Basic Statistical Analysis in Life and Environmental Sciences

Chapter 1 - Basic notions of probability theory

Rodrigo Labouriau [1]

2022

[1] Applied Statistics Laboratory, Department of Mathematics, Aarhus University.

Contents

Chapter 1

Basic notions of probability theory

R.Labouriau [1]
Draft - Please do not circulate. [2]

1.1 Basic probability

1.1.1 A simple example

One of the simplest experiments is a trial with only two possible outcomes, called binary trial. An example of binary trial is a game in which there are two boxes, one containing an object and one empty. We do not have any information on which box contains the object. The experiment consists in choosing one of the boxes, opening it and verifying whether the object is in the box or not. We say that the result of this experiment is a *random quantity* because it is impossible to predict it with certainty.

Another classical example of binary trial is an experiment consisting on tossing a coin and observing the result: head or tail. In principle, it would be

[1] Applied Statistics Laboratory, Department of Mathematics, Aarhus University.
[2]Last revised: February, 2022.

possible to predict the result of such experiment, provided it is known with a good precision a range of conditions at which the experiment was performed, such as: the force used to throw the coin, the weight and the symmetry of the coin, the magnitude of the air resistance, and so on. However, in practice these conditions are not accessible and we must admit that it is not feasible to calculate with certainty the result of such experiment. Therefore, we say that the result of this experiment is random.

In both of the experiments described above we cannot predict exactly the result, but we can still observe some regularity when we perform the experiment sufficiently many times. The probabilistic and statistical tools we will study will take advantage of this regularity in order to obtain information about the phenomena we observe. For example, if we toss the coin many times, the ratio between the number of times we observe a head and the number of times the trial is performed, tend to stabilize around a certain value. In other words, the relative frequencies of events approach a limit when the number of trials increases. This suggests that there is an intrinsic characteristic common to all the realizations of the experiment, the plausibility of the event in play. Moreover, the limit of the relative frequency gives a measure of how likely, or how probable, is an event. This limit is referred as the *probability* of the event. [3]

When discussing binary experiments it is convenient to call one of the possible outcomes "success" and the other "failure". Here the choice of what is success and what is failure is completely arbitrary. The mathematical structures that arises from one choice or another are absolutely equivalent. This experiment will be called the *basic binary trial.* We speak of the probability of success, which we denote here by p, and the probability of failure in the binary trial, say q. A little of reflection shows that the probability of failure is equal to one minus the probability of success, *i.e.*, $q = 1 - p$.

Let us consider now a slightly more complex experiment. The experiment

[3] This is the so called frequencist notion of probability. There are other ways of defining probability.

consists in performing twice the basic binary trial with probability of success, p, and then register how many times success is observed. Then there are three possible results for the experiment: 0, 1 and 2. This experiment will be referred as the *binomial experiment*. We will describe below how can this yet simple experiment be modelled. The idea is to calculate the probability of each of the possible outcomes of this experiment. This will illustrate several key notions of the probability theory.

It is convenient at this stage to represent the outcome of the composed experiment by a variable, say Y, taking the values 0, 1 or 2. Since the result of the experiment occurs at random, the variable Y takes also its values at random, *i.e.*, it is not possible to predict completely the value that the variable Y will take, before realizing the experiment. We call Y a *random variable*. The event "not observing any success" in our experiment is written $[Y = 0]$, the event event "observing one success" is denoted by $[Y = 1]$, etc. The probability of the event $[Y = 0]$ is denoted by $P(Y = 0)$ and so on.

Let us determine the probability of each of the possible results of the experiment described. First, we calculate the probability of $[Y = 2]$. Clearly $[Y = 2]$ occurs if and only if we observe success in both basic binary trials. We know that the probability of success in each binary trial is p and, according to the frequencist definition of probability, this number represents a limit which the relative frequency approaches. Now we need to make some assumptions that enable us to connect the probabilities of the two binary basic trials. Let us assume that the results of the first binary basic trial does not influence at all the result of the second basic binary trial [4] and vice-versa. In this case the relative frequency of the event "observing success in both binary basic trials" is given by the product of the relative frequencies of observing a success in the first and in the second basic binary trial. Therefore the relative frequency of the event "observing success in both binary basic trials" approaches the product of the probability of observing success in the first and the second basic trial. That is the probability of observing success in both basic trials is

[4] Think in the example of tossing a coin twice, would it then be a reasonable assumption?

the product of the probability of observing success in each of the trials. Here we come to an important notion. We say that two events are *independent* when the probability of observing both events at the same time is the product of the probabilities of each of the two events. We are then assuming that the events "observing a success in the first basic trial" and the event "observing a success in the second basic trial" are independent. Now it is easy to see that the probability of $Y = 2$ is given by

$$\begin{aligned} P(Y=2) &= P(\text{"success in the first trial"}) \cdot P(\text{"success in the second trial"}) \\ &= p^2. \end{aligned}$$

A completely analogous argument can be used to conclude that

$$P(Y = 0) = (1 - p)^2 \, .$$

Now, the calculation of the probability of the of $[Y = 1]$ requires a bit more of argumentation, since this event can occur in two different ways: first if one observes success in the first basic trial and failure in the second trial, or secondly when one observes failure in the first basic trial and success in the second. Let us call these two possibilities event A and B, respectively. Clearly, using the argument of independency of the results of the two basic binary trials, the probability of the event A is $P(A) = p \cdot (1 - p)$ and the probability of event B is $P(B) = (1 - p) \cdot p = p \cdot (1 - p)$. Now, the events A and B are *mutually exclusive*, *i.e.*, if A occurs then B does not occur and vice-versa. We denote that A and B are mutually exclusive by $A \cap B = \emptyset$. We write $A \cup B$ for describing the event "A occurs or B occurs". Using again the frequecist definition of probability as a limit of relative frequencies, it is easy to see that the if the events A and B are mutually exclusive, then the probability of A and B is the sum of the probability of A and the probability of B. This can be written in symbols in the following way: if $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$. Applying this principle to the calculation of $P(Y = 1)$ we obtain

$$P(Y = 1) = P(A \cup B) = P(A) + P(B) = 2p(1 - p) \, .$$

Optional reading We give now an instructive alternative way to calculate $P(Y = 1)$. Given an event, say C, its *complementary* event is the event "not occurring C ", which is denoted C^c. The reader is invited to use our frequencist definition of probability to argument that the probability of the complement of an event C is equal to one minus the probability of C, *i.e.*,

$$P(C^c) = 1 - P(C) .$$

It is not difficult to see that the event $[Y = 1]$ is equal to the complementary of the event "$[Y = 0]$ and $[Y = 2]$", *i.e.*,

$$[Y = 1] = \{[Y = 0] \cup [Y = 2]\}^c .$$

Moreover, the events$[Y = 0]$ and $[Y = 2]$ are mutually exclusive. Therefore, the probability of $[Y = 1]$ is given by

$$\begin{aligned} P(Y = 1) &= P(\{[Y = 0] \cup [Y = 2]\}^c) = 1 - P\left([Y = 0] \cup [Y = 2]\right) \\ &= 1 - (P(Y = 0) + P(Y = 2)) = 1 - p^2 - (1 - p)^2 \\ &= 2p(1 - p) . \end{aligned}$$

1.1.2 Another simple example with continuous variable

Let us consider another simple example of experiment with random results. Suppose that we perform the following experiment: one person places an object in a path between two locations. We dispose of no additional information about the position where the object is placed apart from the fact that the person has no special preference for the position of the object. The result of the experiment is the position of the object. Clearly, we cannot predict precisely the position of the object before the experiment is performed, therefore we say that the result of the experiment is a random quantity.

To simplify matters let us assume that the path has length one. The outcome of the experiment can be represented by a variable, say X, taking values at random in the interval $[0, 1]$ (*i.e.*, the set of numbers between 0 and 1). The typical question here is: What is the chance that the selected point lies in a certain region of the interval $[0, 1]$? For example, what is the chance that the point lies between 0 and 1/2? It is intuitively clear that, if all the regions of the regions of the intervals are equiprobable, then repeating the experiment many times, the chosen point will lie between 0 and 1/2 in approximately one half of the cases. That is, the relative frequency of the event "the chosen points lies between 0 and 1/2" approaches 1/2. As before, we say that the probability of the event "the chosen points lies between 0 and 1/2" is 1/2. Using a analog rationale we conclude that, if the region is one interval contained in $[0, 1]$, say $[a, b]$ (with $0 \leq a < b \leq 1$), then the probability that the chosen point is contained in $[a, b]$ is the length of this interval, *i.e.*, $b - a$. Now, if two intervals are disjoint (*i.e.*, have no elements in common), then the events "the point is contained in the first interval" and "the point is contained in the second interval" are mutually exclusive. Therefore the probability of the event "the point is contained in the first interval or in the second interval" is the sum of the probability of the event "the point is contained in the first interval" and the probability of the event "the point is contained in the second interval". That is, the sum of the length of the two intervals. If we have an enumerable collection of disjoint intervals, say $I_1, I_2, \ldots$, the probability of the union of all the intervals, is the sum (*i.e.*, the series) of the length of the intervals in the collection. Moreover, it is easy to see that the probability of the complement of an interval is one minus the length of the interval. Using these properties we can calculate the probability of many regions in the interval $[0, 1]$. The law of probability associated with this experiment is called the *uniform distribution* (on the interval $[0, 1]$).

1.1.3 The basic probabilistic model

In this section we collect together and generalise some of the results obtained for the two simple particular examples studied above.

The first step in the construction of a mathematical model for experiments with random results is to identify the set of all possible results of the experiment. This set is called the *sample space* and we denote it by Ω. For example, in the case of the binary trial the sample space was a set with two elements (termed "success" and "failure"). The sample space of the example of continuous variable was the interval $[0, 1]$.

The next step in the construction is to determine the class of events to which we will attribute probability. For example, in the experiment of choosing a point in the interval $[0, 1]$ the events are the intervals contained in $[0, 1]$, their complements and the (enumerable) union of intervals. Next we defined a probability of each event as the limit of the relative frequency of the event obtained by a large number of repetitions of the experiment. Using this frequencist notion of probability we could justify the following properties of the probabilities:

1. Given an event A, its probability is a positive number, *i.e.*, $P(A) \geq 0$;

2. The probability of the sample space is one, *i.e.*, $P(\Omega) = 1$;

3. Given a sequence of mutually exclusive (disjoint) events, say $A_1, A_2, \ldots$, the probability that one of the events occurs is the sum of the probability of each of the events, i.e. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.

These three intuitive properties were used by Kolmogorov as axioms for defining formally a probability. Indeed, it is possible to construct the whole theory of probability using only the three axioms listed above (and the notion of independency). [5]

We defined now the important notion of independency. Two events, say A and B, are *independent* if the probability that A and B occurs is the product of the probability of A and the probability of B, *i.e.*, $P(A \cap B) = P(A) \cdot P(B)$.

[5] See the exercise 1.1.

1.1.4 Random variables

The results of a probabilistic experiment as described before are usually registered by mean of a variable taking values at random, called a *random variable*. For example the random variables Y and X were used to describe the binomial experiment and the continuous variable example above. Using the close correspondence between the results of the probabilistic experiment and the associated random variable, it is possible to establish what is the probability of a random variable assumes a certain value or take values in a certain region. That is, we can attribute a *probability law* or *distribution* to random variables.

The way to describe the distribution of a random variable differs slightly according to the type of the variable. A random variable taking values in an countable set[6] is called a *discrete random variable* (*e.g.*, Y, is the results of counting something). In this case the law of probability of the random variable is described by a function that associates each possible value of the random variable with its probability. This function is called the *probability function* of the random variable. For example, in the binomial experiment the probability function is given by, for each possible value t,

$$f_Y(t) = P(Y = t) \begin{cases} p^2 & , \quad \text{if } t = 2, \\ 2p(1-p) & , \quad \text{if } t = 1, \\ (1-p)^2 & , \quad \text{if } t = 0\,. \end{cases} \tag{1.1}$$

A random variable X is called *absolutely continuous* or simply *continuous* if there is a function f taking positive values (*i.e.*, $f(x) \geq 0$, for each x) such that, for each real number x,

$$P(X \leq x) = \int_{-\infty}^{x} f(x)dx\,. \tag{1.2}$$

For the reader not acquainted with the integration, the right hand sign in the expression above is the area between the graph of the function f and the

[6]A countable set is a set is finite or can be enumerated; *e.g.*, the set $1, 2, 3, \ldots$ is enumerable but the set of the real numbers is not (you cannot count them).

horizontal axis, measured up to the point x. The function f is called the *probability density of* X, or in short the *density of* X.

In the example of the uniform distribution (*i.e.*, the example on continuous variable) the probability density of the random variable Y is the function

$$f(y) = \begin{cases} 1 \quad , & \text{if } 0 \leq y \leq 1\,; \\ 0 \quad , & \text{otherwise.} \end{cases}$$

The calculation of the probability of Y takes a value between 0 and 1/2 can be alternatively calculated in the following way:

$$P([0 \leq Y \leq 1/2]) = P([Y \leq 1/2]) = \int_{-\infty}^{1/2} f(y)dy = \int_{-\infty}^{1/2} dy = 1/2\,.$$

Now, changing the form of the density, the law of probability changes also. Here there are two restrictions: the density function should assume only non-negative values and should integrate 1 (that is, the area between the graph of the density and the horizontal axis should be 1).

Another important function related to the law of probability of a random variable X is the *(cumulative) distribution function*,say F, which associates to each real number x the probability that the random variable X is less or equal to x. In symbols, for each x,

$$F(x) = P(X \leq x)\,.$$

It can be shown that (but we will not do that here!) the distribution function of a random variable characterizes its distribution.

It is possible to define the notion of independency of two random variables in a somehow similar way as we did for the independency of events. This is a very important notion because several forms of independency of the observations will be a customary assumption in the models that we will study. The intuitive idea behind the notion of independency is that, if the knowledge of the value taken by a random variable does not affect the distribution of another random variable, then they are independent. Formally we say that the random variables X and Y are independent when, for each par of numbers x and y,

$$P(X \leq x \text{ and } Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \ .$$

That is, the event "X is less or equal x" is independent of the event "Y is less or equal y", for each x and each y.

1.1.5 Expectation and variance of random variables

In this section we briefly present the notions of expectation and variance of a random variable.

The idea of expectation can be easily understood for discrete variables taking a finite number of values. Suppose that X is a random variable taking n values, $x_1, x_2, \ldots, x_n$, with probabilities $p_1, p_2, \ldots, p_n$, respectively. The *expectation* or *expected value* of X is the sum of the possible values of X multiplied by their probabilities. We use the symbol $\mathrm{E}(X)$ to denote the expectation of X and write

$$\mathrm{E}(X) = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n \ .$$

The expectation can be interpreted in the view of the intuitive frequencist approach in the following way: If the random variable X represents the result of an experiment, suppose that we (at least hypothetically) repeat independently the experiment r times, obtaining many values for the random variable X. If the number of repetitions r is very large, then the variable X will take a value x_i with relative frequency approximately equal to the probability of the result x_i, say p_i. That is, x_i will appear approximately rp_i times in the r observations. Therefore, the mean value of the observed results in the r repetitions will be approximately

$$\frac{1}{r}\left(rp_1 x_1 + rp_2 x_2 + \ldots + rp_n x_n\right) = \left(p_1 x_1 + p_2 x_2 + \ldots + p_n x_n\right) \ ,$$

which is the expectation of X. In other words, the mean value of the results obtained in r hypothetical independent repetitions of the experiment approaches the number $\mathrm{E}(X)$, as r increases. This is a rough version of the *law of large numbers*, a classic result of the probability theory[7].

In the example of the binary trial the random variable X takes the values 0 and 1 with probabilities $(1-p)$ and p, respectively. The expectation of X is then

$$\mathrm{E}(X) = (1-p)0 + p1 = p \ .$$

In the example of the binomial variable X, taking the values $0, 1$, and 2 with probabilities $(1-p)^2$, $2p(1-p)$ and p^2, the expectation is

$$\mathrm{E}(X) = (1-p)^2 0 + 2p(1-p)1 + p^2 2 = 2p \ .$$

In the case of a continuous variable Y with probability density f the expectation is given by

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f(y) dy \ .$$

The law of large numbers is (under very general assumptions) also valid for continuous variables. Therefore, we have essentially the same interpretation of the expectation for continuous variables. It must be remarked, however, that it is not always that the expectation exists. There are examples of random variables without expectation, *i.e.*, for which the integral or the series involved in the definition of expectation is not convergent.

There is another intuitive interpretation of the expectation. This comes from an analogy with the mechanics. Indeed, the integral used in the definition of expectation is analogue to the centre of mass of a body with the distribution of masses described by the probability density function (see Figure 1.1).

In the example of the uniform distribution the expectation is calculated as follows:

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f(y) dy = \int_0^1 y dy = \frac{y^2}{2} \Big|_0^1 = \frac{1}{2} \ .$$

[7]Note that we only gave an informal intuitive argument here. It is possible to define the law of large numbers precisely and to prove that as a real theorem.

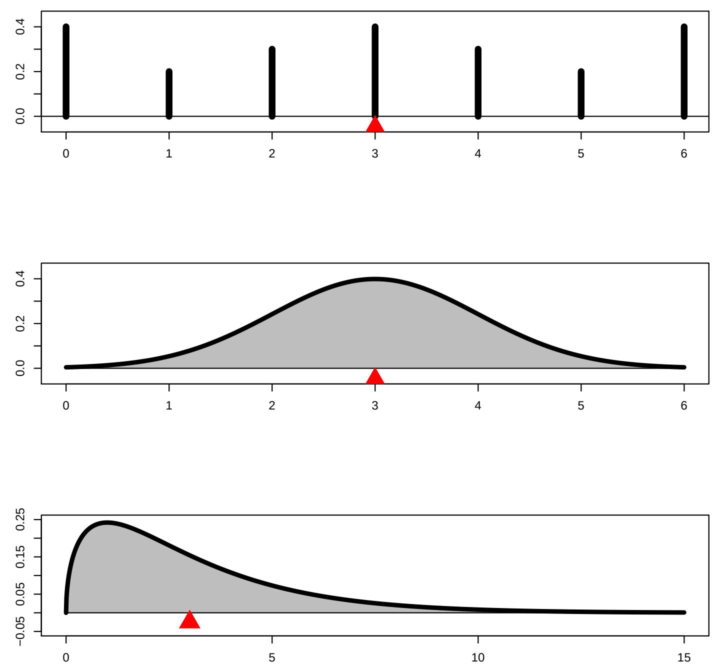


Figure 1.1: Analogy between the expectation and the centre of mass. Above, a discrete variable taking values $0, 1, \ldots, 6$; imagine that masses were put in a bar at the positions $0, 1, \ldots, 6$ proportional to the probability of observing the corresponding value; since the distribution is symmetric around the value 3, the bar will be in balance if we put a keel beneath the position corresponding to the value 3 (which plays the rule of the expectation). Middle, a continuous distribution symmetric around 3; imagine that a would shape is constructed with the form of the graph of the density; to keep the shape balanced a keel should be placed beneath the position corresponding to the value 3 (the expectation). Below, a continuous asymmetric distribution; the keel should be placed in a position such that the shape is kept in balance.

Using the analogy with mechanics, it is clear that $1/2$ is the center of mass of a body with mass uniformly distributed in the interval $[0, 1]$.

The expectation has the following basic properties:

1. If the random variable X is equal to a constant c with probability 1, then $\mathrm{E}(X) = c$;

2. If X and Y are random variables (with expectation well defined) and a, b are constants, then $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$;

3. If X and Y are random variables (with expectation well defined) such that $X \leq Y$ with probability 1, then $\mathrm{E}(X) \leq \mathrm{E}(Y)$.

4. (Jensens inequality) If ϕ is a convex real function and X is a random variable with finite expectation, then

$$\mathrm{E}\left\{\phi(X)\right\} \geq \phi\left\{\mathrm{E}(X)\right\} .$$

The notion of variance The expectation gives an idea of the position of a central value of the random variable. We introduce next the notion of variance, which will give an idea of how much disperse are the values of the random variable. The *variance* of a random variable X is defined by

$$\mathrm{Var}(X) = \mathrm{E}\left\{X - \mathrm{E}(X)\right\}^2 = \mathrm{E}(X^2) - \left\{\mathrm{E}(X)\right\}^2 .$$

Clearly, $\left\{X - \mathrm{E}(X)\right\}^2$ is a measure of the distance between the random variable X and its expectation. Therefore, the expected value of this distance, *i.e.*, the variance, is a measure of the dispersion of the data around its expected value. The larger is the variance the more disperse is the data.

The variance of the binary variable X taking values 0 and 1 with probabilities $(1.p)$ and p is

$$\mathrm{Var}(X) = \mathrm{E}\left\{X - \mathrm{E}(X)\right\}^2 = \mathrm{E}\left\{X - p\right\}^2 = \ldots = \mathrm{E}(X^2) - p^2 .$$

To complete the calculation above we must compute the expectation of the random variable X^2. Note that $X^2 = X$, since X takes only the values 0 and 1. Therefore $\mathrm{E}(X^2) = \mathrm{E}(X)$. Replacing that in the last equation yields

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - p^2 = p - p^2 = p(1-p) \ .$$

In the case of the uniform variable Y the variance can be calculated in the following way

$$\mathrm{Var}(Y) = \int_{-\infty}^{\infty} (y - \mathrm{E}(Y))^2 f(y) dy = \int_0^1 (y - 1/2)^2 dy = \ldots = \frac{1}{12} \ .$$

The variance has the following basic properties:

1. If the random variable X is equal to a constant with probability 1, then $\mathrm{Var}(X) = 0$;

2. If the random variable X has finite variance and b is a constant, then $\mathrm{Var}(bX) = b^2 \mathrm{Var}(X)$;

3. If the random variables X and Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

1.2 Probability distributions

We consider next some useful examples of probability distributions. The distributions described below will depend on some parameters. That is, there will be a list of numbers, called parameters (typically one or two), such that for each combination of values of them, there will correspond one (and only one) distribution. The most famous example is the normal distribution, which is determined by two parameters: the mean and the variance (note that one could have used other parametrizations!). We say that a collection of distributions indexed by a set of parameters is a *family of distributions*. One should say then "the family of the normal distributions" when talking about the normal distributions in general and "the normal distribution with mean 0 and variance 1" when talking about this specific distribution. However, it is a common abuse of language to say "**the** normal distribution" when talking about the whole family of normal distributions. Since there are only little chances for confusion we will adopt this abuse of language and write "the binomial distribution", "the Poisson distribution" and so on.

1.2.1 Some common key distributions

The next examples will be families of distributions sharing some mathematical properties that will characterize them as "exponential dispersion models". We adopt a reversed order here presenting first the examples and defining what we mean by exponential dispersion models in the next section (which is an optional reading).

The binomial distribution The binomial distribution is related to the experiment: perform independently n times a basic binary trial with probability p of success. The distribution of the number of successes, say X, is a random variable distributed according to a binomial distribution. The binomial distribution is discrete (taking the values 0,1, ...,n) and depends on two parameters: n and p. The probability function of X is given by

$$\begin{aligned} P(X=x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} , \qquad \text{for } x = 0, 1, \ldots, n \ . \end{aligned}$$

Here $\frac{n!}{x!(n-x)!}$ is the number of subsets with x elements of a set with n elements, called the binomial coefficient. It is customary to write $X \sim Bi(n, p)$, when the random variable X is binomially distributed with parameters n and p. It is not difficult to show that if $X \sim Bi(n, p)$, then the expectation and the variance of X are

$$\mathrm{E}(X) = n\,p \quad \text{and} \quad \mathrm{Var}(X) = n\,p\,(1-p) \ .$$

Models based on the binomial distribution will be studied in chapter **??**.

The Poisson distribution The Poisson distribution is a discrete distribution taking non-negative integer values (*i.e.*, 0,1,2,3, ...). Many models for counting are based on the Poisson distribution. The probability function of a random variable X depends on one parameter, $\lambda > 0$, and is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \text{ for } x = 0, 1, 2, \dots \ .$$

We write $X \sim Po(\lambda)$ when X is Poisson distributed with parameter λ. The expectation and the variance of a Poisson distributed random variable are both equal to the parameter λ, *i.e.*, if $X \sim Po(\lambda)$, then

$$\mathrm{E}(X) = \mathrm{Var}(X) = \lambda \ .$$

Models based on the Poisson distribution will be studied in chapter **??**.

The normal distribution The normal distributions constitute, with no doubt, the most famous family of distributions. It is a family of continuous distribution depending on two parameters, μ and σ^2 and probability density given by, for each real number x,

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} .$$

Here μ is a real number and σ is a positive number ($\sigma > 0$). If a random variable X is normally distributed with parameters μ and σ^2, we write $X \sim N(\mu, \sigma^2)$ and in this case X has expectation and variance given by

$$\mathrm{E}(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2 \ .$$

Some models based on the Normal distribution will be studied in chapter **??**.

The gamma distribution* The gamma distribution is a continuous distribution useful for analyzing non-negative data. The family of the gamma distribution is indexed by two indices, μ and λ, both positive. The density of a gamma distributed variable X with is given by

$$f(x; \mu, \lambda) = \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\mu}\right) x^{\lambda-1} \exp\left(-\frac{\lambda}{\mu}x\right) \quad \text{for } x \geq 0 \ .$$

The mean and the variance of the a gamma distributed random variable X with parameters μ and λ are

$$\mathrm{E}(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \frac{\mu^2}{\lambda} \; .$$

The parameter μ is usually called the mean parameter. On the other hand, the parameter λ is called the shape parameter, since the form of the density depends on μ. Indeed, if $0 < \lambda < 1$, then the density decreases as x increases and increases indefinitely as x tends to zero (see Figure **??**). If $\lambda = 0$, the distribution is called the exponential distribution (do not misunderstand with the exponential family!) and the density decreases as x increases. If $\lambda > 1$ then the density takes the value 0 at $x = 0$ is positive skewed and has mode (ie. the assumes its highest value) at $x = \mu - \mu/\lambda$. The density approaches the density of a normal distribution when λ tends to infinity.

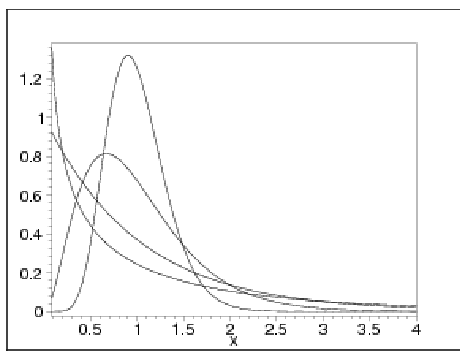


Figure 1.2: Probability density of a gamma distribution with $\mu = 1$ and $\lambda = 0.5, 1, 3$ and 10.

The inverse gaussian distribution * Another example of family positive continuous distribution are the inverse gaussian distributions. These distributions appear in physics as the distributions for the time for a random walk hits a barrier. They have been used for modelling particle movements, survival times, duration of strikes, pharmaco-kinetic and remote sensing studies, etc. The inverse gaussian distributions have a probability density depending on two parameters μ and λ given by

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left\{-\lambda\frac{(x-\mu)^2}{2\mu^2 x}\right\} \ , \quad \text{for } x \geq 0 \ .$$

Figure 1.3 shows the form of the density of inverse gaussian distributions with different values of the parameter λ. The mean and the variance of the a inverse gaussian distributed random variable X with parameters μ and λ are

$$\mathrm{E}(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \frac{\mu^3}{\lambda} \ .$$

The distribution of a random variable following an inverse gaussian distribution with parameters μ and λ approximates the normal distribution with mean μ and variance μ^3/λ, when λ increases.

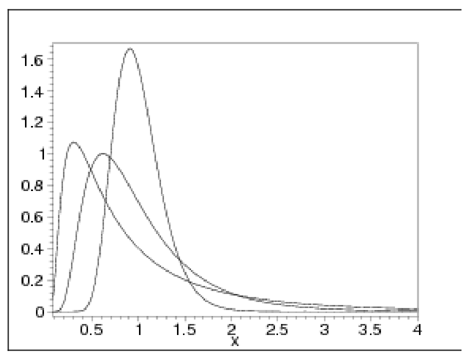


Figure 1.3: Probability density of an inverse gaussian distribution with $\mu = 1$ and $\lambda = 1, 3$ and 15 (from left to right).

1.2.2 Exponential dispersion models *

[8]

In the following we study the notion of exponential dispersion models. They are special families of distributions that have nice properties and that

[8]Optional section. Can be skipped in a first reading.

will be the basis for generalized linear models. A family of distributions indexed by two parameters, θ in Θ and λ in Λ for which the probability density or the probability function can be written in the form

$$f(x;\theta,\lambda) = \exp\left[\lambda\left\{x\theta - b(\theta) - c(x,\lambda)\right\}\right] \quad , \tag{1.3}$$

is called an *exponential dispersion model*. Here b and c are suitable functions, λ is a positive real number and θ is a real number. There are many examples of exponential dispersion models, among them all the families of distributions considered in the last section.

We write $X \sim ED(\theta,\lambda)$ to denote that the random variable X follows a distribution contained in an exponential dispersion model, which is specified by the parameters θ and λ. It can be shown that if $X \sim ED(\theta,\lambda)$, then its expectation and variance are given by

$$\mathrm{E}(X) = b'(\theta) \quad \text{and} \quad \mathrm{Var}(X) = \frac{b''(\theta)}{\lambda} \; .$$

Here, $b'(\theta)$ and $b''(\theta)$ are the first and the second derivative of the function b at the point θ, respectively.

Since the derivative of the function $b'(\theta)$ (i.e. $b''(\theta)$) is proportional to the variance of X, which in turn is a positive quantity, we conclude that the function b' is monotone and has an inverse, say $u = \{b'\}^{-1}$. That is, $u\{b'(\theta))\} = \theta$ (for each θ) and $b'\{u(\mu)\} = \mu$. In other words, there is a one-to-one correspondence between θ and the expected value of a random variable $X \sim ED(\theta,\lambda)$. Let us denote the expectation of X by μ.

The function $V(\mu) = b''\left[\{b'\}^{-1}(\mu)\right] = b''\left\{u(\mu)\right\}$ is called the *variance function*. This function plays a fundamental role in the theory of exponential dispersion models, as will be clear from the following.

It will be convenient to introduce a new parametrization of the family of distributions $ED(\theta,\lambda)$ by using the parameters $\mu = u(\theta)$ and $\sigma^2 = \frac{1}{\lambda}$. Doing so we can express the expectation and the variance of $X \sim ED(\theta,\lambda)$ as

$$\mathrm{E}(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2 V(\mu) \; .$$

Family	Symbol	Natural parameter	Scale parameter	Distributions support
Binomial	$Bi(n,p)$	$\log(p/1-p)$	1	$\{0,1,\ldots,n\}$
Poisson	$Po(\lambda)$	$\log(\lambda)$	1	$\{0,1,\ldots\}$
Normal	$N(\mu,\sigma^2)$	μ	σ^2	R
Gamma	$Ga(\mu,\lambda)$	$-\mu^{-1}$	λ^{-1}	R_+
Inv.Gauss.	$IG(\mu,\lambda)$	$-(2\mu^2)^{-1}$	λ^{-1}	R_+

Family	$b(\theta)$	μ	$V(\mu)$
Binomial	$n\log(1+e^\theta)$	$ne^\theta/(1+e^\theta)$	$n^{-1}\mu(n-\mu)$
Poisson	e^θ	e^θ	μ
Normal	$\theta^2/2$	θ	1
Gamma	$-\log(-\theta)$	$-\theta^{-1}$	μ^2
Inv.Gauss.	$-(-2\theta)^{1/2}$	$-1/\theta^{1/2}$	μ^3

Table 1.1: Basic characteristics of some standard exponential dispersion models.

Note that with this new parametrization the relation between the parameters and the expectation and the variance of X is very clear. The expectation is the parameter μ and the variance is proportional to the variance function applied to the parameter μ, the constant of proportionality being the parameter σ^2. The parameters μ and σ^2 are called the *mean* and the *scale* parameters, respectively.

It can be shown that the variance function characterizes the exponential dispersion model, *i.e.*, if V is the variance function for an exponential dispersion model, then there is no other exponential dispersion models with the same variance function. The main quantities related with some classic exponential dispersion models are given in Table 1.1.

1.2.3 Distributions derived from the normal

We will use very often three families of distributions that are related to the normal distribution: the chi-square, the F and t distribution. They will be used in the construction of confidence intervals and in many tests. Therefore we briefly review the definitions and very basic properties of them.

The chi-square distribution Suppose that X is a random variable following the normal distribution with mean 0 and variance 1. The random variable X^2 is positive valued with asymmetric distribution. This distribution is called the *chi-square distribution with 1 degree of freedom.* It can be shown that the expectation and the variance of a chi-square distribution with 1 degree of freedom are 1 and 2, respectively.

Now, suppose that $X_1, \ldots, X_\nu$ are ν *independent* random variables each of them distributed according to a normal distribution with mean 0 and variance 1. The distribution of the of the sum of the squares of $X_1, \ldots, X_\nu$, *i.e.*, $X_1^2 + \ldots + X_\nu^2$, is called the *chi-square distribution with ν degree of freedom.* This distribution is obviously positive, asymmetric and has mean ν and variance 2ν (why?). The chi-square distribution with ν degrees of freedom has density function given by

$$f(x;\nu) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\frac{nx}{2}\right) , \quad \text{for } x \geq 0 .$$

As ν tends to infinity (*i.e.* increases unlimited) the chi-square distribution tends to the normal distribution. This tendency is, however, very slow. A better approximation is: If Z is a chi-squared distributed random variable with ν degrees of freedom, then $(Z/\nu)^{1/3}$ is approximately normal distributed with mean $1 - 2/(9\nu)$ and variance $2/(9\nu)$, for large values of ν (see Kendall and Stuart, 1977, pp 399).

The F distribution Suppose that X_1 and X_2 are two independent random variables. Moreover, X_1 is assumed to be distributed according to a chi-square distribution with n degrees of freedom and X_2 is chi-square distributed with m degrees of freedom. The distribution of the random variable

$$\frac{X_1/n}{X_2/m}$$

is said to be the F *distribution with* n *and* m *degrees of freedom.*

The t distribution Suppose that X is a random variable distributed according to a chi-square distribution with n degrees of freedom and Z is a random variable normally distributed with mean zero and variance 1. Assume, moreover that X and Z are independent. Then, the distribution of the random variable

$$\frac{Z}{\sqrt{X/n}}$$

is said to be a t *distribution with* n *degrees of freedom.*

1.3 Some fundamental results in probability theory

Here we briefly discuss two general results of the probability theory: the law of large numbers and the central limit theorem.

1.3.1 Law of large numbers

The idea of the law of large numbers was already mentioned when we discussed the frequencist definition of probability. A rough form of the law of large numbers says that, if we repeat independently many times an experiment generating the same random variable, then the mean of the observed values approximates the expectation of the random variable.

We make the above statement of the law of large number above more precise. Suppose that $X_1, X_2, \ldots$ is a sequence of random variables independent and following the same distribution. We say that these random variables are *independent and identically distributed* (and some times denote that by *iid*). The Kolmogorov law of large numbers says that if X_1 has finite expectation μ (and hence the other random variables also have the same expectation) then with probability 1 the sequence of means converge to the expectation μ, *i.e.*

$$\frac{X_1 + \ldots + X_n}{n} \longrightarrow \mu \tag{1.4}$$

as $n \to \infty$ (*i.e.* as n increases arbitrarily).

The law of large numbers can be proved from the axioms of the probability, without using the simplified frequencist approach we used to define probability intuitively. [9]. On the other hand the frequencist interpretation of the notion of probability is a consequence of the law of large numbers. To see that consider the case where the random variables $X_1, X_2, \ldots$ are binomially distributed with $X_1 \sim Bi(1,p)$. In this case the expectation of each of the variables is $E(X_1) = 1.p = p$. Moreover, since the nth variable, X_n, of the binomial variables above takes the value 1 if the nth binomial trial is successful and 0 otherwise, then the sum $X_1 + \ldots + X_n$ is equal to the number of successes occurred in the first n binomial trials. Therefore, the left hand side of (1.4) is equal to the relative frequency of success in the first n trials, and the right hand side of (1.4) is the probability p of success in one trial. We conclude that the law of large numbers imply that the relative frequency of independent identical trials approximates to the probability of occurrence of an event with probability 1. A slightly weaker version of the law of large number for binomial trials [10] was published by one of the members of the Bernoulli family in the *Ars Conjectandi* in 1713. A stronger form of the law of large numbers, close to the version presented here was proved by Kolmogorov. There are other versions of the law of large numbers using less restrictive assumptions that can be used for approximations in more general contexts.

[9] This is far beyond the mathematical level of this text.

[10] Using a different form of convergence called convergence in probability.

1.3.2 Central limit theorem

The central limit theorem is another useful general approximation result of the theory of probability. Essentially this theorem says that under certain special circumstances the sum of random variables after a proper standardization follow approximately a standard normal distribution. One of the (many) versions of the central limit theorem states that if $X_1, X_2, \ldots$ are independent and identically distributed random variables for which $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$, where $0 < \sigma^2 < \infty$ (*i.e.* the variances are not zero and are not infinite), then

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

follows approximately a standard normal distribution (*i.e.* $N(0, 1)$), for n sufficiently large.

There other versions of the central limit theorem that use less restrictive hypotheses. The first version of the central limit theorem was proved first by De Moivre for binomial essays, latter Gauss proved the famous version of the central limit theorem for distribution of errors in measurements. The central limit theorem has been used, and misused in some cases, to justify the use of statistical models based on the normal distribution.

1.4 Exercises

Exercise 1.1 *The following properties of the probabilities were used by Kolmogorovas axioms to formally define probability:*

1. *Axiom 1: Given an event A, its probability is a positive number, i.e., $P(A) \geq 0$;*

2. *Axiom 2: The probability of the sample space is one, i.e., $P(\Omega) = 1$;*

3. *Axiom 3: Given a sequence of mutually exclusive (disjoint) events, say $A_1, A_2, \ldots$, the probability that one of the events occurs is the sum*

of the probability of each of the events, i.e. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.

1- *Please, give a justification of the three axioms of probability in terms of the (informal) frequencist definition of probability.*

2- *Show, using the axioms of probability that the probability of the complementary of an event is given by one minus the probability of the event, i.e.,* $P(A^c) = 1 - P(A)$.

Exercise 1.2 *Consider the simple experiment:"Toss a fair dice and observe the result".*

1- *If the event* A *is "observe a 5 or a 6", what is* A^c*? What is the probability of* A *and of* A^c*?*

2- *Define the events*
$B =$ *"observe a 6"*
$C =$ *"observe an odd number" and*
$D =$ *"observe an even number".*
Which pairs of these events are mutually exclusive? What is $P(B)$, $P(C)$, $P(D)$ *and* $P(B \text{ or } C)$*?*

3- *What is the probability of flipping a fair dice four times in a row and get a 6 each time?*

Exercise 1.3 *A simple experiment was performed by throwing two dices. The result of throwing the first dice and the second dice are represented by the random variables* X *and* Y *respectively. You may assume* X *and* Y *to be independent.*

1- *What is the probability that* X *is an odd number? What is the probability that* X *and* Y *are both odd numbers? Justify your calculations.*

2- *Calculate the expectation and the variance of* X *(i.e.,* $E(X)$ *and* $Var(X)$*).*

3- Calculate the expectation and the variance of the sum of X and Y (i.e., $E(X+Y)$ and $Var(X+Y)$). Justify your calculations.

4- Calculate the expectation and the variance of $2X+Y$.

Exercise 1.4 *Here we analyse the results of an experiment involving the height of maize plants. Two measurements were made per plant and analysed separately: 1) the height from the ground to the first leave insertion H_1 and the distance between the first leave insertion and the plant apice H_2, both expressed in centimetres. From the analyses used the expected value of the variables H_1 and H_2 were estimated to be 89.2 and 38.2, respectively. Moreover, the variance of H_1 and H_2 were estimated to be 2 and 1 respectively.*

1- After analysing the data and reported the analyses in a manuscript subject to publication one of the reviewer asked to express the measurements in millimetres (instead of centimetres). What would be then the estimates of the expected values and the variances when you convert to H_1 and H_2 to millimetres?

2- A second (nasty) reviewer asked you to estimate the expected value of the plant height (i.e., H_1+H_2). What would be your estimate?

3- [11] You estimated also the covariance between H_1 and H_2 as 1.5 (in the original scale in cm). Assuming that the covariance between H_1 and H_2 is 1.5, what would be then the variance of plat height (expressed in cm and expressed in mm)?
(Hint: $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$)

Exercise 1.5 *In this exercise we consider the number of worms found in a portion of one cubic decimetre of soil (1l). The number of worms encountered in a portion of soil was in mean 3.*

1- Is it reasonable to assume that the number of worms per portion of soil is Poisson distributed?

[11] Optional item.

2- *Assuming the number of worms per litre of soil to be Poisson distributed with expectation 3, what is the probability of observing no worms in a portion of 1l of soil? What is the probability of observing exactly one worm in a portion of 1l of soil? What is the probability of observing more than five worms in a portion of 1l of soil?*

3- *What is the distribution of the number of worms per cubic metre of soil?*

4- *What is the probability of finding no worms a cubic metre of soil?*

5- *Simulate an experiment where 30 independent repeated counts of the number of worms per litter of soil are performed (with different mean numbers of worms per litre of soil, say 1, 2 and 3 worms per litre). Calculate then the sample mean and the sample variance. Make a normal Q-Q-plot of the counts. Now, simulate the same experiment were the number of worms per cubic meter is counted. Calculate the sample mean and the sample variance and draw a Q-Q normal plot. (Hint: rpois(n= 30, lambda=1) simulates 30 samples from a Poisson distribution with parameter intensity $\lambda = 1$).*

Exercise 1.6 *In this exercise, we discuss a modelling problem that arises from cell culture. Suppose that we want to produce mono-clonal tissues (i.e., a tissue that arises from the multiplication of* **one** *single cell). A standard technique to do that is to prepare a suspension of the cells, place a small aliquot of this suspension in an incubation plate were a tissue is formed after cultivation. Here the hope is that if the suspension is diluted enough (i.e., has only a small number of cells per unit of volume), then probably the tissue formed will be mono-clonal. Note that after cultivation of several plates one typically get a number of empty plates (the transferred aliquot did not contain any cells) and some plates containing a tissue (the transferred aliquot contained at least one cell). The plates containing tissue and the plates without tissue will be termed "positive-tested" and "negative-tested" respectively.*

Normally, it is not possible to distinguish whether a single cell or more than one cell formed the tissue in a positive-tested plate; so we cannot claim with certainty whether a positive-tested plate produced mono-clonal a tissue or not. It is therefore of interest to determine the probability that a positive-tested plate received more than one cell.

In an experiment it was reported that 17 plates were tested positive out of 96 plates (from an array of 12x8 plates). It might be argued that (see exercise ?? in chapter 4) the number of cells transferred to the plate is Poisson distributed and that the intensity parameter is reasonably estimated to be $\hat{\lambda} = -\log(79/96) \approx 0.1949$*, which we will assume for the rest of the exercise.*

1- *What is the probability of having transferred no cells to the plate?*

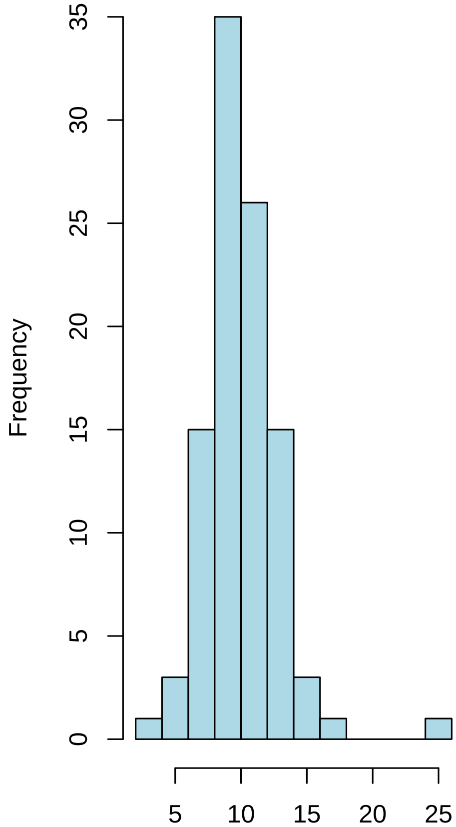
2- *What is the probability of having transferred exactly one cell to the plate?*

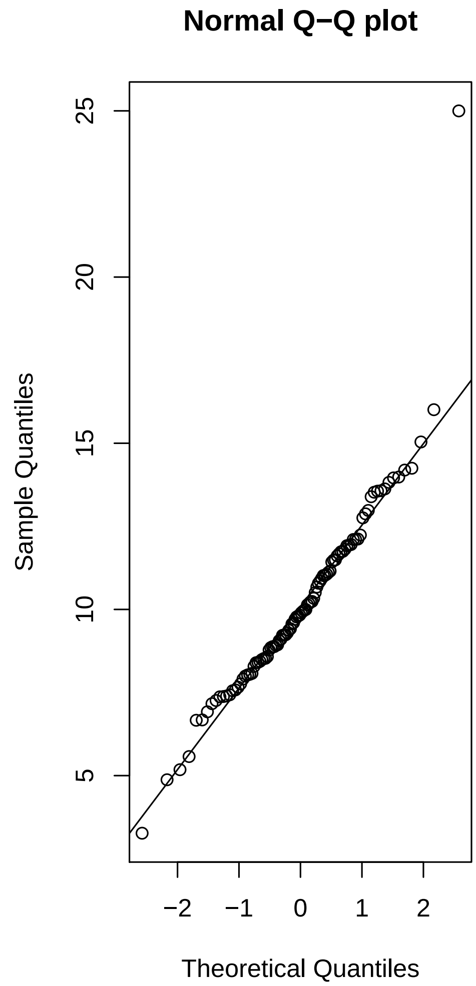
3- *What is the probability of having transferred more than one cell to the plate?*

4- *If you repeat the experiment 96 times (using an array of 12x8 plates), what would be the expected number of plates with mono-clonal tissues? What would be the expected number of plates with tissues formed by more than one cell (i.e. positive-tested but with not mono-clonal tissue)?*

Exercise 1.7 *This exercise regards some basic calculations using the normal distribution. The weights (dry matter) of plants of Amaranthus cruentus (a type of Amaranth used for alimentation) were determined for 100 plants placed far away for each other in a homogeneous cultivation field.*

1- *The figure displays a histogram and a normal Q-Q plot of the 100 observed weights. Is it reasonable to assume that the data is normally distributed?*

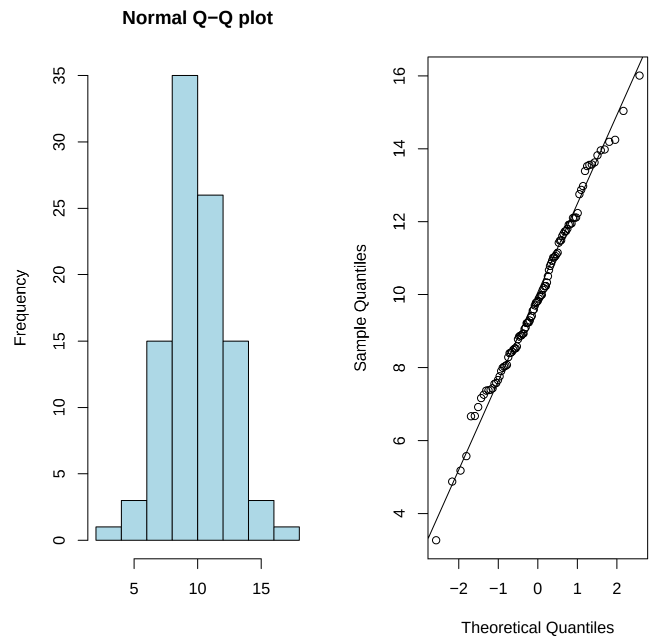




2- A descriptive summary of the 100 observations is given below

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  Variance
 3.268   8.431   9.864  10.170  11.740  25.000   7.832
```

After some analyses of the data it was identified that the first observation presents a value of 25 (the maximum). After removing the first *observation the histogram and the normal Q-Q-plot of the other 99 values is displayed below.*



A descriptive summary of the 99 observations (eliminating the first observation) is given below

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Variance
3.268	8.423	9.832	10.020	11.700	16.010	5.644

Why the variance became smaller? Is it reasonable to assume this data to be normally distributed?

3- *Assuming the data to be normally distributed with expectation 10.020 and variance 5.644 (i.e. equal to the sample mean and sample variance obtained after eliminating the first observation), what is the probability of observing a value larger or equal 25 (i.e. the value of the first* *observation)? (Hint: In R use the function pnorm to calculate this probability).*

4- *Assuming the data to be normally distributed with expectation 10.020 and variance 5.644, what would be the distribution of the measurements after subtracting 10.020 from them and dividing the result by the square root of 5.644?*

Exercise 1.8 *This exercises is related to the standard normal distribution and its uses.*

1- *If* $X \sim N(\mu, \sigma^2)$*, what is the expectation of* $X - \mu$*? What is the expectation of* $\frac{X-\mu}{\sqrt{\sigma^2}}$*? (Hint:* $E(Y - c) = E(Y) - c$ *and* $E(kY) = kE(Y)$*, for any r.v.* Y *and constants* c *and* k*).*

2- *If* $X \sim N(\mu, \sigma^2)$*, what is the variance of* $X - \mu$*? What is the variance of* $\frac{X-\mu}{\sqrt{\sigma^2}}$*? (Hint:* $Var(Y - c) = Var(Y)$ *and* $Var(kY) = k^2 Var(Y)$*, for any r.v.* Y *and constants* c *and* k*).*

3- *Suppose that* $Z \sim N(0, 1)$*. What is the probability of* Z *be larger than the following values:* $0.5, 0.55, 0.93, 1.96, 2$ *and* 3.49*? What is the probability of* Z *be smaller than the following values:* $0.5, 0.55, 0.93, 1.96, 2$ *and* 3.49*? What is the probability of* Z *be smaller than the following values:* 0.45*? (Hint: Use the function pnorm in R).*

4- *If* $X \sim N(30, 9)$ *what is the probability of* X *be larger than the following values:* $30, 33, 36, 40$*? What is the probability of* X *be smaller than the following values:* $30, 27, 24$ *and* 20*? (Hint: Use the function pnorm in R)*

Exercise 1.9 *This exercise is about the distribution of the sample mean.*

1- *If* $X_1, \ldots, X_n$ *are identically distributed with expectation* μ *(i.e.* $E(X_1) = \mu$*), what is the expectation of* $\bar{X} = 1/n \sum X_i$*?*
(Hint: recall that if c *is a constant and* X *and* Y *are a random variables, then* $E(cX) = cE(X)$ *and that* $E(X + Y) = E(X) + E(Y)$ *).*

2- *If* $X_1, \ldots, X_n$ *are* **independent** *identically distributed (iid), with variance* σ^2*, what is the variance of* $\bar{X} = 1/n \sum X_i$*?*
(Hint: recall that if c *is a constant and* X *is a random variable, then* $Var(cX) = c^2 Var(X)$*, moreover if* X *and* Y *are* **independent** *random variables, then* $Var(X + Y) = Var(X) + Var(Y)$*).*

3- If $X_1, \ldots, X_n$ *are* **independent** *identically distributed (iid), with* $X_1 \sim N(\mu, \sigma^2)$*, what is the distribution of* $\bar{X} = 1/n \sum X_i$*?*

4- If $X_1, \ldots, X_n$ *are* **independent** *identically distributed (iid), with* $X_1 \sim N(10, 9)$*, what is the probability of* X_1 *be larger than 13? If* n *is 100, what is the probability of* $\bar{X}$ *be larger than 13?*

1.4.1 Answer to selected exercises

Exercise 1.2

1- $P(A) = 2/6 = 1/3$, $P(A^c) = 1 - 1/3 = 2/3$.

2- $P(B) = 1/6$, $P(C) = 3/6 = 1/2$, $P(D) = 1/2$, $P(B \text{ or } C) = 2/3$.

3- Probability of observing 6 four times in four troughs is $(1/6)^4 = 1/1296 = 0.0007716049$.

Exercise 1.3

1- $P(X \text{ is odd}) = 1/2$, $P(X \text{ and } Y \text{ are odd}) = 1/4$.

2- $E(X) = 21/6 = 3.5$,
$Var(X) = 91/6 - (21/6)^2 = 11.666\ldots \approx 11.67$.

3- $E(X + Y) = 2 * 21/6 = 7$
$Var(X + Y) = 2\,[91/6 - (21/6)^2] \approx 26.83$.

3- $E(2X + Y) = 7 + 3.5 = 10.5$
$Var(2X + Y) = 4Var(X) + Var(Y) = 5Var(X) \approx 58.35$.

Exercise 1.4

1- $E(H_1 \text{ measured in milimeters}) = 892$,
$E(H_2 \text{ measured in milimeters}) = 382$,
$Var(H_1 \text{ measured in milimeters}) = 200$,
$Var(H_2 \text{ measured in milimeters}) = 100$.

2- $E(H_1 + H_2) = 89.2 + 38.2$

3- $Var(H_1 + H_2) = 2 + 1 + 2 \cdot 1.5 = 6$ when expressed in cm and $Var(H_1 + H_2) = 600$ when expressed in mm.

Exercise 1.5

2- $X \sim Po(3)$, then $P(X = 0) = 0.04978707...$, $P(X = 1) = 0.1991483$ and $P(X > 5) = 0.08391794... \approx 0.084$.

3- $X \sim Po(3) \Rightarrow Z = 1000X \sim Po(3000)$

4- $Z \sim Po(3000) \Rightarrow e^{-3000} \approx 0$

Exercise 1.6 Preliminaire remark: Let Y be a random variable representing the number of transferred cells to one plate. Note that, if $Y \sim Po(\lambda)$, then $P(Y = 0) = e^{-\lambda}$. Now, out of the 96 transfers of cells we know that $96 - 17 = 79$ did not contain any cells. Therefore, it is natural to estimate $P(Y = 0)$ by 79/96. Equating $P(Y = 0)$ to 79/96 we get the equation $e^{-\lambda} = 9/96$; solving for λ we obtain the estimate $\hat{\lambda} = -\log(79/96) \approx 0.1949$. [12]

1- $P(Y = 0) \approx e^{-0.1949} = 79/96 \approx 0.8229$.

2- $P(Y = 1) \approx e^{-0.1949} 0.1949 \approx 0.1604$

3- $P(Y > 1) = 1 - P(Y = 0) - P(Y = 1) \approx 1 - 0.8229 - 0.1604 = 0.0167$

4- The number of plates with mono-clonal tissue can be seen as a binomial random variable where the probability of success is $p = P(Y = 1) \approx e^{-0.1949} 0.1949 \approx 0.1604$ and the number of trials is $n = 96$. Such variable has expectation given by $n.p \approx 0.1604 \cdot 96 = 15.3984$. Analogously, the expected number of plates with tissue formed by more than one cell is $\approx 0.0167 \cdot 96 = 1.6032$.

Exercise 1.7

3-
$$
\begin{aligned}
X \sim N(10.02, 5.644) \Rightarrow P(X > 25) &= 1 - P(X \leq 25) \\
&\approx 1.436489 \cdot 10^{-10}.
\end{aligned}
$$

4- The distribution is the standard normal distribution, *i.e.*, $N(0, 1)$.

Exercise 1.8 Assume that $X \sim N(\mu, \sigma^2)$.

1- $E(X - \mu) = 0$ and $E(\frac{X-\mu}{\sqrt{\sigma^2}}) = 0$.

[12] This estimation technique is called moment estimation (of first order).

2- $Var(X - \mu) = Var(X) = \sigma^2$ and $Var\left(\frac{X-\mu}{\sqrt{\sigma^2}}\right) = 1$.

3- If $Z \sim N(0,1)$, then the probability of Z be larger than $0.5, 0.55, 0.93, 1.96, 2$ or 3.49 are 0.3085, 0.2912, 0.1762, 0.0250, 0.0228, 0.0002, respectively. Detailed calculation of one case: $P(Z > 0.5) = 1 - P(Z \leq 0.5) 1 - \Phi(0.5) \approx 0.3085$, where Φ is the cumulative diste?ribution function of a standard normal distribution.

4- If $X \sim N(30, 9)$, then $P(X > 30) = P(X - 30 > 0) = P((X - 30)/3 > 0) = P(Z > 0) = 1 - \Phi(0) = 1/2$. The othe probabilities are calculated in the same way.

Exercise 1.9

1- $E(\bar{X}) = \mu$

2- $Var(\bar{X}) = n\sigma^2/n^2 = \sigma^2/n$

3- $X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$

4- $P(X_1 > 13) \approx 0.1586553$, $P(\bar{X} > 13) \approx 0$.

1.4.2 Some R codes related to the exercises

```
######################################################################
# Exercise 1.6
######################################################################

#######################
# Item 2-
#######################
  # Calculating the probability of 0 and the probability of 1
  dpois(0,lambda=3); dpois(1,lambda=3)
  # Calculating the probability of 0 and the probability of 1 (alternative)
  dpois(c(0,1),lambda=3)
  # Calculating the probability of 0 or 1
  sum(dpois(c(0,1),lambda=3))
  # Calculating the probability of 0 or 1 or 2 or 3 or 4 or 5
  sum(dpois(as.numeric(0:5),lambda=3))
  # Calculating the probability of more than 5
  1-sum(dpois(as.numeric(0:5),lambda=3))
  # Rounding
  round(1-sum(dpois(as.numeric(0:5),lambda=3)), digits=3)
#######################
# Item 4-
#######################
dpois(0,lambda=3000)
exp(-3000)
```

```
#####################################################################
# Exercise 1.8
#####################################################################

#######################
# Item 3-
#######################
1-pnorm(25, mean=10.02, sd=sqrt(5.644))

#####################################################################
# Exercise 1.9
#####################################################################
Values <- c(0.5, 0.55, 0.93, 1.96, 2, 3.49)
round(1-pnorm(Values), digits=4)

#####################################################################
# Exercise 1.10
#####################################################################

#######################
# Item 4-
#######################
pnorm(13, mean=10, sd=3 , lower.tail=F)
round(pnorm(13, mean=10, sd=3/100 , lower.tail=F), digits=15)
```