# Probability and Statistics

Radjesvarane ALEXANDRE

October 7, 2016

## Chapter 1

# Probability Spaces

# Chapter 2

# Random Variables

We may keep in mind that a random variable is a quantity which depends only on the result of the (random) experiment:

- number of "6" obtained while throwing 3 dices;
- number of phone calls during one hour...

are examples of random variables.

More precisely, we have

**Definition 2.0.1**

*Let $(\Omega, \mathcal{T}, P)$ be a probability space. Let $E$ be any set. Any map*

$$X : \Omega \to E, \omega \in \Omega \to X(\omega)$$

*is called a random variable on the probability space, with values in $E$.*

If $E = \mathbb{R}$, we say that we have a real random variable;
If $E = \mathbb{R}^d$, we say that we have random vector.

If $B \subset E$, we use the following equivalent notations

$$X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} \equiv X \in B$$

Note carefully that it is a subset of $\Omega$.

In this chapter, we shall only consider rrv.

**Example 2.0.1**

*Consider throwing two well balanced dices. Work out this example with the rv X= sum of the two results.*

$\odot$

**Example 2.0.2**

*i) Le $E_i$ be the following experiment: a transistor is taken at random in a box containing 10 of brand A and 20 of brand B, and we keep in mind its brand. Here the space of events is $\Omega = \{A, B\}$. Let $X$ be the rv taking the value 1 if the transistor is of brand $A$ and 0 otherwise. We have*

*$X(\Omega) = \{0,1\}$. Here the rv $X$ is the indicator function of the event F: a transistor of brand A has been drawn.*

*ii) Let $E_{ii}$ be the experiment where we choose at random a number in the interval $[0,10)$. Let $X$ be the rv taking the value of obtained number and $Y$ the rv equals to the integer part of this number. Then $\Omega = [0,10)$, $X(\Omega) = [0,10)$ and $Y(\Omega) = \{0,1,...,9\}$.*

*Here, in fact, the rv $X$ is the identity function.*

$\odot$

Note that if $\Omega$ is discrete, then $X(\Omega)$ is also discrete. However, if $\Omega$ is continuous, then $X(\Omega)$ could be discrete as well as continuous (as for $Y$ in the previous example).

## 2.1 Distribution function

**Definition 2.1.1**

*The distribution function of the rv X is defined by, for any real $x$*

$$F_X(x) = P(X \leq x)$$

**Proposition 2.1.1**

*Properties: We have*

1. $0 \leq F_X(x) \leq 1$;
2. $\lim_{+} \infty F_X = 1$;
3. $\lim_{-\infty} F_X = 0$;
4. $F_X$ *is increasing.*
5. $F_X$ *is right continuous, that is* $F_X(x) = F_X(x^+)$ *(limit on the right at $x$).*

Distribution functions are an important tool, at least because their knowledge enable to reconstruct the law of X.

**Proposition 2.1.2**

*We have*

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

We deduce

**Proposition 2.1.3**

*We have*

$$P(X = x) = F_X(x) - F_X(x^-)$$

*where $F_X(x^-)$ denotes the left limit at $x$.*

It is important to note that if $F_X$ is a continuous function, then the probability $P(X = x)$ is zero for any real $x$. In that case, that is if the distribution function is continuous, we have

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

**Definition 2.1.2**

*Let $X$ be a rv taking at most a countable number of values, $X(\Omega) = \{x_1, x_2, ...\}$. Then we say that $X$ is a discrete rv.*

**Example 2.1.1**

*Let us observe the lifetime of an electric bulb which does work. Assume that the probability of the lifetime in hours of this kind of bulb belonging to the interval $[a, b)$ is given by*

$$P(a \leq T < b) = e^{-a/100} - e^{-b/100}, 0 < a < b \leq \infty.$$

*Let $X$ be the number of complete periods of 100 hours. Then*

$$X(\Omega) = \{0, 1, 2, ..\}.$$

*That is $X$ is a discrete rv.*

*As $X$ only takes positive values, then*

$$F_X(x) = P(X \leq x) = 0, \forall x < 0$$

*Morover, for any $x \in [0, 1)$, we have*

$$F_X(x) = P(X \in \{0\}) = P(X = 0) = P(T < 100) = 1 - e^{-1}$$

*where we used the fact that $P(T = 0) = 0$.*

*Similarly for any $x \in [n, n+1)$, with $n \in \mathbb{N}$, we have*

$$F_X(x) = 1 - e^{-n-1}$$

*Thus*

$$F_X(x) = 0 \text{ si } x < 0 \text{ et } F_X(x) = 1 - e^{-[x]-1} \text{ si } x \geq 0$$

*Therefore, $F_X$ is a step function.* $\odot$

**Definition 2.1.3**

*Let $X$ be a rv taking an infinite non countable number of values. If $F_X$ is continuous, we say that $X$ is a continuous rv.*

**Example 2.1.2**

*In the previous example, we may write $T(\Omega) = (0, +\infty)$ and*

$$F_T(t) = P(0 < T \leq t) = P(0 \leq T < t) = 0 \text{ if } t \leq 0 \text{ and } 1 - e^{-t/100} \text{ if } t > 0$$

*since $P(T = 0) = 0$ and $P(T = t) = 0$ (continuous space $\Omega$). As $F_T$ is continuous, it follows that $T$ is a continuous rv. Note that $F_T$ is derivable everywhere except at 0.* ⊙

## 2.2 Mass and density functions

**Definition 2.2.1**

*Let $X$ be a discrete rv. Then $X(\Omega) = \{x_1, x_2, ...\}$. The function $p_X$ defined by*

$$p_X(x_k) = P(X = x_k)$$

*for $k = 1, 2, ...$, is called the probability mass function of X.*

It is also called probability law of X.

**Remark 2.2.1**

*Note that $p_X$ satisfies*

$$P_X(x_k) \geq 0, \forall k, \sum_k p_X(x_k) = 1$$

**Remark 2.2.2**

*When $X(\Omega)$ is a finite set, the function $p_X$ is also given by a table. For example, if*

$$p_X(x) = 1/4 \text{ if } x = -1, 1/2 \text{ if } x = 0, 1/4 \text{ if } x = 1$$

*we may write*

| $x$ | $-1$ | $0$ | $1$ |
|---|---|---|---|
| $p_X(x)$ | $1/4$ | $1/2$ | $1/4$ |

*and similarly for the distribution function.*

Consider now the case of a continuous rv X.
Let $x$ be a point where $F_X$ is derivable. Then we may write

$$P(x < X \leq x + \varepsilon) = F_X(x + \varepsilon) - F_X(x)$$

$$= \frac{F_X(x + \varepsilon) - F_X(x)}{\varepsilon}\varepsilon \sim f_X(x)\varepsilon$$

with $\varepsilon > 0$ small enough and $f_X(x)$ beging the derivative of $F_X$ at $x$.
Then we can set the following definition

**Definition 2.2.2**

*Let $X$ be a continuous rv. The function $f_X$ defined by*

$$f_X(x) = \frac{d}{dx}F_X(x)$$

*if it exists at $x$, is called the probability density function of $X$.*

Thus keep in mind that $f_X(x)\varepsilon$ is almost equal to the probability that the rv X takes a value in a small interval of length $\varepsilon$ around $x$.

Careful: $f_X(x)$ is not the probability that X takes the value $x$, as in fact this probability is zero, as X being continuous.

We have the following properties, if for example $F_X$ is $C^1$:
i) $f_X(x) \geq 0$, as $F_X$ is increasing;
ii) Summing the definition of $f_X$, we find

$$\int_{-\infty}^{x} f_X(t)dt = F_X(x)$$

In particular

$$\int_{-\infty}^{+\infty} f_X(x)dx = F_X(+\infty) = 1$$

A positive function with this property is called a density (function).

We have the following important relation

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$

Thus the probability that X belongs to the interval $(a, b]$ is given by the area under the curve of $f_X$ from $a$ to $b$.

**Remark 2.2.3**

*If $X(\Omega)$ is not the full set $\mathbb{R}$, we define $f_X$ by extending by zero out of this interval.*

## 2.3 Important examples of discrete RV

### 2.3.1 Bernoulli Law

Assume that we are performing a random experiment E. Let$A \subset \Omega$ and X the indicator rv of this event A, that is $X = 1$ if A occurs, and 0 otherwise.

We say that X follows a Bernoulli law with parameter $p$, where $p = P(A)$ is the probability of a success. We have

| $x$ | 0 | 1 | $\Sigma$ |
|---|---|---|---|
| $p_X(x)$ | $1-p$ | $p$ | 1 |

We may also write

$$p_X(x) = p^x q^{1-x}, x = 0, 1$$

with $q = 1 - p$. Here $p$ is a given known parameter.

### 2.3.2 Bernoulli Trials and Binomial Law

We now assume that we repeat the experiment E $n$ times. We then say that the trials $E_1$, ..., $E_n$ form Bernoulli trials if

a) these trials are independent and

b) if the probability of a success is the same for each of these trials.

For example, assume that an individual should cross $n$ cross-roads, equipped with traffic lights (for example to go to work, not in Hanoi of course ...). Let $E_k$ be the random experiment which consists in observing if that individual should wait for green light when arriving to the $k$- cross-road. Then these trials $E_1$, ... , $E_n$ form Bernoulli trials iff we may assume that the probability that this person could cross a cross-road without stopping is the same for each cross-road and does not depend on what happened arriving to the previous cross-roads.

In practise, unless the traffic lights are not too far from each other, the independance assumption is not exactly satisfied. Similarly, for assumption b), we need that the duration of the green light in the full cycle of traffic lights should be the same for each cross-road, which is not true in general.

**Binomial Law**

Let X be the number of success with $n$ Bernoulli trials.

We say that X follows a binomial law with parameters $n$ and $p$, where $p$ is the probability of a success. One has $X(\Omega) = \{0, 1, ..., n\}$. We write $X \sim B(n, p)$.

The probability mass function of $X \sim B(n, p)$ is given by

$$p_X(k) = C_k^n p^k q^{n-k}, \text{ pour } k = 0, 1, ..., n$$

**Preuve :** Indeed, let E be the random experiment which consists in observing if event A occurs or not for each of the $n$ Bernoulli trials. Then, we have

$$\Omega = \{A....A, (n \text{ times }, X = n), ..., A^c...A^c(n \text{ times }, X = 0)\}$$

There are $2^n$ elementary events in all. Let $\omega$ be one of these elementary events, for which $X = k$. Without any difficulty, one can show that $P(\{s\}) = p^k q^{n-k}$ using independence. There are $C_k^n$ different elementary events in $\Omega$ with exactly $k$ success and $(n-k)$ failures. Thus the formula. ■

### 2.3.3 Geometric Law

Let X be the number of Bernoulli trials necessary in order to get the first succes. Then $X(\Omega) = \{1, 2, ...\}$. We then say that X follows a geometric law with parameter $p$. We write $X \sim Geom(p)$. In this case, we have

$$p_X(k) = q^{k-1}p, \text{ pour } k = 1, 2, ...$$

Indeed let E be the experiment which consists in observing the result of Bernoulli trials. We have

$$\Omega = \{A(''X = 1''), A^c A(''X = 2''), A^c A^c A(''X = 3''), ...\}$$

Then we can write

$$p_X(k) = P(X = k) = P(A^c...A^c A) =_{ind} q^{k-1}p$$

Note that function $p_X$ is positive and

$$\sum_{k=1}^{\infty} p_X(k) = ... = 1$$

The distribution function is given, for integer $n$, by

$$F_X(n) = ... = 1 - q^n$$

In particular

$$P(X > n) = q^n, n = 1, 2, ...$$

To end up, sometimes on defines X as being the number of Bernoulli trials before getting the first success. In that case, we have $X(\Omega) = \{0, 1, ..\}$ and function $p_X$ becomes

$$p_X(k) = q^k p,\ k = 0, 1, ..$$

If one denotes this rv by Y, then $Y = X - 1$.

The choice between X and Y depends on the context. For example, if we are looking for a stochastic model for the number of people involved in an accident, it is preferable to take X with $X(\Omega) = \{1, 2, ...\}$. On the other hand, if the rv X denotes the number of injuried people, we must choose a rv X such that $X(\Omega) = \{0, 1, ...\}$.

### 2.3.4 Poisson law

**Definition 2.3.1**

*Let X be a discrete rv with $X(\Omega) = \{0, 1, ..; \}$ and*

$$p_X(k) = \frac{e^{-\alpha}\alpha^k}{k!} \text{ for } k = 0, 1, ...$$

*Then we say that X follows a Poisson law with parameter $\alpha > 0$. We write $X \sim Poi(\alpha)$.*

**Example 2.3.1**

*We assume that the number X of errors in a program submitted for the first time to a central computer by each student from a fixed group is a rv following a Poisson law with parameter $\alpha = 3$. We assume that the submitted programs are independent from each other.*

*a) What is the probability that an arbitrary program does not contain any error?*

*b) What is the probability that among 20 submitted programs, there are at least 3 without any error?*

*Solutions:*

*a) We look for $P(X = 0) = e^{-3} \simeq 0,05$ (note that this is an approximation, as X cannot follow a Poisson law, as the number of errors in a programm cannot go to infinity).*

*b) Let Y be the number of programs, among the 20 submitted, without any error. Then Y* *follows a binomial law with parameters $n = 20$ and $p = P(X = 0) \simeq_{a)} 0,05$. We look for*

$$P(Y \geq 3) = 1 - P(Y \leq 2) \simeq 1 - \sum_{k=0}^{2} C_k^{20} (0,05)^k (0,95)^{20-k} \simeq 0,0754$$

⊙

**Example 2.3.2**

*In a lottery, 6 balls are drawn at random and without reset among 49 balls numbered from 1 to 49. We win a prize if the choosen combination has at least three good numbers. A player decides to buy a ticket for each play until he wins a prize. What is the probability that he has to buy at least 10 tickets?*

*Let M be the number of good numbers in the choosen combination. Let X be the number of tickets that the player will need to buy to get his first prize. We have*

$$P(M \geq 3) = 1 - P(M \leq 2) = 1 - \sum_{k=0}^{2} \frac{C_k^6 C_{6-k}^{43}}{C_6^{49}} \simeq 0,0186$$

*Then, we can write that $X \sim Geom(p = 0,0186)$. We look for*

$$P(X < 10) = P(X \leq 9) \simeq 1 - (1 - 0,0186)^9 \simeq 0,1555$$

⊙

## 2.4 Examples of continuous rv

### 2.4.1 Uniform Law

We assume that a number X is choosen at random in the interval $[a, b]$. In that case, the probability that X belongs to an interval of length $\varepsilon$, small enough around a point $x$, with $a < x < b$, should be the same for all $x$. Thus the density function of X should be constant. Since we must have a density, it follows that

$$f_x(x) = \frac{1}{b-a}, a \leq x \leq b$$

We then say that X follows an uniform law in the interval $[a, b]$ and we write $X \sim U(a, b)$.
We deduce that the distribution function of X is:

$$F_X(x) = 0 \text{ if } x < a, \frac{x-a}{x-b} \text{ if } a \leq x \leq b, 1 \text{ if } x > b$$

In particuliar, we find also that if $[c, d] \subset [a, b]$, then

$$P(c < X \leq d) = \frac{d-c}{b-a}$$

Thus the probability that X belongs to a given sub interval does depend only on the length of this interval.

### 2.4.2 Exponential law

We say that a rv X defined on $[0,+\infty)$ follows an exponential law with parameter $\lambda$ if

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$$

We write $X \sim Exp(\lambda)$.

**Remark 2.4.1**

*Note that this is indeed a density. On the other hand, the associated distribution function is given by*

$$F_X(x) = e^{-\lambda x} \text{ if } x \geq 0 \text{ and } 0 \text{ if } x < 0$$

*In particular*

$$P(X > 0) = e^{-\lambda x}, x \geq 0$$

**Example 2.4.1**

*We assume that the lifetime X (in years) of a device follows an exponential law with parameter $\lambda = 1/3$. What is the probability that a device of 3 years old still works after three consecutive additional years?*

*Here, we look for*

$$P(X > 6|X > 3) = P(X > 6-3) = e^{-1} \simeq 0,3679$$

*Note that the answer would be the same whether the device would be new or 6 years old ... If we apply this property to the lifetime of a car, it is clear that this assumption is not realistic. But it can be accepted for a very short period.* ⊙

### 2.4.3 Gamma Law

Let us first introduce the Gamma function, denoted by $\Gamma(.)$

$$G(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

for $\alpha > 0$. Performing an integration by parts, we find that

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1), \text{ if } \alpha > 1$$

In particular, we find that if $\alpha = n = 2, 3, ...$, then

$$\Gamma(n) = (n-1)!$$

One can also show that

$$\Gamma(1/2) = \sqrt{\pi}$$

Now, let X be a positive rv. If its density function is of the form

$$f_X(x) = \frac{(\lambda x)^{\alpha-1}\lambda e^{-\lambda x}}{\Gamma(\alpha)} \text{ for } x \geq 0,$$

we say that X follows a Gamma law with parameters $\alpha > 0$ and $\lambda > 0$.

We write $X \sim G(\alpha, \lambda)$.

**Remark 2.4.2**

*i) Parameter $\alpha$ is a shape parameter, while $\lambda$ is a scale parameter. As $f_X$ has a shape which varies rapidly when parameter $\alpha$ takes different values, Gamma law is often used in practise.*

*ii) For $\alpha = 1$, we recover the exponential law.*

*iii) Case $\alpha = n/2, n \in \mathbb{N}$ and $\lambda = 1/2$.*

*Then Gamma law is also called khi-square law with $n$ degrees of freedom.*

### 2.4.4 Gaussian law

Let X be a rv with $X(\Omega) = \mathbb{R}$. If the density function of X is of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, x \in \mathbb{R}$$

we say that X follows a gaussian or normal law with parameters $\mu$ and $\sigma^2$ with $\sigma > 0$.

We write $X \sim N(\mu, \sigma^2)$.

It is also called Laplace-Gauss law.

One can check that $f_X$ is a density. Parameter $\mu$ is a position parameter and $\sigma$ a scale parameter. All gaussian laws have a bell shape.

One can check that $f_X$ is symmetric wrt $\mu$, that is

$$f_X(x+\mu) = f_X(-x+\mu)$$

and moreover $f_X$ has its max at $x = \mu$ and have two inflexion points at $x = \mu \mp \sigma$.

If $\mu = 0$ and $\sigma = 1$, we say that X follows a standard normal law. Its density function is given by

$$\phi(z) \equiv \frac{1}{\sqrt{2\pi}} exp(-z^2/2), z \in \mathbb{R}$$

and its distribution function denoted by $\Phi$, is

$$\Phi(z) = \int_{-\infty}^{z} \phi(y)dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2}dy$$

If $X \sim N(\mu, \sigma^2)$, we find that its distribution function can be expressed through $\Phi$:

$$F_X(x) = \Phi(\frac{x-\mu}{\sigma})$$

and that its density function is given by

$$f_X(x) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$$

That is we may obtain everything from the law $N(0,1)$.

**Remark 2.4.3**

*It is useful to know some properties of these two functions. First of all, one can show that*

$$\int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}}$$

*Moreover, as $\phi$ is an odd function, we get*

$$\Phi(-x) = 1 - \Phi(x)$$

To estimate $\Phi$, we can use tables or programs. For example, a table will give the value of $\Phi(z)$ for $z = 0,00...$, $(0,01)$, ...., $3,99$. By symmetry, we have $\Phi(-z) = 1 - \Phi(z)$. Starting from $z = 3,90$, we may write $\Phi(z) \simeq 1,0000$. With more precisions, we find that

$$\Phi(4) \simeq 0,99997, \Phi(5) \simeq 0,9999997 \text{ and } \Phi(6) \simeq 0,999999999$$

Finally, note that function $\Phi$ is usuall expressed in terms of the error function

$$erf(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-y^2/2} dy = \Phi(z) - \frac{1}{2}$$

We shall see later on that the gaussian law is often used due to the central limit theorem.

**Example 2.4.2**

*Assume that the lifetime $X$ of a CPU follows a gaussian law with parameters $\mu$ and $\sigma^2$. We can then compute, for $k = 1, 2, ...$*

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) =_{cont} F_X(\mu + k\sigma) - F_X(\mu - k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

*which is independent of $\mu$ and $\sigma$. For example, if $k = 1$, we find that $\Phi(1) \simeq 0,8413$ so that the sought probability is nearly $0,683$.*

⊙

## 2.5 Expectation and variance

**Definition 2.5.1**

*The expectation or the mean of a rv $X$, denoted by $E(X)$ or by $< X >$ is defined by (if meaning-*

*ful)*

$$E(X) \equiv \mu_X = \begin{cases} \sum_{k=1}^{\infty} x_k p_X(x_k) \text{ if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx \text{ if } X \text{ is continuous} . \end{cases}$$

**Example 2.5.1**

*For example, if X is uniform on the interval* $(x_1, x_2)$*, we find*

$$E(X) = \frac{x_1 + x_2}{2}$$

**Example 2.5.2**

*Cauchy rv A warning light is fixed at point G, emits a light flash in the random direction uniform with angle* $\theta$*. We look for the abscissa X of the impact point of this beam on an infinite plane screen located at a distance* 1 *from point G.*

*By assumption, the angle* $\theta$ *is a rv with density* $q(\theta) = \frac{1}{\pi}\mathbb{I}_{]-\pi/2,+\pi/2[}(\theta)$*. The abscissa is given by* $X = \tan\theta$*. We deduce that its distribution function is given by*

$$F(x) = P(X \leq x) = P(\theta \leq \arctan x) = \int_{-\infty}^{\arctan x} q(\theta)d\theta = \frac{1}{\pi}\arctan x + \frac{1}{2}.$$

*Its derivative is given by*

$$f(x) = \frac{1}{\pi(1+x^2)}, \; x \in \mathbb{R}$$

*and thus X is a so called Cauchy rv, Note that* $E(X) = +\infty$*, that is X has no expectation.*

On can show that the expectation is a linear operation and that moreover, if $a$, $b$ and $c$ are constants, we have

$$E(c) = c \text{ et } E(aX + b) = aE(X) + b$$

**Example 2.5.3**

*Let* $X \sim Poi(\alpha)$*. Then*

$$E(X) = ... = \alpha$$

⊙

**Example 2.5.4**

*Let* $X \sim Exp(\lambda)$*. Then*

$$E(X) = \frac{1}{\lambda}$$

⊙

We have

**Proposition 2.5.1**

*Let $X$ be a rv and $Y = g(X)$. Then we have*

$$E(Y) = \begin{cases} \sum_{k=0}^{\infty} g(x_k) p_X(x_k) \text{ if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx \text{ if } X \text{ is continuous} \end{cases}$$

**Example 2.5.5**

*Let $X \sim U(0,2)$. Then we have*

$$E(X^2) = \int_0^2 x^2 f_X(x) dx = ... = 4/3.$$

⊙

**Definition 2.5.2**

*The variance of a rv $X$ is defined (if meaningful)*

$$Var(X) = \sigma_X^2 = E((X - E(X)^2)$$

This is a positive quantity. This measures the variability of $X$ around its mean. The variance is zero iff X is a constant (...).

If $X$ has a density $f_X$, then we have

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$$

with $\mu = E(X)$.

Standard deviation is defined by

$$STD(X) = \sigma_X = \sqrt{Var(X)}$$

One can show that

$$Var(aX + b) = a^2 Var(X)$$

et

$$Var(X) = E(X^2) - (E(X))^2$$

**Example 2.5.6**

*Let $X$ following a Bernoulli law with parameter $p$. We have*

$$E(X^n) = 1^n p + 0^n q = p, \text{ for all integer } n \geq 1$$

*Then*

$$Var(X) = pq$$

⊙

**Example 2.5.7**

*Let $X \sim U(a,b)$. Then*

$$E(X) = \frac{a+b}{2} \text{ and } Var(X) = \frac{(b-a)^2}{12}$$

$\odot$

**Example 2.5.8**

*The density of a normal law $N(\mu,\sigma^2)$ is given by*

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

*Let us show that the parameters are resp. the mean and variance of this law. As $f$ is symmetric wrt to $\mu$, then $E(X) = \mu$. As the area under $f$ is 1 (this is a density), we have*

$$\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma\sqrt{2\pi}$$

*Taking derivative wrt $\sigma$, multiplying the result by $\sigma^2/\sqrt{2\pi}$, then $E((X-\mu)^2) = \sigma^2$.* $\odot$

**Example 2.5.9**

*If $\sim Poi(a)$, that is X takes on values 0, 1, ... with*

$$P(X = k) = e^{-a} \frac{a^k}{k!}$$

*then*

$$E(X) = a,\ E(X^2) = a^2 + a \text{ et } \sigma_X^2 = a$$

*using series theory.*

$\odot$

Table for some discrete rv

| Law | Parameters | Mean | Variance |
|---|---|---|---|
| Bernoulli | $p$ | $p$ | $pq$ |
| Binomial | $n$ and $p$ | $np$ | $npq$ |
| Geometric | $p$ | $1/p$ | $q/p^2$ |
| Poisson | $\alpha$ | $\alpha$ | $\alpha$ |

Table for some continuous rv

| Law | Parameters | Mean | Variance |
|---|---|---|---|
| Uniform | $[a,b]$ | $(a+b)/2$ | $(b-a)^2/12$ |
| Exponential | $\lambda$ | $1/\lambda$ | $1/\lambda^2$ |
| Gamma | $\alpha$ and $\lambda$ | $\alpha/\lambda$ | $\alpha/\lambda^2$ |
| Gaussian | $\mu$ and $\sigma^2$ | $\mu$ | $\sigma^2$ |

We have just seen that the variance $\sigma^2$ enables to measure the concentration of a rv X around its mean $\mu$. The next result shows that in fact the probability that X is out a an arbitrary interval $(\mu-\varepsilon, \mu+\varepsilon)$ is negligible, if the ratio $\sigma/\varepsilon$ is sufficiently small.

## 2.6 Some inequalities

**Theorem 2.6.1**

*Bienaymé-Tchebychev inequality* *Let X be a rv with mean* $E(X)=\mu$ *and variance* $Var(X)=\sigma^2$. *Then for all* $a>0$, *we have*

$$P(|X-\mu|\geq a)\leq \frac{\sigma^2}{a^2}$$

**Remark 2.6.1**

*1. We can deduce that if* $\sigma=0$, *then the probability that X is out of an interval* $(\mu-\varepsilon, \mu+\varepsilon)$ *is zero for all* $\varepsilon>0$. *Thus* $X=\mu$ *with probability* 1. *Similarly if* $\mu=0$ *and* $E(X^2)=\sigma^2=0$, *then* $X=0$ *with probability* 1.

*2. In fact, it appears that the bound given by BT is too large. For example, if X is a normal rv, then*

$$P(|X-\mu|\geq 3\sigma)=2-2\Phi(3)\simeq 0,0027$$

*On the other hand, BT gives* $P(|X-\mu|\geq 3\sigma)\leq 1/9$. *The interest of BT is in fact that it is true for any density* $f$, *without any knowledge of it.*

**Theorem 2.6.2**

*Markov inequality* *Let X be a positive rv. Then for all* $a>0$

$$P(X\geq a)\leq \frac{E(X)}{a}$$

## 2.7 Additional Sections: Why the definition of expectation (can be skipped)

Though this definition is quite intuitive, the theoretical idea is to construct a notion of integration of this rv, wrt to the probability.

The key idea is to work over the target space, here $\mathbb{R}$, instead of the too much abstract space $\Omega$. If one wants to define a notion of expectation, that is a sort of integration, the key idea is to divide the target space instead of the state space. This is one of the fundamental differences with Riemann integration theory, which gives in particular the notion of Lebesgue integration.

As in the Riemann case, we start with much more simple rv, the step rv. A step rv $X$ is a rv which only takes a finite number of values $a_1$, ...,$a_p$.

For this kind of rv, we define the expectation by the formula

$$E(X) \equiv \sum_{i=1}^{p} a_i P(X = a_i).$$

Then we consider the case of general (and good) positive rv $X$.

In that case, we consider a sequence of step rv $X_n$ which approximates $X$ in an increasing way. For example, one can take

$$X_n(\omega) = k2^{-n} \text{ if } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{ and } 0 \leq k \leq n2^n - 1, \text{ and } n \text{ otherwise.}$$

As $X_n \leq X_{n+1}$, we deduce that $E(X_n) \leq E(X_{n+1})$, and thus we can write the following definition

$$E(X) \equiv \lim_{n \to +\infty} E(X_n).$$

The above limit exists, eventually being infinite, as we work whit increasing sequences. One can show that this limit does not depend on the specific sequence which approximates $X$ [of course there are many mathematical details that we skip here]. Then, if $X$ is any (good) rv with an arbitrary sign, we write $X = X^+ - X^-$, wher $X^+ = \sup(X, 0)$ and $X^- = \sup(-X, 0)$. Then we have $|X| = X^+ + X^-$, and $X^+$ et $X^-$ are positive rv.

In that case, we say that the rv $X$ is integrable iff $E(X^+)$ and $E(X^-)$ are two finite numbers. In that case, by definition, the expectation of $X$ is the number

$$E(X) \equiv E(X^+) - E(X^-)$$

which is also denoted by

$$E(X) \equiv \int_\Omega X(\omega) dP(\omega).$$

Note carefully that all these facts depend on the initial probability $P$.

What is the link with the definition of the expectation given previously?

Let us keep in mind the definition of the expectation in the continuous case, and assume we have a Riemann integral. We divide the axis $x$ into small intervals $(x_k, x_{k+1}))$ of common length $\Delta x$. If $\Delta x$ is small enough, and considering that we have a Rieman integral in the definition of the expectation, it can be approximated by a sum

$$\int_{-\infty}^{+\infty} \sum_{k=-\infty}^{\infty} x_k f(x_k) \Delta x$$

As

$$f(x_k)\Delta x \simeq P(x_k < X < x_k + \Delta x)$$

we get that

$$E(X) \simeq \sum_k x_k P(x_k < X < x_k + \Delta x)$$

And when $\Delta x \to 0$ it goes to

$$E(X) = \int_\Omega X(\omega) dP(\omega)$$