leq X/2\}$?

P.2.14 Find the pdf of RV $Z = \min(X, Y)$, where X and Y are two independent negative exponential RVs' with parameters λ and μ respectively. (Hint: observe that $\min(X, Y) > z$ if $x > z$ and $Y > z$. Also, we may take the condition $Y = y$...)

P.2.15 Take interval $[0, X]$, where X is a RV Erlang-2. Then take a point P uniformly within the preceding interval. Find the pdf of Y, length of $\overline{0P}$.

P.2.16 n points are uniformly taken within $[0; T]$. Find the probability that k out of n point lie within an interval $[0; X]$ where RV X is uniform in $[0; T]$.

P.2.17 Two RVs' X and Y are independent and uniformly distributed in $[0; 1]$. Find $f_X(x|X > Y)$, $f_{XY}(x, y|X > Y)$ and $P(X > 2Y|X > Y)$.

Chapter 3

Functions of RV's

3.1 The sum of two continuous RV's

Given the two continuous RV's X e Y, whose joint pdf is known, we want to find the pdf of their sum

$$Z = X + Y \tag{3.1}$$

To this purpose, we note that

$$f_Z(z/X = x) = f_Y(z - x/X = x) \tag{3.2}$$

From the total probability theorem we have

$$f_Z(z) = \int f_Z(z/X = x) f_X(x) dx = \int f_Y(z - x/X = x) f_X(x) dx \tag{3.3}$$

which provides the final formula

$$f_Z(z) = \int f_{XY}(x, z - x) dx \tag{3.4}$$

Symmetrically we have

$$f_Z(z) = \int f_{XY}(z - y, y) dy \tag{3.5}$$

If X and Y are statistically independent the two above become

$$f_Z(z) = \int f_X(x) f_Y(z - x) dx \tag{3.6}$$

$$f_Z(z) = \int f_X(z - y) f_Y(y) dy \tag{3.7}$$

The operation in (3.6) and (3.7) are known as the convolution of pdf's. In fact, the convolution of functions $f(x)$ and $g(y)$ (need not to be pdf's) is defined as

$$f(z) * g(z) = \int f(x)g(z-x)dx = \int f(z-x)g(x)dx$$

Example (32)
Find the pdf of RV $Z = X + Y$ where X and Y are independent RVs with the same pdf, namely

a) $f(x) = \dfrac{1}{a} \qquad (0 < x < a)$

b) $f(x) = \lambda e^{-\lambda x} \qquad (x > 0)$

a) The integrating function in (3.6) is different from zero when both the following conditions apply:

$$\begin{cases} 0 < x < a \\ 0 < z - x < a \end{cases}$$

or

$$\begin{cases} 0 < x < a \\ z - a < x < z \end{cases}$$

Such conditions depend on z and, therefore, we must distinguish the following cases:

- for $z \leq 0 \quad f_Z(z) = 0$
- for $0 \leq z < a \quad$ condition $0 < x < z$ holds, and therefore we have
$$f_Z(z) = \frac{1}{a^2} \int_0^z dz = \frac{z}{a^2};$$
- for $a < z \leq 2a$ condition $z - a < x < a$ holds, and therefore we have
$$f_Z(z) = \frac{1}{a^2} \int_{z-a}^{a} dx = \frac{2-z}{a^2};$$
- for $z > 2a \quad f_Z(z) = 0;$

The seeked pdf is shown in Figure 3.1.

b) The integrating function in (3.6) is different from zero when $\begin{cases} x > 0 \\ z - x > 0 \end{cases}$ that is $\begin{cases} x > 0 \\ x < z \end{cases}$ and, therefore, we have

$$f_Z(z) = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx = \lambda^2 z e^{-\lambda z} \quad (z > 0)$$

The pdf we get is called Erlang-2. This is a pdf out of family

$$E_k(x) = f_k(x) = \frac{(\lambda x)^{k-1}}{(k-1)!} \lambda e^{-\lambda x}, \qquad x \geq 0 \tag{3.8}$$

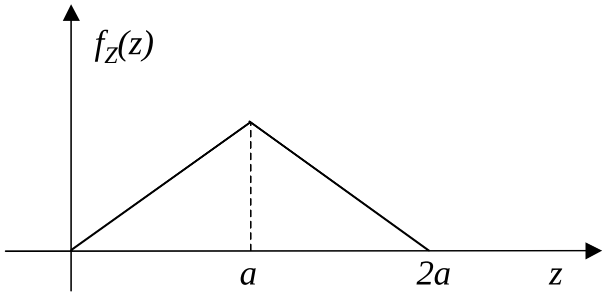

Figure 3.1:

called the Erlang family. Actually the Erlang-1 coincides with the negative exponential. Indeed we have

$$E_m = E_1^{*m} \tag{3.9}$$

$$E_m * E_n = E_{m+n} \tag{3.10}$$

Property (3.10) can be stated saying that the Erlang family is closed with respect to the convolution.

3.2 The sum of two integer RV's

Much as in the previous case we can write the distribution of the sum $Z = X + Y$, with X and Y integer RV's as

$$P(Z = k) = \sum_j P(X = j, Y(k-j) = \sum_j P(X = k-j, Y(j), \tag{3.11}$$

which, again, becomes the convolution if the two RV's are independent.

Example (33)
Find the distribution of RV $Z = X + Y$ where X and Y are Binomial RV's of order n and m with the same success probability p.

A Binomial RV X of order n represents the number of successes in n Bernoulli independent trials and as such can be seen as a sum of n binary independent RV's V_i whose distribution is

$$P(V_i = 1) = p \qquad P(V_i = 0) = q = 1 - p$$

where p is the probability of a successful trial. It follows that the sum of two V.C. Independent binomial of order h and k with the same success probability p is still a binomial of order $k + h$, and, therefore, also the Binomial family is closed respect to the operation of convolution.

It is left to the reader to verify the above by expressly applying the convolution operation.♣

By the definition we see that the average function is interchangeable with a linear operation and, therefore, we have

$$E[X+Y] = E[X] + E[Y] \tag{3.12}$$

We have also, and the proof is left to the reader,

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y] + \mathrm{COVAR}[XY] \tag{3.13}$$

Of course, RVs are independent, or just uncorrelated, we have

$$\mathrm{VAR}[X+Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y] \tag{3.14}$$

In particular, the variance of the average of n values, $\overline{X}_n$, is

$$\mathrm{VAR}[\overline{X}_n] = \frac{\mathrm{VAR}[X]}{n} \tag{3.15}$$

which explains the law of large numbers, as the variance of $\overline{X}_n$ decreases as $n \to \infty$

3.3 Problems for solution

Chapter 3

P.3.1 f_X, f_Y are uniform within intervals respectively $[0;5]$ $[-3,-1]$. Find the pdf of RVs' (6.1)

1. $Z = X + Y$
2. $W = X - Y$

P.3.2 Let X e Y be independent RVs' with negative exponential pdfs' and average value $\frac{1}{\lambda}$. Find the pdf of RVs (6.2)

1. $Z = X - Y$
2. $W = X + \dfrac{Y}{2}$

P.3.3 Find $P(Z = n)$ where $Z = X + Y$ is the sum of the numbers that appear in the rolling of two dices. (6.6)