Jan Vecer

Statistics and Econometrics, Lecture Notes

CRC PRESS
Boca Raton London New York Washington, D.C.

Chapter 1

Elements of Probability Theory

1.1 Probability Basics

Probability is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set of outcomes (also known as sample space), $\mathcal{F}$ is a set of events, and $\mathbb{P}$ is the probability measure. We denote by ω individual outcomes from the set Ω. The set of events $\mathcal{F}$ includes combinations of outcomes, and thus each event from $\mathcal{F}$ is a subset of Ω. Moreover, the set of events $\mathcal{F}$ includes the set of all outcomes Ω, and is closed under complements and countable unions. In mathematical notation

$$A \in \mathcal{F} \Rightarrow (\Omega \setminus A) \in \mathcal{F}, \qquad A_i \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

A probability measure assigns a value on the interval $[0, 1]$ to events in $\mathcal{F}$ with the following restrictions: the probability of the set of all outcomes Ω is one, and when the sets A_i are disjoint ($A_i \cap A_j = \emptyset,\ i \neq j$), probability of $\bigcup_i A_i$ is the sum of probabilities of individual events:

$$\mathbb{P}(\Omega) = 1, \qquad \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Example 1.1

Consider a coin toss. The two possible outcomes are "Head" and "Tail," so $\Omega = \{\text{Head},\ \text{Tail}\}$. The set of events is $\mathcal{F} = \{\emptyset, \{\text{Head}\}, \{\text{Tail}\}, \Omega\}$. The probability measure $\mathbb{P}$ is determined by the value of $\mathbb{P}(\{\text{Head}\}) = p \in [0, 1]$, because $\mathbb{P}(\{\text{Tail}\}) + \mathbb{P}(\{\text{Head}\}) = \mathbb{P}(\Omega) = 1$, and thus $\mathbb{P}(\{\text{Tail}\}) = 1 - p$. Note that we may have different probability measures determined by the value p, that comes with the same set of outcomes Ω and the same set of events $\mathcal{F}$. □

What is the purpose of assigning probability to events $\mathcal{F}$ rather than assigning the probability directly to individual outcomes ω? We have seen in Example 1.1 that the probability of the individual outcomes determines the probability of all events, but this does not work in general. One example of why it is not sufficient to assign probability just to individual outcomes is a situation when the set of outcomes is uncountable (such as the real line).

In this case the probability of each individual outcome may be zero, and it would be impossible to reconstruct the probability of events from outcomes with zero probability. For that one has to start with events that contain more than a critical fraction of outcomes, such as intervals in the case of continuous distributions on the real line.

Note that the set of outcomes may have nonnumerical values, such as in the case of the coin toss when the set of outcomes is $\Omega = \{\text{Head}, \text{ Tail}\}$. When the outcomes are assigned a number X, then we say X is a **random variable.** Formally, a random variable is a mapping

$$X : \Omega \to \mathbb{R}$$

with the property that for each Borel set B,

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\}$$

is an event from the set of events $\mathcal{F}$. This assures that the probability of the event $\{X \in B\}$ is well defined. Every Borel set can be obtained from closed intervals $[a, b]$ by taking complements, or countable unions or intersections of these sets.

We call

$$F(x) = \mathbb{P}(X \leq x),$$

a cumulative distribution function. The cumulative distribution function already determines $\mathbb{P}(X \in B)$ for all Borel sets, in which case we talk about the distribution of a random variable. The derivative f of the cumulative distribution function, if it exists, is known as a **density**

$$f(x)dx = dF(x).$$

The cumulative distribution function is always well defined, but not every distribution admits a density function. For instance when the cumulative distribution function has jumps, which is the case of discrete distributions, the density function does not exist in the classical sense as a function. Discrete distributions can be described by a finite or a countable list of probabilities $\mathbb{P}(X = x)$.

REMARK 1.1 The important difference between probability and statistics as scientific field is the following. Probability studies idealized models with parameters that are given. For instance, when considering a roll of a fair die, it is assumed that the probability of each outcome equals $\frac{1}{6}$. However, the real die may be biased, and some of these numbers can appear more or less frequently than in the idealized probability model. Statistics addresses the question whether given data support the original probabilistic model or not, which model to fit, and what are the best parameters that explain the observed data. For instance one can ask a question whether a given die is fair or not based on the numbers observed from consecutive die rolls. □

1.2 Conditional Probability and Independence

Let A and B be two events from $\mathcal{F}$. The conditional probability is defined by the formula

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Example 1.2 Drawing Aces:
Consider a standard deck of 52 cards, and let

$$A = \{\text{The first card is an ace}\},$$

and

$$B = \{\text{The second card is an ace}\}.$$

The probability that the first drawn card is an ace is

$$\mathbb{P}(A) = \frac{4}{52} = \frac{1}{13}.$$

The probability that the second drawn card is an ace is also

$$\mathbb{P}(B) = \frac{4}{52} = \frac{1}{13}.$$

Many people have a hard time to accept this fact as the second card comes after the first card, and this information has an impact on this probability in the following way. If the first card is an ace, there are 3 aces left in the remaining 51 cards, and thus the probability that the second card is an ace is $\frac{3}{51}$. But this is a conditional probability on the event A

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\frac{\binom{4}{2}}{\binom{52}{2}}}{\frac{4}{52}} = \frac{\frac{\frac{4\cdot 3}{2}}{\frac{52\cdot 51}{2}}}{\frac{4}{52}} = \frac{3}{51}.$$

If the first card is not an ace, there are 4 aces left in the remaining cared, and thus the probability that the second card is an ace is $\frac{4}{51}$. But this is a conditional probability on the event A^c

$$\mathbb{P}(B|A^c) = \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} = \frac{\frac{48}{52} \cdot \frac{4}{51}}{\frac{48}{52}} = \frac{4}{51}.$$

To compute the probability $\mathbb{P}(B)$, observe that

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c).$$

This is also known as the **law of total probability.** Applying this formula, we get

$$\mathbb{P}(B) = \frac{3}{51} \cdot \frac{4}{52} + \frac{4}{51} \cdot \frac{48}{52} = \frac{3 \cdot 4 + 4 \cdot 48}{52 \cdot 51} = \frac{4}{52}$$

as previously stated. To get a further inside of this result, consider that a given card can appear anywhere in the shuffled deck with the same probability, so the chance that the ace appears as the first or as the second card must be the same. ▯

When

$$\mathbb{P}(B|A) = \mathbb{P}(B),$$

or in other words when the conditional probability of the event B after observing the event A remains the same as the original unconditional probability of the event B, we call A and B to be **independent events.** We can equivalently define independence as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

which follows immediately from the definition of the conditional probability. Two events A, B that are not independent are called **dependent events.**

Example 1.3 Coin Tossing:
Consider tossing a coin two times. Let

$$A = \{\text{The first coin toss is a head}\},$$

and

$$B = \{\text{The second coin toss is a head}\}.$$

Assuming that $\mathbb{P}(A) = p$ (a biased coin is possible), we also have $\mathbb{P}(B) = p$ and $\mathbb{P}(A \cap B) = p^2$. Thus the two events A and B are independent. ▯

Example 1.4 Drawing Aces:
Recall the problem of drawing two aces from a standard deck of 52 cards with events

$$A = \{\text{The first card is an ace}\},$$

and

$$B = \{\text{The second card is an ace}\}.$$

We have seen that $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{13}$, and $\mathbb{P}(A \cap B) = \frac{4 \cdot 3}{52 \cdot 51} \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$, and thus the two events A and B are dependent. ▯

1.3 Short Introduction to Python

Our following text uses the programming language Python to illustrate computational concepts in probability and statistics. This text does not have an ambition to teach the readers Python in detail, but it can serve as a quick starting point. As a first step, one needs to install the language on the computer. We suggest to use the distribution called "Anaconda" from

`www.anaconda.com.`

This distribution includes not only Python, but a number of important libraries (including all that are used in this text) and some of the editors. I personally use Spider as an editor of choice.

1.3.1 Lists, Sets, Arrays

Let us introduce several concepts used in the text. The first concept is a list. List can be defined by explicit members, such as in

```
mylist = [1,2,3,4,5,6]
```

The same result is obtained if we define it with a range of values, such as in

```
mylist = list(range(1,7))
```

Note that the Python list is indexed from 0 to $n-1$ rather than from 1 to n, so in particular,

```
mylist[0]
```

gives 1,

```
mylist[5]
```

gives 6. This is a reason why we have to call the range of values with a number 7 in order to have the maximal element to be equal 6 in the above situation. We can also extract range of values by calling

```
mylist[1:3]
```

that gives [2,3]. Note that the value 4 that corresponds to mylist[3] is missing. We can also simply extract the elements from the end of the list by using negative indices. For instance,

```
mylist[-1]
```

gives the last value of the list (which is equal to 6). The last 3 elements are extracted as

```
mylist[-3:]
```

which gives [4,5,6]. Note that the same list can be created by a sloppy for cycle construction

```
mylist = list()

for i in range(1,7):
    mylist.append(i)
```

which starts with an empty list and inserts the elements 1 up to 6 by adding them one by one.

One can call a function on a given list. It is a rather powerful concept as one can use a single command that applies in principle on entire list, thus bypassing a need for a sloppy use of iterative cycles (like for cycle or while cycle). The objects in Python (and in other programming languages) that can be operated on are called *iterables.* Basic list operations include (on a list called mylist):

```
len(mylist)
```

This is the length of the list.

```
x in mylist
```

Tests if x is a part of the list.

```
x not in mylist
```

Tests if x is not a part of the list.

```
mylist.append(x)
```

Appends x to the list. Equivalent to mylist[len(mylist):] = [x].

```
mylist.count(x)
```

Counts the number of instances of x in the list.

```
mylist.pop(i)
```

Returns mylist[i] and removes it from mylist.

```
mylist.insert(i, x)
```

Inserts x to position i.

```
mylist.remove(x)
```

Removes the first instance of x from mylist.

An example of a more specific call on an iterable, specifically on mylist defined above is

```
squares = [x**2 for x in mylist]
```

that gives [1,4,9,16,25,36]. One can do many other such operations.

The second closely related concept to a list is a set. Set is formally a collection of unique elements. For instance

```
myset = set([1,1,2,2,3])
```

gives myset = {1,2,3}. One can apply the usual set tests (length, membership) or operations (inclusions, unions, intersections, differences) using the following commands (and using a second set t):

```
len(myset)
x in myset
x not in myset
myset.issubset(t)
myset.issuperset(t)
myset.union(t)
myset.intersection(t)
myset.difference(t)
myset.symmetric_difference(t)
```

The above described Python list has one particular shortage, namely it can include elements with very different types, such as in

```
mylist = [1,'cat','dog']
```

This is a completely legitimate definition of the list, which mixes both numbers and strings. However, such an object generally occupies a relatively large memory space and it takes more time to access the elements in a large list in comparison to arrays that are defined with a specific element type. The manipulation of arrays is a part of a library called numpy, which has to be called in the code preamble as

```
import numpy as np
```

The part "as np" is optional and it shortens calling the commands from numpy. Using this library, we can create an array [1,2,3,4,5,6] by calling

```
x = np.array([1,2,3,4,5,6])
```

or simply by defining the range of values by

```
x = np.arange(1:7)
```

1.3.2 Loops

Computers are especially powerful in well defined repetitive tasks. These repetitive tasks are realized by loops. Python has several loop methods, one of them uses "for" cycle and one uses "while" cycle. As an example, consider that we would like to add up squares of the integers

$$\sum_{i=1}^{n} i^2$$

for $n = 10$. A simple for loop in Python that realizes this sum is

```
k=0
for i in range(1,11):
 k=k+i**2
print(k)
```

The code initiates the running sum of squares k to be zero and then it adds i^2 to it with each run of the "for" loop. The commands that are a part of the loop are slightly shifted to the right. The last command, print(k), is not a part of the loop. It prints the resulting number, which is 385.

A good programming practice is to avoid loops if not absolutely necessary by applying functions on iterables (such as list or arrays). In fact, this question can be answered without a "for" loop by

```
x = np.arange(1,11)
sum(x**2)
```

This creates an array $[1, 2, \ldots, 10]$, apply a square to the elements and sum them up. However, "for" loops are very useful in situation when such an alternative solution is not readily available.

As an example for using the "while" loop, consider that we want to add up the squares of the integers until the sum hits some predefined value, such as the previously computed 385. The code that does the job looks like

```
k=0
i=0
upper = 385
while k < upper:
   i=i+1
   k=k+i**2
print(i,k)
```

It initiates the running sum k, the integer i as zeros and then enters the loop. The body of the loop is executed if the condition in the "while" loop is true. It first increments the integer i and adds i^2 to the running sum k. Once the condition is false, the loop stops and the current values are printed. The results is (10, 385) as expected.

1.3.3 Functions, Plotting

We often need to work with functions which map some values to some other values. As an example, we can explicitely define a function that corresponds to the sum of the squares

$$\sum_{i=1}^{n} i^2 = \frac{n(1+n)(1+2n)}{6}.$$

This is a known formula that represents a "square pyramidal number" as the consecutive squares form a pyramid if stacked on each other, with the smallest square being on the top. The Python implementation is called by "def" to define a function. The body of the definition can be used for some intermediate calculations, only the part called by "return" is given back:

```
def sumofsquares(n):
    return n*(1+n)*(1+2*n)/6
```

For instance, sumofsquares(10) returns 385. If we want to plot this function, we can use the matplotlib library by importing it:

```
import matplotlib.pyplot as plt
```

The plot needs both x and y variables, so let us first initiate the x values to represent an array $[0, 1, \ldots, 10]$ and subsequently call the plotting function

```
x = np.arange(11)
plt.plot(x,sumofsquares(x), 'bo', ms=5)
```

We can simply call plot by plt.plot(x,y) for given arrays x and y with the same length, but the resulting plot is by default continuous. Thus we use optional plotting parameters, such as 'bo', where 'b' stands for color (blue) and 'o' stands for the plotting point type circle. ms is the size of the plotting point. The resulting graph is given in Figure 1.1.

1.3.4 Probability in Python

Most functions related to probability are implemented in the Python library called scipy.stats. It is imported by invoking the following command:

```
import scipy.stats as sp
```

It includes a large number of specific distributions, in particular binomial, Poisson, geometric, uniform, normal, exponential, t, chisquare and others used in the following text. Let us illustrate the library function on binomial distribution that is explained in detail in the following section. The specification of the particular distribution (binomial in this case) is done by calling

```
sp.binom
```

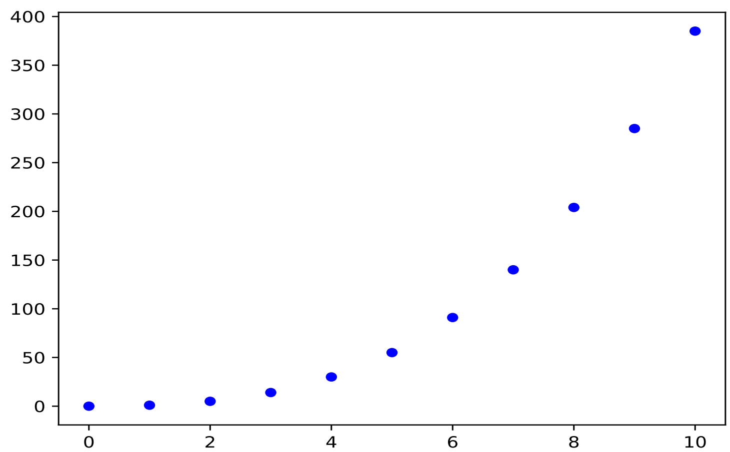

FIGURE 1.1: Square pyramidal number as a function of n.

followed by the specific function. For other distributions, one has to change the 'binom' part and replace it with the name of the distribution of interest. For a complete list of the available distributions implemented in the scipy library, refer to

```
https://docs.scipy.org/doc/scipy/reference/stats.html
```

Binomial distribution is a random variable that gives number of successes in n independent trials, where probability of success in a given trial is equal p. Thus the distribution depends on two parameters. Scipy.stats library gives the user the following functions:

```
sp.binom.rvs(n, p, loc=0, size=1, random_state=None)
```

This generates one random variable with the specific distribution. The first two parameters, n and p have to be specified, while the remaining three are set by default to the values in the bracket. For instance, consider that we toss 1000 times a fair coin and count the number of heads. This corresponds to a binomial random variable with $n = 1000$ and $p = \frac{1}{2}$. Thus

```
sp.binom.rvs(1000, 0.5)
```

generates one random variable representing the number of heads in 1000 such coin tosses. The number is random, so calling this function repeatedly results in different values. These values tend to be close to 500 as we would expect that approximately half of the values will result in a head, but there tends to be some little discrepancy from the exact value. We can generate more than one value at a given time by specifying a larger size (number of the generated random variables). The variable random_state is useful in testing the code as fixing this value will results in the same generated random variables. The random number generator in Python is in fact deterministic (or pseudorandom) and fixing the random_state will give the same results in all implementations. This is good for reproducibility of the results. In our text, we simply put 12345 as the random_state of our choice. For instance,

```
sp.binom.rvs(1000, 0.5, size = 10, random_state = 12345)
```

outputs

```
array([529, 502, 507, 497, 493, 507, 511, 497, 492, 475])
```

and the reader should get exactly the same numbers.

The probabilities of having exactly k successes in n such trials with probability p is given by a probability mass function (called in Python pmf):

```
sp.binom.pmf(k, n, p, loc=0)
```

In particular,

```
sp.binom.pmf(500, 1000, 0.5)
```

gives 0.025225018178354687, and thus the probability of having exactly 500 heads in 1000 tosses is in the order of 2.5%.

Cumulative distribution function cdf gives probability that the random variable is smaller or equal to a given value k:

```
sp.binom.cdf(k, n, p, loc=0)
```

For instance,

```
sp.binom.cdf(525,1000, 0.5)
```

gives 0.94662522898285784, which represents the probability that the number of heads in 1000 coin tosses will be smaller or equal to 525. In other words, it is rather likely to see the actual number of heads to be below 525. A closely related concept to cdf is a survival function defined as sf = 1 - cdf, the complementary probability:

```
sp.binom.sf(k, n, p, loc=0)
```

Calling

```
sp.binom.sf(525,1000, 0.5)
```

gives 0.053374771017142109, so the probability that one gets more than 525 heads in 1000 coin tosses is just slightly above 5%.

Percent point function ppf represents a quantile of the distribution. It is a smallest value of q_α, where $\mathbb{P}(X \leq q_\alpha) = F(q_\alpha) \geq \alpha$. It is called by

```
sp.binom.ppf(q, n, p, loc=0)
```

For instance, a choice of $\alpha = 0.95$ in the above example is called

```
sp.binom.ppf(0.95, 1000, 0.5)
```

gives 526, so 95% of all possible values of the random variable representing the number of heads in 1000 tosses are below or equal to 526. This is expected as the probability of being smaller or equal to 525 was nearly 95%, but fell slightly short of that value. A special case of a quantile that corresponds to $\alpha = 0.5$ is called median and it has its own function

```
sp.binom.median(n, p, loc=0)
```

In particular,

```
sp.binom.median(1000, 0.5)
```

gives 500. This is expected from the symmetry of the distribution.

Basic statistics descibed later in the text, such as the mean (expected value), variance and standard deviation are obtained by calling

```
sp.binom.mean(n, p, loc=0)
sp.binom.var(n, p, loc=0)
sp.binom.std(n, p, loc=0)
```

In our case,

```
sp.binom.mean(1000, 0.5)
sp.binom.var(1000, 0.5)
sp.binom.std(1000, 0.5)
```

give values 500, 250 and 15.811388300841896 respectively. These statistics can be extracted from a similar command

```
sp.binom.stats(n, p, loc=0, moments='mv')
```

where 'm' stands for mean and 'v' stands for variance. In addition, one can also extract two additional statistics, namely 's' (skewness) and 'k' (kurtosis). The function call

```
sp.binom.stats(1000, 0.5, moments='mvsk')
```

outputs

```
array(500.0), array(250.0), array(0.0), array(-0.002))
```

that agrees with our previous findings.

The last function we mention here is a confidence interval of size α, called

```
sp.binom.interval(alpha, n, p, loc=0)
```

For instance, 95% confidence interval in our coin tossing case is called by

```
sp.binom.interval(0.95, 1000, 0.5)
```

and it outputs (469.0, 531.0). Thus 95% of the values of the random variable fall in the range between 469 and 531.

1.3.5 If Else Statements

Our text also works with conditional statements. The code is executed condition on that the statement is true or false. The idea is that 'if' condition is true, the code should do something, otherwise do something else after the 'else' statement. The 'else' part is optional, in which case the code executes only the 'if' part if the test condition is true, otherwise it does nothing.

Let us illustrate the conditional statement on a situation when we want to generate a random variable with the following values:

$$X = \begin{cases} +1, & p = \frac{18}{37}, \\ -1, & 1-p = \frac{19}{37}. \end{cases}$$

This represents a profit/loss distribution of betting one dollar on red in roulette. Standard roulette has 37 numbers, from which 18 are red. Thus one wins a dollar with probability $\frac{18}{37}$ and loses one dollar with probability $\frac{19}{37}$. Such random variable can be obtained from a binomial distribution with $n = 1$, where the random variable Y takes only two values, one and zero measuring success in one trial. We have

$$Y = \begin{cases} 1, & p = \frac{18}{37}, \\ 0, & 1-p = \frac{19}{37}. \end{cases}$$

This random variable Y is also called a Bernoulli random variable. Note that X can be written as $X = 2Y - 1$, which maps $(0, 1)$ to $(+1, -1)$. If we want to generate 10 random values representing the variable X, we could simply call

```
2*sp.bernoulli.rvs(18/37, size = 10, random_state = 12345)-1
```

which gives array([1, -1, -1, -1, 1, 1, 1, 1, 1, 1]).

A second possible approach to generate random variables X is the following. First, generate a uniform random number from an interval [0,1], meaning that each such value is equally likely. If the resulting value is below $\frac{18}{38}$, assign it $+1$, otherwise assign it -1. This procedure will create a random variable with the desired value and the desired probability. A uniform random variable is generated by calling

```
sp.uniform.rvs(random_state = 12345)
```

which gives 0.92961609281714785. Fixing the random_state means that this result is reproducible. Omission of this parameter would lead to an unpredictable value between [0,1].

Using the above logic, one can define a function 'pl' that would output plus one if the uniform random variable falls below $\frac{18}{37}$ and output minus one otherwise. The exact if else statement inside the function has the following form:

```
def pl(x):
    if x<= 18/37:
        a = 1
    else:
        a= -1
    return a
```

In particular

```
pl(sp.uniform.rvs(random_state = 12345))
```

outputs -1. One little shortage of this approach is that the function operates only on one number rather than a full array. In order to apply it to arrays, one can call

```
pl = np.vectorize(pl)
```

after which the function will apply arrays as well. For instance, generating 10 uniform random variables by

```
sp.uniform.rvs(size =10,random_state = 12345)
```

gives

```
array([ 0.92961609,  0.31637555,  0.18391881,  0.20456028,  0.56772503,
        0.5955447 ,  0.96451452,  0.6531771 ,  0.74890664,  0.65356987])
```

and applying the 'pl' function on this array

```
pl(sp.uniform.rvs(size =10,random_state = 12345))
```

results in array([-1, 1, 1, 1, -1, -1, -1, -1, -1, -1]).

1.4 Binomial Distribution

This section introduces some basic discrete distributions. The discrete distributions are given by a list of probabilities $\mathbb{P}(X = x)$ for a finite or countable list of values of x. The most basic random variable X marks the success of a given outcome in a particular trial. This random variable takes only values 1 or 0 depending on the success of the outcome:

$$X = \begin{cases} 1 & \text{if successful (with probability } p\text{)}, \\ 0 & \text{if not successful (with probability } 1-p\text{)}. \end{cases}$$

This random variable is also known as a **Bernoulli random variable.** For instance, a random variable that indicates the success in a fair coin toss (say tossing a head), takes the value 1 with probability $p = \frac{1}{2}$. When rolling a die and marking number 6 as a success leads to a Bernoulli random variable with $p = \frac{1}{6}$.

Let us consider a situation when the trial is repeated, for instance when the coin is tossed a number of times. Define $Z = \sum_{i=1}^{n} X(i)$, the total number of successes in n consecutive trials. Let us determine $\mathbb{P}(Z = k)$. In order to illustrate the concept, take $n = 5$ and $k = 2$, so the question is what is the probability of exactly 2 successes in 5 trials. Let us denote by Y success and by N a fail in a given trial. An example how to get 2 successes in 5 trials is a sequence

$$YYNNN.$$

The probability of getting this sequence in this particular order is given by

$$p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p) = p^2 \cdot (1-p)^3.$$

However, this is not the only possible way how to get 2 successes in 5 trials, other orders, such as

$$NYNNY$$

are possible. Note that the probability of the above sequence is also $p^2 \cdot (1-p)^3$, so we just need to determine in how many different orders of 2 successes in 5 trials one can get. If all 5 outcomes were distinguishable (like numbers 1, 2, 3, 4, 5), one gets $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ different ways how to order these elements. Let us start with a hypothetical situation that we can distinguish between different successes and fails, and index them with a subscript. Then one possible ordering out of the total of $5! = 120$ is $Y_1Y_2N_1N_2N_3$, which is different from $Y_2Y_1N_2N_3N_1$. However, we do not care in which order different successes Y and different fails N come. This is for instance similar to matching numbers in a lottery. When choosing the numbers, it matters if there is a match with the drawn numbers regardless of the order in which they were drawn.

Returning to our example, there are two successes Y_1 and Y_2, and there are $2! = 2 \cdot 1 = 2$ number of orders of these two elements. Thus half of the total $5! = 120$ orders has Y_1 preceding Y_2 and the other half Y_2 preceding Y_1, but since we do not care about their ordering, we have to scale down the total number of 5! by 2! if we do not care about ordering of the successful outcomes. Similarly, we get $3! = 3 \cdot 2 \cdot 1$ number of ways how to order fails N, and when we do not care about ordering of these outcomes as well, we have to scale the number of orderings by additional 3!. In conclusion, there are

$$\frac{5!}{2! \cdot 3!} = 10$$

different ways how to get 2 successes in 5 trials. Since this is a reasonably small number, we can list all the possibilities here:

$$YYNNN, \quad YNYNN, \quad YNNYN, \quad YNNNY, \quad NYYNN,$$

$$NYNYN, \quad NYNNY, \quad NNYYN, \quad NNYNY, \quad NNNYY.$$

For general values n and k, the number of k successes in n trials is given by

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}.$$

This number is known as a binomial coefficient. Thus the probability of getting k successes in n trials is given by a formula

$$\mathbb{P}(Z = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The random variable Z has a **binomial distribution** and we will write $Bin(n, p)$ to represent it.

Number of heads in n coin tosses, number of a particular outcome in a die roll or a number of wins in betting (like in a roulette) follow the binomial distribution. Figure 1.2 illustrates the distribution of the number of heads in 1000 tosses of a fair coin. One should expect to see around 500 heads in 1000 coin tosses. In fact, with 95% probability, the number of heads will be in the interval $[469, 531]$. This represents a typical range of the outcomes and we will revisit this concept in a later text. Figure 1.3 shows the distribution of the number of 6s in 600 die rolls. One should expect to see around 100 6s in 600 die rolls. The typical range of the outcomes is in the interval $[82, 118]$, which covers 95% of all the possible outcomes. Figure 1.4 shows the distribution of

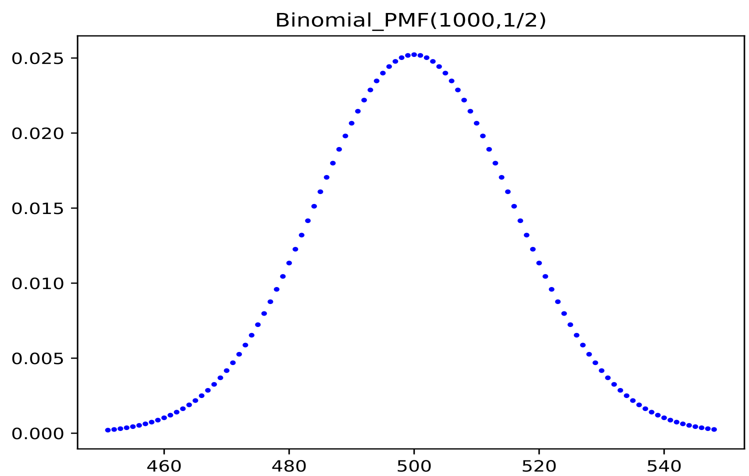

FIGURE 1.2: Binomial distribution $Bin(1000, \frac{1}{2})$ that represents the probability of a given number of heads in 1000 coin tosses of a fair coin.

the number of wins in 100 bets on a single number in a roulette. The probability of success in each individual bet is $\frac{1}{37}$. One should see about $\frac{100}{37} \approx 2.7$ wins in 100 bets, the typical range of number of wins is $[0, 6]$ which covers more than 95% of all probability.

Example 1.5 Python Code:
This code is used in generating the figures in this section, but called with different parameters n and p. Note that the range of values for plotting is computed from the 0.001 and 0.999 quantiles, so the values plotted represent 99.8% of all values of interest.

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

n=1000; p=1/2

x = np.arange(sp.binom.ppf(0.001, n, p),
              sp.binom.ppf(0.999, n, p))
plt.figure(1)
plt.plot(x, sp.binom.pmf(x, n, p), 'bo', ms=2)
plt.title('Binomial_PMF(1000,1/2)')
```

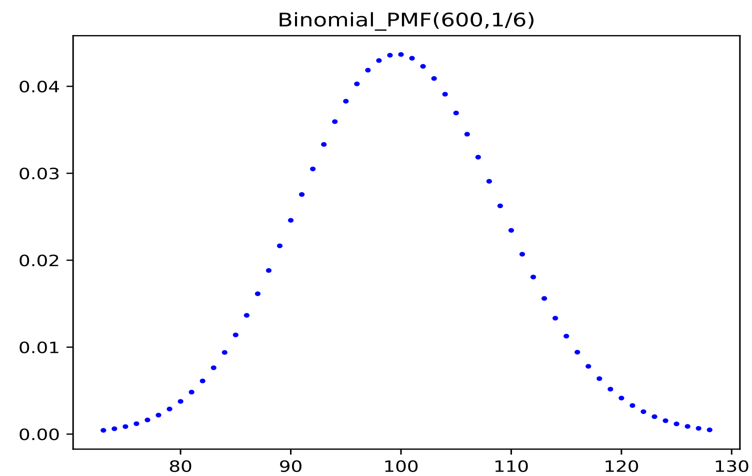

FIGURE 1.3: Binomial distribution $Bin(600, \frac{1}{6})$ that represents the probability of a given number of 6s in 600 die rolls of a fair die.

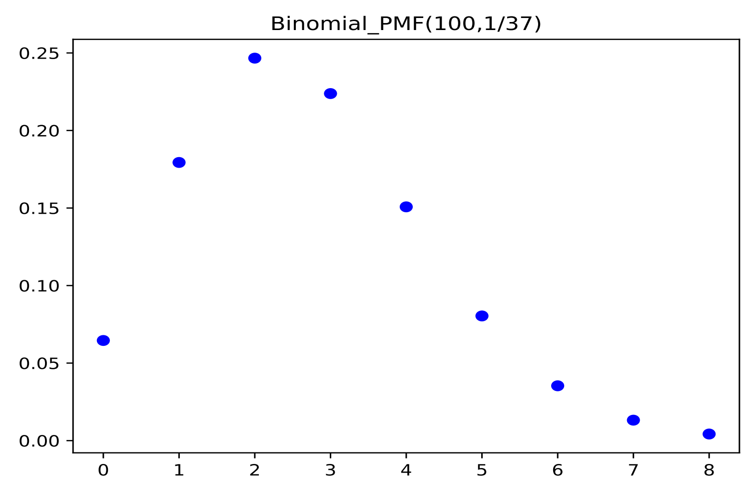

FIGURE 1.4: Binomial distribution $Bin(100, \frac{1}{37})$ indicates the number of wins in 100 bets on a single number in a roulette.

```
alpha = 0.95
print(sp.binom.interval(alpha, n, p))
```

□

1.5 Poisson Distribution

Consider the situation when the probability of the success is relatively small, but the number of trials is relatively large. This happens for instance in the following example. Winning a jackpot in a given lottery has usually an astronomically small probability, but there are millions of people who bet in a given draw. What is the probability that there are exactly k winners of the jackpot? The answer is (assuming that the tickets filed are random) that this probability follows binomial distribution

$$\mathbb{P}(Z = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where n is the number of tickets sold and p is the probability of winning the jackpot. For instance, a major lottery Euromillions has the probability of winning the jackpot equal to $p = \frac{1}{116{,}531{,}800}$. Let us consider a draw with $n = 60,000,000$ tickets sold. Computing the exact value of this probability for typical values of k can pose a significant computational challenge, even the best software has problems in evaluation of the above probability.

However, it is relatively simple to get a very fast and accurate approximation of the binomial probability when n is relatively large ($n \to \infty$), p is relatively small ($p \to 0$), and $np \to \lambda$. Note that

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!n^k} \cdot (np)^k \cdot \left(1 - \frac{np}{n}\right)^n \cdot (1-p)^{-k}$$
$$\to \frac{1}{k!}\lambda^k e^{-\lambda}.$$

We have used the facts that $\frac{n-l}{n} \to 1$, $np \to \lambda$, $\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$, and $(1 - p)^{-k} \to 1$. Hence,

$$\mathbb{P}(Z = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}.$$

This is know as a **Poisson approximation** and the distribution given by $\mathbb{P}(Z = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ is called a **Poisson distribution.** We will denote by

$Po(\lambda)$ the Poisson random variable with the parameter λ. The Poisson approximation says that

$$Bin(n,p) \approx Po(np).$$

Let us return to our example of winning the jackpot in Euromillions lottery and consider the probability that $k = 2$ tickets win the jackpot out of $n = 60,000,000$ random tickets. In this case,

$$\lambda = n \cdot p = \frac{60,000,000}{116,531,800} \approx 0.514881,$$

and

$$\mathbb{P}(Z = 2) \approx e^{-0.514881} \cdot \frac{0.514881^2}{2!} \approx 0.0792088.$$

Thus the probability that exactly two tickets win the jackpot is approximately 7.92% in this case.

The Poisson approximation works quite well even for relatively small values of n and relatively large values of p. Consider the previously mentioned example of betting 100 times on a single number in roulette. The number of wins follows binomial distribution $Bin(100, \frac{1}{37})$. According to the Poisson approximation, this should be close to the random variable $Po(\frac{100}{37})$. Table 1.1 compares the two probability distributions. As seen from the table, the probabilities differ at most by 0.003. Figure 1.5 shows the two distribution graphically.

TABLE 1.1: Comparison of the binomial distribution $Bin(100, \frac{1}{37})$ and its Poisson approximation $Po(\frac{100}{37})$.

k	$Bin(100, \frac{1}{37})$	$Po(\frac{100}{37})$
0	0.065	0.067
1	0.179	0.181
2	0.247	0.245
3	0.224	0.221
4	0.151	0.149
5	0.080	0.081
6	0.035	0.036

The corresponding code that generates Figure 1.5 follows. Note that the scipy library calls the Poisson distribution as 'poisson'.

```
import numpy as np
import scipy.stats as sp
```

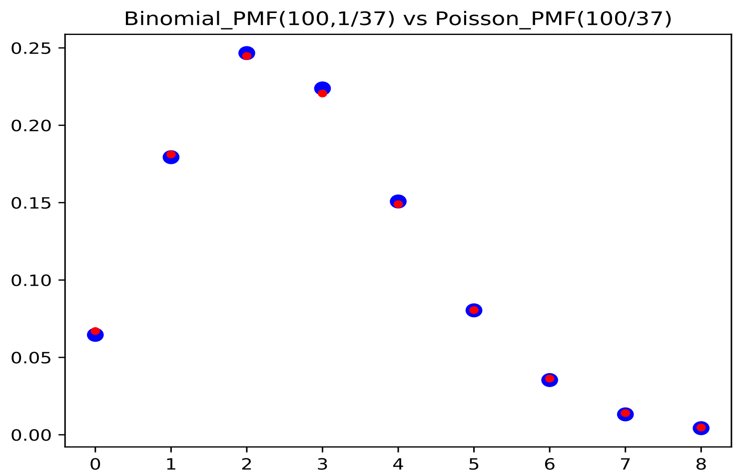

FIGURE 1.5: Comparison of the binomial distribution $Bin(100, \frac{1}{37})$ (blue points) vs Poisson distribution $Po(\frac{100}{37})$ (red points).

```
import matplotlib.pyplot as plt

n=100; p=1/37

x = np.arange(sp.binom.ppf(0.001, n, p),
              sp.binom.ppf(0.999, n, p))
plt.figure(1)
plt.plot(x, sp.binom.pmf(x, n, p), 'bo', ms=8)
plt.plot(x, sp.poisson.pmf(x, n*p), 'ro', ms=4)
plt.title('Binomial_PMF(100,1/37) vs Poisson_PMF(100/37)')
```

1.6 Geometric Distribution

Let us consider the situation when we are waiting for the first success in repeated independent trials, where the probability of success in a given trial is p. The success in i-th trial is given by a Bernoulli random variable $X(i)$, where $\mathbb{P}(X(i) = 1) = p$, and $\mathbb{P}(X(i) = 0) = 1 - p$. Examples of this situation include waiting for the first head in coin tossing, waiting for the first 6 in die tossing, waiting for the first win when betting on a single number in roulette, or waiting for the first jackpot win in a major lottery. Let T denote the first when we get the first success, or in other words,

$$T = \min\{n > 0 : X(n) = 1\}.$$

What is the distribution of the random variable T? In order to have $T = k$, we must have a fail in the first $k-1$ trials followed by the first success in the k-th trial. Probability of this event is given by

$$\mathbb{P}(T = k) = (1-p)^{k-1} \cdot p, \qquad k = 1, 2, \ldots$$

This distribution is called a **geometric distribution.**

Example 1.6 Memoryless property of geometric distribution:
Consider two events $A = \{T > k\}$ and $B = \{T > k + l\}$. The event A means that the success did not come in k trials, the event B means the success did not come in $k + l$ trials. It is not difficult to see that

$$\mathbb{P}(A) = \mathbb{P}(T > k) = \sum_{i=k+1}^{\infty} \mathbb{P}(T = i) = \sum_{i=k+1}^{\infty} (1-p)^{k-1} \cdot p = (1-p)^k.$$

This is intuitive as the probability of k consecutive fails is exactly $(1-p)^k$. Similarly, $\mathbb{P}(B) = (1-k)^{k+l}$. Now consider the case that one has observed the event A (k consecutive fails). How does it influence the probability of l additional fails (event B given A)? We get

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{(1-k)^{k+l}}{(1-p)^k} = (1-p)^l.$$

We have used the fact that in this particular case we have $B \subset A$, and thus $\mathbb{P}(A \cap B) = \mathbb{P}(B)$.
of numbers $1, 2, \ldots, 50$. We have seen that $Z_i \sim Bin(440, \frac{1}{10})$. Using

What is the consequence of the above result? If somebody waited k units of time for a success and it did not come, he has the same probability of success in additional l units of time as somebody who just started the game. For instance, if someone played a lottery for one year and did not win, this fact alone does not give the player any better odds of success as to someone who just bought the ticket for the first time. In other words, the system does not keep track of the history.

Many people have a hard time to take it as a fact, many lottery players believe that buying lottery tickets increases the chances of winning in a long run, but this is not the case. On a similar note, some people believe that if a given number in a lottery was not drawn for a while, it should be drawn next time around with a higher probability, which is again not the case. The reader should keep in mind that price returns in the markets do not have memory, although they may not keep the same distribution, it may be influenced by the changing volatility. However, parameters such as the volatility or the interest rate have memory, but one cannot make a profit out of this knowledge as the prices depend on these parameters in a nonlinear fashion. Sports games when played in a tournament do not tend to have a significant memory as well. Individual games may or may not have a memory. A sport that has a strong memory is for instance boxing, a heavy hit in the beginning of the match seems to be well remembered by the affected player. Sports like tennis, football or baseball have almost no memory.

For comparison, check the example of waiting for a subway in the following section. Waiting for a subway has a memory, and the probability of success is increasing with time. On the other hand, a system with a memory may have a decreasing probability of success when waiting for it. Some books on dating suggest that one should start a conversation within 5 seconds after making an eye contact, otherwise the probability of success is getting smaller after that critical time. ▯

Example 1.7 Rolling a die:

Consider waiting for the first '6' when rolling a die. The corresponding distribution is geometric with $p = \frac{1}{6}$ is shown in Figure 1.6. Here is the Python

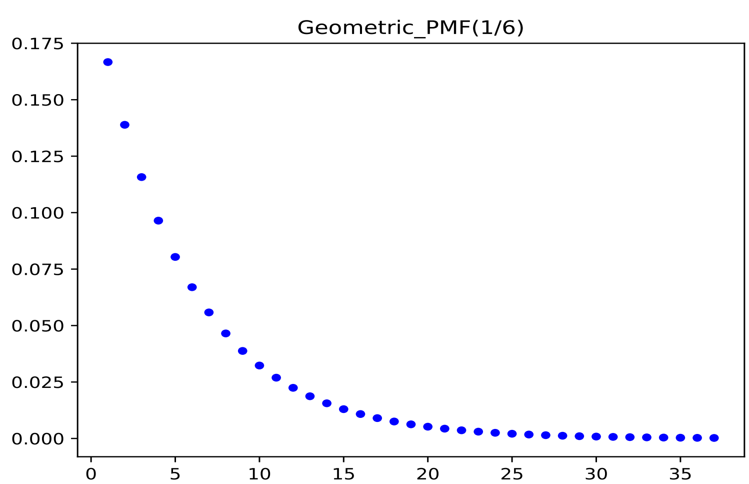

FIGURE 1.6: Geometric distribution with $p = \frac{1}{6}$.

code that generates the probability mass function. Note that the 95% of the probability is below 17, which is an output of the last command line.

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

p=1/6

x = np.arange(1, sp.geom.ppf(0.999, p))

plt.figure(1)
plt.plot(x, sp.geom.pmf(x, p), 'bo', ms=4)
plt.title('Geometric_PMF(1/6)')

print(sp.geom.ppf(0.95,p))
```

□

1.7 Uniform Distribution

Uniform distribution puts an equal probability on every possible outcome. Consider first the case when the number of outcomes is finite and equal to N, and the outcomes take values from 1 to N. Then probability of each outcome is equal to

$$\mathbb{P}(X = k) = \tfrac{1}{N}.$$

Uniform distribution appears in a die roll (6 equally likely outcomes), roulette (37 or 38 equally likely outcomes), or in lotteries where each number is equally likely to be drawn. We are considering idealized model, the question whether a given die, roulette or lottery is biased is a question for statistics discussed later in this text.

Example 1.8 Sampling from uniform distribution:
Let us address the question what is the distribution of a given number when drawn from the uniform distribution. Let Z be the number of times this particular number is drawn from M draws. The probability of success in an individual draw is $p = \frac{1}{N}$, and the number of total successes Z follows binomial distribution

$$\mathbb{P}(Z = k) = \binom{M}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{M-k}$$

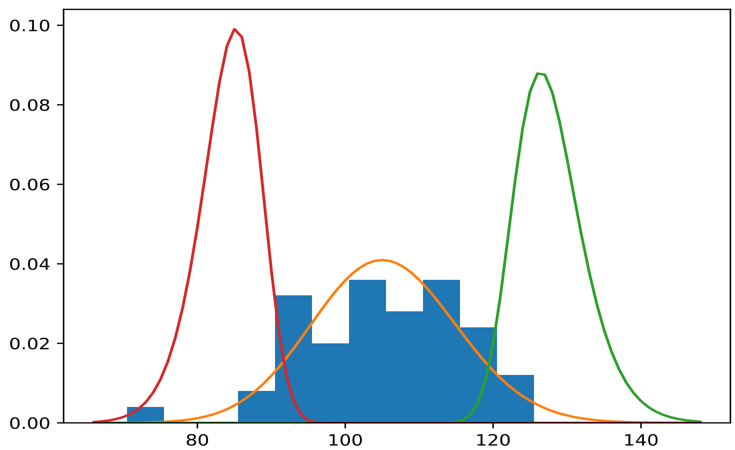

FIGURE 1.7: The expected and the realized frequency of drawn numbers in 440 draws of Euromillions lottery (histogram and the graph in the center) together with the distribution of the maximal frequency (right) and the minimal frequency (left).

For instance consider the Euromillions lottery where 5 numbers are being drawn from the total of 50. The probability that a given number is drawn in a given draw is thus $\frac{1}{10}$. As of October 27, 2017, there were $n = 1054$ draws in total. Thus the distribution (expected frequency) of a given number follows binomial distribution $Bin(1054, \frac{1}{10})$. The graph of this distribution together with the realized frequencies of drawn numbers (in the form of a histogram) is given in Figure 1.7. In addition, we plot the theoretical distribution of the maximum and the minimum value which is derived later. The realized frequency is close enough to the theoretical binomial frequency with one noticeable outlier, namely the number 46 was drawn only 75 times, which seems to be in the very low end of the binomial distribution. The probability that a random variable with $Bin(1054, \frac{1}{10})$ distribution falls at or below 75 can be computed from the binomial cumulative distribution function as

```
sp.binom.cdf(75,n,p)
```

which gives 0.00067954023224671596, which is approximately 1:1500. This may seem rather unlikely, even if the situation when we deliberately chose the number with the smallest frequency. However, how exactly this value fits the distribution of the minimum frequency is answered in the text that follows. $\square$

Example 1.9 Distribution of the maximum and the minimum:
A good fit to a given distribution should be also realized for the observed maximum and the observed minimum. The previous example of the frequencies of the numbers drawn in the Euromillions lottery motivates to check what exact values should we see for the maximum and minimum statistics. In particular, the frequency of 75 drawn values of number 46 in 1054 seems rather small.

Let us start with the theoretical distribution of the maximum from n independent identically distributed observations $Z_1, Z_2, \ldots, Z_n$. Determination of the distribution of the maximum is relatively simple. Let us determine the cumulative distribution function of the maximum $F_{max}(x)$ from n random observations. We have

$$\begin{aligned} F_{max}(x) = \mathbb{P}(Z_{max} \leq x) &= \mathbb{P}(Z_1 \leq x, Z_2 \leq x, \ldots, Z_n \leq x) \\ &= \mathbb{P}(Z_1 \leq x) \cdot \mathbb{P}(Z_2 \leq x) \cdot \cdots \cdot \mathbb{P}(Z_n \leq x) = [F(x)]^n, \end{aligned}$$

where $F(x)$ is the cumulative distribution function of the random variable Z_i. The density function (if it exists) is given by the derivative with respect to x:

$$f_{max}(x) = n \cdot f(x) \cdot [F(x)]^{n-1}.$$

In our case we are dealing with a discrete distribution, so the density function does not exist, but we can get the probabilities from

$$\begin{aligned} \mathbb{P}(Z_{max} = x) &= \mathbb{P}(Z_{max} \leq x) - \mathbb{P}(Z_{max} \leq x - 1) \\ &= F_{max}(x) - F_{max}(x-1) \\ &= [F(x)]^n - [F(x-1)]^n \end{aligned}$$

Let us apply this analysis on numbers drawn in Euromillions lottery. Strictly speaking, while each frequency of a number i, Z_i ($n = 50$ different values), follows $Bin(1054, \frac{1}{10})$, the results on the distribution of the maximum from the above discussion do not apply exactly as the Z_i's are dependent (they satisfy linear dependence $\sum_{i=1}^{50} Z_i = 5 \cdot 1054$) and the previous discussion assumed independence. However, this dependence is rather negligent and thus the result that we obtained earlier is close enough to the true distribution.

Before we take specific numbers, let us give some prior estimate where to expect to see the maximal value. Given that we have 50 observations, the observation with the largest value is expected to be somewhere in the range of 2% of the possible values. To be more exact, top 2% values start at the value 126 as we can see from the following command

```
sp.binom.cdf(0.98, 1054, 0.1)
```

This means that 126 is the breaking point between the bottom 98% of the values and top 2% of the values. Indeed, we compute that

$$\mathbb{P}(Z_{max} = 126) = 0.087879100491850104,$$

that we get from

```
sp.binom.cdf(126,1054,0.1)**50 - sp.binom.cdf(125,1054,0.1)**50
```

while

$$\mathbb{P}(Z_{max} = 125) = 0.083380425741693354,$$
$$\mathbb{P}(Z_{max} = 127) = 0.08756022182893175,$$

computed from

```
sp.binom.cdf(125,1054,0.1)**50 - sp.binom.cdf(124,1054,0.1)**50
sp.binom.cdf(127,1054,0.1)**50 - sp.binom.cdf(126,1054,0.1)**50
```

so indeed the number 126 is the most likely value of the maximum from 50 observations coming from $Bin(1054, \frac{1}{10})$, but the maximum itself has some dispersion, so the values slightly away from 126 are also quite possible. The most frequently chosen number in Euromillions, number 50, appeared in 124 draws. This is indeed a value within the previously discussed range, even on a slightly smaller side. In other words, this maximum is not so large. We have

$$\mathbb{P}(Z_i \geq 124) = 1 - \mathbb{P}(Z_i \leq 123) = 0.033807439830801221,$$

that can be computed from

```
1 - sp.binom.cdf(123,1054,0.1)
```

meaning that the probability of observing a value greater than 124 is just around 3.4%, but this falls somewhat short from being in the top 2%. In terms of the distribution of the maximum,

$$\mathbb{P}(Z_{max} \geq 124) = [F(x)]^n = 0.2532957851333949,$$

computed from

```
sp.binom.cdf(124,1054,0.1)**50
```

confirming that the maximum is relatively small. Only about 25% of the maxima would be smaller or equal to 124, other scenarios would lead to higher values.

Similarly for the distribution of the minimum Z_{min}, the least chosen frequency,

$$Z_{min} = \min\{Z_1, Z_2, \ldots, Z_{50}\},$$

we have

$$\begin{aligned}
F_{min}(x) &= \mathbb{P}(Z_{min} \leq x) = 1 - \mathbb{P}(Z_{min} > x) \\
&= 1 - \mathbb{P}(Z_1 > x, Z_2 > x, \ldots, Z_n > x) \\
&= 1 - \mathbb{P}(Z_1 > x) \cdot \mathbb{P}(Z_2 > x) \cdot \cdots \cdot \mathbb{P}(Z_n > x) \\
&= 1 - [1 - F(x)]^n,
\end{aligned}$$

where $F(x)$ is the cumulative distribution function of the random variable Z_i.

In our specific situation, the smallest observed frequency corresponds to number 46 that was chosen only 75 times. We already computed that probability of being at that level all smaller is about 1:1500. This seems even smaller on the scale of the minimum that we would expect to see in the bottom 2% values (probability 1:50). To see if this minimum is indeed unusual, let us compute Thus

$$\mathbb{P}(Z_{min} \leq 75) = 1 - [1 - F(75)]^{50} = 0.033417438898618812.$$

that we get from

```
1-(1-sp.binom.cdf(75,1054,0.1))**50
```

The standard approach in statistics is to reject events with less than 5% probability as typical, so according to this cutoff, we should not see such a small minimal frequency if the numbers were chosen from the uniform distribution. On the other hand, the chances of this minimal frequency are not far from 5% (about 1 in 30), so this is still within the range of possible, although not very typical values of the minimal frequency. One would need to have much higher odds (in the scale of 1 in 1000 or more) in order to have a more solid evidence that something is wrong with the randomness of the lottery draw.

Here is the corresponding Python code that generates Figure 1.7:

```
n=1054; p=1/10

freqs = [[1, 107],[2, 91],[3, 105],[4, 119],[5, 109],
         [6, 105],[7, 110],[8, 95],[9, 99],[10, 118],
         [11, 112],[12, 104],[13, 105],[14, 112],[15, 108],
         [16, 98],[17, 114],[18, 93],[19, 121],[20, 104],
         [21, 113],[22, 94],[23, 119],[24, 106],[25, 114],
         [26, 116],[27, 113],[28, 104],[29, 114],[30, 114],
         [31, 99],[32, 91],[33, 86],[34, 94],[35, 103],
         [36, 101],[37, 118],[38, 115],[39, 100],[40, 100],
         [41, 86],[42, 109],[43, 102],[44, 121],[45, 110],
         [46, 75],[47, 91],[48, 93],[49, 116],[50, 124]]

frequencies=np.transpose(freqs)[1]

x=np.arange(sp.binom.ppf(0.0001,n,p)-0.5,
            sp.binom.ppf(0.9999,n,p)+0.5,5)

plt.figure(1)
plt.hist(frequencies, normed = True, bins = x)
```

```
x=np.arange(sp.binom.ppf(0.00001,n,p),
            sp.binom.ppf(0.99999,n,p),1)
plt.plot(x,sp.binom.pmf(x,n,p))

def max_pmf(x,k):
    return sp.binom.cdf(x,n,p)**k - sp.binom.cdf(x-1,n,p)**k

plt.plot(x,max_pmf(x,50))

def min_pmf(x,k):
    return (1-sp.binom.cdf(x-1,n,p))**k -(1-sp.binom.cdf(x,n,p))**k

plt.plot(x,min_pmf(x,50))
```

□

Example 1.10 Sampling uniform points in a square:
A similar argument applies when sampling uniformly distributed points in a square. One should not expect uniform distribution as a result, but rather binomial distribution, or its Poisson approximation. Visually, one should see some departure from the average number of points expected. In a given small area, it is quite possible that one would get more points than expected or less points that expected, thus producing areas of void and clusters.

Figure 1.8 shows nonrandom points that are uniformly filling the space, but they do not come from randomly sampled points. Such sequence is deterministic. People usually regard this situation as perfectly random. Such non-random sequences are called low discrepancy sequences and this particular example comes from so-called Sobol sequences. Using such sequences may be preferable in situations when we want to generate scenarios that are filling the space better than randomly chosen scenarios. This is called a quasi Monte-Carlo simulation, where the lower discrepancy is an important advantage.

On the contrast, Figure 1.9 shows points that were sampled randomly from the uniform distribution. Note the regions of void and clusters which are expected to appear. For the reason that uniform sampling does not uniformly fill the space, for pseudorandom points like in Figure 1.8 are preferred for Monte Carlo simulation (scenario generation).

Low discrepancy sequences are not a part of the standard Python libraries, but one can simply install a specialized library for Sobol sequences called 'sobol_seq'. It can be installed by a command

```
pip install sobol_seq
```

in the command line in the terminal. The command cannot be executed in

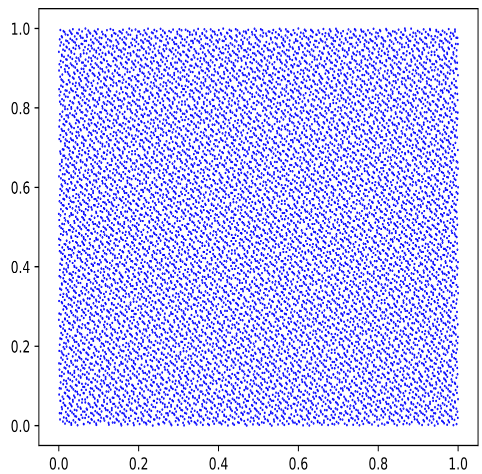

FIGURE 1.8: 12,000 pseudorandom points in a square that correspond to Sobol sequence. People usually perceive the fact that the points fill in the space in a uniform fashion as a quality associated with randomness.

Python. Once we have the library installed, the Python code corresponding to this example is

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt
import sobol_seq

a = sobol_seq.i4_sobol_generate(2,12000)
a=np.transpose(a)
plt.figure(1)
plt.plot(a[0],a[1],'b.', ms=1)

b=sp.uniform.rvs(size = (2,12000),random_state = 12345)
plt.figure(2)
plt.plot(b[0],b[1], 'b.', ms=1)
```

$\square$

Example 1.11 Waiting for a subway:
Consider the case of waiting for a subway when arriving at a random time

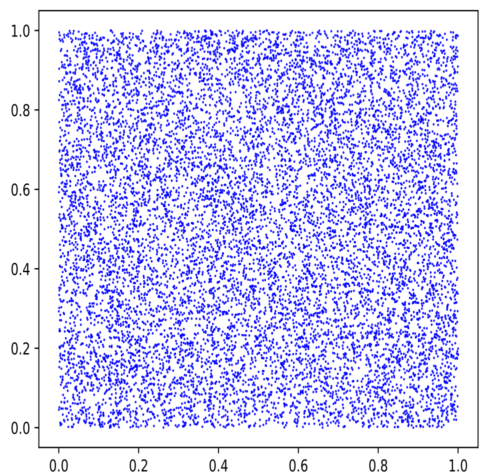

FIGURE 1.9: 12,000 random points in a square. Random points tend to cluster in some parts of the space and leave the void in other parts.

point. Suppose that the interval between two arrivals is N, and the passenger arrives at the platform at a uniformly distributed time T in the interval $[0, N]$. In other words, the random time has a density

$$f(x) = \frac{1}{N}, \qquad x \in [0, N].$$

Let A be the event $T > k$ and B the event $T > k + l$. The event A means that the subway does not arrive in k minutes, the event B means that the subway does not arrive in $k + l$ minutes. Note that $B \subset A$, and thus $A \cap B = B$. Let us consider the case that $k \leq N$ and $k + l \leq N$, otherwise the subway arrives for sure by time N. It is not difficult to see

$$\mathbb{P}(A) = 1 - \frac{k}{N}, \qquad \mathbb{P}(B) = 1 - \frac{k+l}{N}.$$

Let us determine the probability that the subway does not arrive in additional l minutes given that the passenger has already waited k minutes:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{1 - \frac{k+l}{N}}{1 - \frac{k}{N}} < 1 - \frac{l}{N}.$$

As one can see, this conditional probability is smaller than the probability of not getting the subway after just arriving at the platform and waiting for l

minutes from the scratch which is equal $1 - \frac{l}{N}$. The chances of the subway arriving are increasing as the passenger waits, so this is clearly a system with a memory. ▯

1.8 Expectation and Variance

For further analysis, it is good to know where is the mean value of the random variable, and what is its dispersion. Formally, the **expectation** of a random variable is defined as

$$\mathbb{E}[X] = \sum_{x} x \cdot \mathbb{P}(X = x)$$

for the case of the discrete distributions and as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx$$

for the case of continuous distributions. One can interpret the expectation as the center of mass of the random variable, where the weighted with their respective probabilities. The reader should distinguish between the expectation and the average. For a given distribution, expectation is a given fixed number. The average is a property of a random sample, and thus it is itself a random variable. The average itself is an estimator of the mean, but there is some inherent uncertainty as we will show in the text that follows.

Example 1.12 Expectation of Bernoulli random variable:
A Bernoulli random variable X takes only two values, 1 with probability p, and 0 with probability $1 - p$. Its expectation is given by

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

▯

REMARK 1.2 Expectation is Linear Expectation is linear in the following sense:

$$\mathbb{E}[a \cdot X + b \cdot Y] = a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y],$$

where X and Y are any two random variables and a and b are constants. The result holds regardless of the dependency of the random variables X and Y.

The linearity follows directly from the definition of the expectation:

$$\begin{aligned}\mathbb{E}[a \cdot X + b \cdot Y] &= \sum_{x,y} [a \cdot x + b \cdot y] \cdot \mathbb{P}(X = x, Y = y) \\ &= \sum_{x,y} [a \cdot x \cdot \mathbb{P}(X = x, Y = y)] + \sum_{x,y} [b \cdot y \cdot \mathbb{P}(X = x, Y = y)] \\ &= a \sum_{x} x \cdot \mathbb{P}(X = x) + b \sum_{y} y \cdot \mathbb{P}(Y = y) \\ &= a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y].\end{aligned}$$

The proof for continuous distribution is analogous. ∎

Example 1.13 Expectation of binomial distribution:
Binomial random variable Z is a sum of n Bernoulli random variables $X(i)$ that each has an expectation p. Using the linearity of expectation, we get

$$\mathbb{E}[Z] = \mathbb{E}[\sum_{i=1}^{n} X(i)] = \sum_{i=1}^{n} \mathbb{E}[X(i)] = \sum_{i=1}^{n} p = n \cdot p.$$

This can be also computed directly from

$$\mathbb{E}[Z] = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k} = n \cdot p.$$

∎

Example 1.14 Expectation of uniform distribution:
The continuous uniform distribution on the interval $[a, b]$ has the density

$$f(x) = \frac{1}{b-a} \cdot \mathbb{I}([a, b])(x);$$

i.e., it is equal to $\frac{1}{b-a}$ on the interval $[a, b]$ (1 over the length of the interval), and 0 everywhere else. The expectation is equal to

$$\begin{aligned}E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{a}^{b} \frac{x\, dx}{b-a} \\ = \frac{1}{b-a} \cdot \left[\frac{x^2}{2}\right]_{x=a}^{x=b} = \frac{1}{b-a} \cdot \frac{1}{2} \cdot (b^2 - a^2) = \frac{b+a}{2}.\end{aligned}$$

∎

Example 1.15 Expectation of Poisson distribution:
The Poisson distribution is given by the following probabilities:

$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!},$$

where $\lambda > 0$ is a parameter. Its expectation is given by

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{(k-1)!} = \lambda \sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

$\square$

Example 1.16 Expectation of geometric distribution:
The geometric random variable T is given by the following probabilities:

$$\mathbb{P}(T = k) = (1-p)^{k-1} \cdot p, \qquad k = 1, 2, \ldots$$

It expectation is equal to

$$\begin{aligned} \mathbb{E}[T] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = p \cdot \sum_{k=1}^{\infty} [-(1-p)^k]' = p \cdot \left[\sum_{k=1}^{\infty} -(1-p)^k \right]' \\ = p \cdot \left(-\frac{1}{p} \right)' = p \cdot \frac{1}{p^2} = \frac{1}{p}. \end{aligned}$$

The expected waiting time for success in Bernoulli random variable is inversely proportional to the probability of success p. $\square$

Example 1.17 Finite random variable with infinite expectation:
It is possible to have a finite random variable with infinite expectation. Consider a random variable X with the density

$$f(x) = \frac{1}{x^2}, \qquad x \geq 1.$$

We have

$$\mathbb{P}(X \leq \infty) = \int_1^{\infty} f(x)dx = \int_1^{\infty} \frac{1}{x^2} dx = \left[-\frac{1}{x} \right]_{x=1}^{x=\infty} = 0 - (-1) = 1,$$

but

$$\mathbb{E}[X] = \int_1^{\infty} x f(x) dx = \int_1^{\infty} \frac{1}{x} dx = [\ln(x)]_{x=1}^{x=\infty} = \infty.$$

$\square$

Variance of a random variable is defined as

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

In the discrete case we have

$$\mathrm{Var}(X) = \sum_x (x - \mathbb{E}[X])^2 \mathbb{P}(X = x),$$

in the continuous case we have

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x) dx.$$

Note that $\mathbb{E}[X]$ is a constant, so the variance measures the squared difference of the random variable X from its fixed mean $\mathbb{E}[X]$, and weights it correspondingly with the probability mass $\mathbb{P}$.

Note that we can expand the expectation in the definition of the variance to obtain

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X \cdot \mathbb{E}[X] + (E[X])^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (E[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.\end{aligned}$$

We have applied the linearity of the expectation several times in the above equation. The relationship

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

may be puzzling. The expression $\mathbb{E}[X^2]$ is the expectation of the random variable X^2, so the variable is squared first before taking its expectation. The expression $(\mathbb{E}[X])^2$ is the expectation of X squared, so first one takes the expectation which is then squared. Note that

$$\mathrm{Var}(X) \geq 0$$

by definition (adding and multiplying nonnegative numbers), so we must also have

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2,$$

which is just a special case of Jensen's inequality

$$E[h(X)] \geq h(E[X])$$

which holds for any convex function h.

REMARK 1.3 Variance of a linear combination of random variables: Consider random variables X and Y and constants a and b. Let us compute $\mathrm{Var}(a \cdot X + b \cdot Y)$. We have

$$\begin{aligned}
\mathrm{Var}(a \cdot X + b \cdot Y) &= \mathbb{E}[((a \cdot X + b \cdot Y) - \mathbb{E}[a \cdot X + b \cdot Y])^2] \\
&= \mathbb{E}[(a(X - \mathbb{E}[X]) + b(Y - E[Y]))^2] \\
&= a^2 \cdot \mathbb{E}[(X - \mathbb{E}[X])^2 + 2ab \cdot \mathbb{E}[(X - \mathbb{E}[X])(Y - E[Y])] \\
&\qquad\qquad + b^2 \cdot \mathbb{E}[(Y - \mathbb{E}[Y])^2 \\
&= a^2 \cdot \mathrm{Var}(X) + 2ab \cdot \mathrm{Cov}(X, Y) + b^2 \cdot \mathrm{Var}(Y).
\end{aligned}$$

We have introduced so called **covariance** of the random variables X and Y defined as

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - E[Y])].$$

For discrete distributions, we have

$$\mathrm{Cov}(X, Y) = \sum_{x,y} (x - \mathbb{E}[X])(y - \mathbb{E}[Y])\mathbb{P}(X = x, Y = y),$$

for continuous distributions we have

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f(x, y) dx dy.$$

The distribution of two random variables (X, Y) as a pair is known as a **joint distribution**. The distribution of the individual random variables X or Y is known as a **marginal distribution**. The marginal distribution is implied from the joint distribution from the relationships

$$\mathbb{P}(X = x) = \sum_{y} \mathbb{P}(X = x, Y = y)$$

in the case of the discrete distributions and

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

in the case of the continuous distributions. On the other hand, knowledge of the marginal distributions is not sufficient to determine uniquely the joint distribution. The **conditional distribution** is defined as

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

in the case of the discrete distributions and

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

in the case of the continuous distributions. ▯

A closely related concept to the covariance is a **correlation** defined as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)}}.$$

It can be shown that $-1 \le \rho(X,Y) \le 1$. The **standard deviation** $\sigma(X)$ of a random variable X is defined as the square root of the variance $\mathrm{Var}(X)$:

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

If the random variables X and Y are independent, meaning that

$$\mathbb{P}(X\in A, Y\in B) = \mathbb{P}(X\in A)\cdot \mathbb{P}(X\in B),$$

we would have $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\cdot\mathbb{P}(Y=y)$, and the covariance would simplify to

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \sum_{x,y}(x-\mathbb{E}[X])(y-\mathbb{E}[Y])\mathbb{P}(X=x,Y=y)\\
&= \sum_{x,y}(x-\mathbb{E}[X])(y-\mathbb{E}[Y])\mathbb{P}(X=x)\cdot\mathbb{P}(Y=y)\\
&= \left(\sum_{x}(x-\mathbb{E}[X])\cdot\mathbb{P}(X=x)\right)\cdot\left(\sum_{y}(y-\mathbb{E}[Y])\cdot\mathbb{P}(Y=y)\right)\\
&= 0.
\end{aligned}$$

For independent random variables, the variance of the sum is the sum of the variances.

Example 1.18 Variances of some selected distributions:

Bernoulli distribution: We have

$$\mathbb{E}[X^2] = 1^2\cdot p + 0^2\cdot(1-p) = p,$$

and thus

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

Binomial distribution: Binomial random variable Z is a sum of n independent Bernoulli random variables $Z=\sum_{i=1}^n X(i)$, so

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\sum_{i=1}^n X(i)\right) = \sum_{i=1}^n \mathrm{Var}(X(i)) = np(1-p).$$

Poisson distribution: The variance of the Poisson distribution is equal to

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

Geometric distribution: The variance of the geometric distribution is equal to

$$\mathrm{Var}(T) = \frac{1-p}{p^2}.$$

$\square$

Example 1.19 Diversification of stakes:
Consider two random variables

$$Y(i) = \begin{cases} +1 & \text{with probability } \frac{1}{2}, \\ -1 & \text{with probability } \frac{1}{2}, \end{cases}$$

for $i = 1, 2$. One can think about the random variables $Y(i)$ as profits or losses from betting on a 50:50 outcome, such as a coin toss. The expected value of the bet is 0 as

$$\mathbb{E}[Y(i)] = \tfrac{1}{2} \cdot +1 + \tfrac{1}{2} \cdot (-1) = 0.$$

The knowledge of the distribution of $Y(1)$ and $Y(2)$ does not imply the knowledge of the joint distribution $Y(1), Y(2)$. If we assume independence like in the case of two different coin tosses, we have

$$\begin{aligned}
\mathbb{P}(Y(1) = +1, Y(2) = +1) &= \mathbb{P}(Y(1) = +1) \cdot \mathbb{P}(Y(2) = +1) = \tfrac{1}{4} \\
\mathbb{P}(Y(1) = +1, Y(2) = -1) &= \mathbb{P}(Y(1) = +1) \cdot \mathbb{P}(Y(2) = -1) = \tfrac{1}{4} \\
\mathbb{P}(Y(1) = -1, Y(2) = +1) &= \mathbb{P}(Y(1) = -1) \cdot \mathbb{P}(Y(2) = +1) = \tfrac{1}{4} \\
\mathbb{P}(Y(1) = -1, Y(2) = -1) &= \mathbb{P}(Y(1) = -1) \cdot \mathbb{P}(Y(2) = -1) = \tfrac{1}{4}
\end{aligned}$$

Consider the random variable $S(2) = Y(1) + Y(2)$, the cumulative win from two bets $Y(1)$ and $Y(2)$. In the case of independence, one can see that

$$S(2) = \begin{cases} +2 & \text{with probability } \frac{1}{4}, \\ 0 & \text{with probability } \frac{1}{2}, \\ -2 & \text{with probability } \frac{1}{4}. \end{cases}$$

One can compute $\mathbb{E}[S(2)]$ and $\mathrm{Var}(S(2))$ using the distribution of $S(2)$ given above, or one can use the rules for the expectation and variance to conclude

$$\mathbb{E}[S(2)] = \mathbb{E}[Y(1) + Y(2)] = \mathbb{E}[Y(1)] + \mathbb{E}[Y(2)] = 0,$$

and

$$\begin{aligned}\mathrm{Var}(S(2)) &= \mathrm{Var}(Y(1)+Y(2)) \\ &= \mathrm{Var}(Y(1)) + \mathrm{Var}(Y(2)) + 2\,\mathrm{Cov}(Y(1),Y(2)) = 2.\end{aligned}$$

The covariance of independent random variables is zero.

As mentioned earlier, the joint distribution $\mathbb{P}(Y(1) = y_1, Y(2) = y_2)$ is not uniquely determined by the marginal distributions $\mathbb{P}(Y(1) = y_1)$ and $\mathbb{P}(Y(2) = y_2)$. For instance, the joint distribution

$$\begin{aligned}\mathbb{P}(Y(1) = +1, Y(2) = +1) &= \tfrac{1}{2} \\ \mathbb{P}(Y(1) = -1, Y(2) = -1) &= \tfrac{1}{2}\end{aligned}$$

is also consistent with the marginal distributions of $Y(1)$ and $Y(2)$. This situation would correspond to the case that the second bet has the same result as the first bet, which can happen if the player places 2 same bets in the first round and does not place any bet in the second round. One can show that in this case, $Y(1)$ and $Y(2)$ are perfectly correlated ($\rho(Y(1), Y(2)) = 1$). We have

$$\begin{aligned}\mathrm{Cov}(Y(1),Y(2)) &= \mathbb{E}[Y(1)\cdot Y(2)] - \mathbb{E}[Y(1)]\cdot\mathbb{E}[Y(2)] \\ &= 1\cdot 1\cdot \tfrac{1}{2} + (-1)\cdot(-1)\cdot\tfrac{1}{2} - 0\cdot 0 = 1,\end{aligned}$$

and thus

$$\rho(Y(1),Y(2)) = \frac{\mathrm{Cov}(Y(1),Y(2))}{\sqrt{\mathrm{Var}(Y(1))\cdot\mathrm{Var}(Y(2))}} = \frac{1}{\sqrt{1\cdot 1}} = 1.$$

The distribution of the cumulative win $S(2) = Y(1) + Y(2)$ is given by

$$S(2) = \begin{cases} +2 & \text{with probability } \tfrac{1}{2}, \\ -2 & \text{with probability } \tfrac{1}{2}. \end{cases}$$

We get

$$\mathbb{E}[S(2)] = \mathbb{E}[Y(1)+Y(2)] = \mathbb{E}[Y(1)] + \mathbb{E}[Y(2)] = 0,$$

and

$$\begin{aligned}\mathrm{Var}(S(2)) &= \mathrm{Var}(Y(1)+Y(2)) \\ &= \mathrm{Var}(Y(1)) + \mathrm{Var}(Y(2)) + 2\,\mathrm{Cov}(Y(1),Y(2)) = 4.\end{aligned}$$

Since the expectation is linear regardless of the dependence or independence of the random variables $Y(1)$ and $Y(2)$, it is not surprising that it is equal zero in both cases. The expecation of the sum must be the same in both cases. What is different is the variance of the outcome. In the case of two independent bets, the variance is equal to 2, in the case of a perfect correlation, it is equal to 4. Thus the diversification of bets leads to a smaller variance of the resulting cumulative profit and loss distribution. $\square$

Example 1.20 Number of denominations in a bridge hand:

A player of bridge get 13 cards out of 52 from the standard card deck. How many different denominations (or ranks) he is expected to get? In order to solve this problem, let us introduce a random variable $Y(i)$ which takes value 1 if the players has the i-th denomination in his hand and 0 if he does not. Each denomination has the same likelihood of appearing in the player's hand and thus $Y(i)$ has the same distribution for each denomination. Probability that one does not receive a particular denomination is equal to

$$\frac{\binom{48}{13}}{\binom{52}{13}} = \frac{6327}{20825} \approx 0.3038$$

and thus $Y(i)$ is given by

$$Y(i) = \begin{cases} 1 & \text{with probability } 1 - \frac{\binom{48}{13}}{\binom{52}{13}}, \\ 0 & \text{with probability } \frac{\binom{48}{13}}{\binom{52}{13}}. \end{cases}$$

The total number of different denominations in a bridge hand is simply

$$S = \sum_{i=1}^{13} Y(i),$$

leading to

$$\mathbb{E}[S] = \mathbb{E}[\sum_{i=1}^{13} Y(i)] = \sum_{i=1}^{13} \mathbb{E}[Y(i)] = 13 \cdot \left(1 - \frac{\binom{48}{13}}{\binom{52}{13}}\right) \approx 9.0504.$$

One should note that the variables $Y(i)$ and $Y(j)$ are dependent and thus S does not have a binomial distribution that requires independent variables. To see that $Y(i)$ and $Y(j)$ are dependent, we can compute

$$\begin{aligned} \mathrm{Cov}(Y(i), Y(j)) &= \mathbb{E}[Y(i) \cdot Y(j)] - \mathbb{E}[Y(i)] \cdot \mathbb{E}[Y(j)] \\ &= \mathbb{P}(Y(i) = 1, Y(j) = 1) - \mathbb{P}(Y(i) = 1)^2 \\ &= \frac{2134661}{4502365} - \left(\frac{14498}{20825}\right)^2 \\ &= -\frac{4945979699}{468808755625} \approx -0.01055. \end{aligned}$$

This follows from

$$\begin{aligned}
&\mathbb{P}(Y(i)=1, Y(j)=1)\\
&= \mathbb{P}(A\cap B) = 1-\mathbb{P}((A\cap B)^c)\\
&= 1-\mathbb{P}(A^c\cup B^c) = 1-(\mathbb{P}(A^c)+\mathbb{P}(B^c)-\mathbb{P}(A^c\cap B^c))\\
&= 1-(\mathbb{P}(Y(i)=0)+\mathbb{P}(Y(j)=0)-\mathbb{P}(Y(i)=0, Y(j)=0))\\
&= 1-2\cdot\frac{\binom{48}{13}}{\binom{52}{13}}+\frac{\binom{44}{13}}{\binom{52}{13}}\\
&= \frac{2134661}{4502365}\approx 0.47412.
\end{aligned}$$

This in turn allows us to compute the variance of the random variable S

$$\begin{aligned}
\operatorname{Var}(S) &= \operatorname{Var}\left(\sum_{i=1}^{13} Y(i)\right)\\
&= \sum_{i=1}^{13}\operatorname{Var}(Y(i)) + \sum_{i\neq j}\operatorname{Cov}(Y(i), Y(j))\\
&= 13\cdot\left(1-\frac{\binom{48}{13}}{\binom{52}{13}}\right)\cdot\frac{\binom{48}{13}}{\binom{52}{13}} - 13\cdot 12\cdot\frac{4945979699}{468808755625}\\
&\approx 2.74966-1.64582 = 1.10385.
\end{aligned}$$

If the variables $Y(i)$ were independent, their sum would have a variance of 2.74966. However, the sum S has a smaller variance 1.10385, meaning that the variable S is distributed closer to its mean 9.0504 than if the variables were independent.

TABLE 1.2: Distribution of S compared to the distribution of the corresponding binomial random variable.

n	S	Binomial
0	0	0.000000
1	0	0.000006
2	0	0.000076
3	0	0.000646
4	0.000001	0.003704
5	0.000151	0.015279
6	0.005511	0.046681
7	0.058308	0.106968
8	0.229885	0.183835
9	0.374231	0.234027
10	0.256202	0.214504
11	0.069422	0.134052
12	0.006182	0.051195
13	0.000105	0.009024

$\square$

Example 1.21 Bridge hand simulation:

Let us illustrate the previous example with the corresponding computer simulation. As the first step, we should be able to simulate how to deal 13 cards out of 52. For the sake of simplicity, let us consider that each number from $[1, 2, 3, 4, 5, \ldots, 49, 50, 51, 52]$ represents one card, namely $[A\clubsuit, A\spadesuit, A\diamondsuit, A\heartsuit, 2\clubsuit, \ldots, K\clubsuit, K\spadesuit, K\diamondsuit, K\heartsuit]$. We can generate a random bridge hand by picking 13 random numbers from 52. This can be done in the library numpy, command random.choice. We have to specify the list of 52 values (np.arange(1,53)), selection to be picked (size = 13) and the fact that the numbers are selected without the replacement (replace = False). In addition, let us select the previously used random state 12345 that makes the random generation reproducible.

```
import numpy as np
np.random.seed(seed=12345)
a= np.random.choice(np.arange(1,53),size=13, replace=False)
```

that gives

```
a= array([27, 40, 39, 26, 7, 29, 31, 13, 36, 45, 48, 9, 22])
```

that corresponds to the hand (unordered)

$$[7\diamondsuit, 10\heartsuit, 10\diamondsuit, 7\spadesuit, 2\diamondsuit, 8\clubsuit, 8\diamondsuit, 4\clubsuit, 9\heartsuit, Q\clubsuit, Q\heartsuit, 3\clubsuit, 6\spadesuit]$$

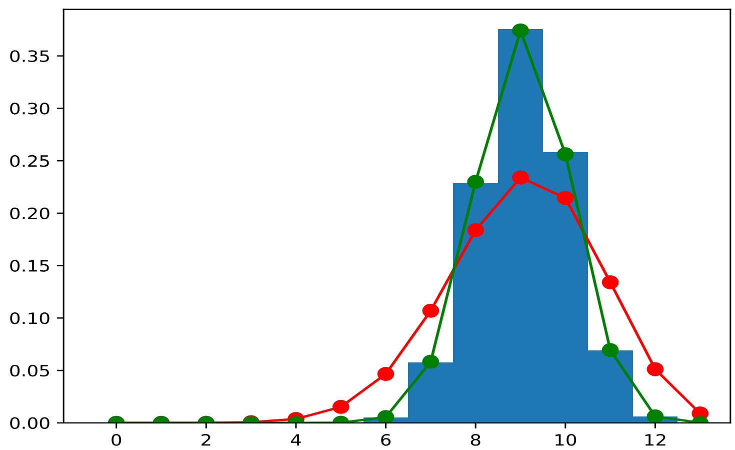

FIGURE 1.10: Results from simulations of the number of denominations in a bridge hand (histogram) in comparison to the expected frequencies (green) and the binomial distribution (red). Note that the distribution of the number of denominations has smaller variability in comparison to the binomial distribution and that the simulated values fit the theoretical frequencies pretty well.

This gives exactly 9 different denominations as we have four pairs: 7's, 8's, 10's and Q's. The mathematical idea that Python can use to count these different denominations is the following. Each denomination corresponds to four numbers: A to [1,2,3,4], 2 to [5,6,7,8], up to K that corresponds to [49,50,51,52]. These groups give the same number if divided by four and rounded up. The corresponding code that applies this idea on the array `a` is

```
ceil=np.vectorize(np.ceil)
ceil(a/4)
```

Function 'ceil' can be applied to arrays after calling command np.vectorize. The above code gives

```
[7., 10., 10.,  7.,  2.,  8.,  8.,  4.,  9.,  12., 12.,  3.,  6.]
```

so we have a one to one correspondence of the card numbers to the denominations. If you recall, we can create an object containing only unique elements by calling

```
set(ceil(a/4))
```

which gives

```
{2.0, 3.0, 4.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.0}
```

The length of this set represents the number of unique denominations in the bridge hand and this is obtained by

```
len(set(ceil(a/4)))
```

which is indeed 9.

Once we are in a position to simulate a random bridge hand and count the number of different denominations, we can repeat this procedure many times and estimate the corresponding frequencies for each possible value of different denominations.

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.special

simulations = 100000
ceil=np.vectorize(np.ceil)
denominations = np.empty(simulations
np.random.seed(seed=12345)
a= np.random.choice(np.arange(1,53),size=13, replace=False)

for i in range(simulations):
    a= np.random.choice(np.arange(1,53),size=13, replace=False)
    denominations[i]=len(set(ceil(a/4)))
    if i%5000 == 0: print(i)

x=np.arange(-0.5,13.5,1)
weights = np.ones_like(denominations)/float(len(denominations))
plt.figure(1)
plt.hist(denominations,weights=weights, bins=x)

x=np.arange(0,14)
pp = 1 - scipy.special.binom(48,13)/scipy.special.binom(52,13)
plt.plot(sp.binom.pmf(x,13,pp),'ro',ms=8)
plt.plot(sp.binom.pmf(x,13,pp),'r')

y=[0,0,0,0,0.000001,0.000151,0.005511,0.058308,
 0.229885,0.374231,0.256202,0.069422,0.006182,0.000105]
plt.plot(x,y,'go',ms=8)
plt.plot(x,y,'g')
```

that results in the following Figure 1.10. Note that the sample mean of the simulated denominations given by

```
np.mean(denominations)
```

gives 9.05515, which is close enough to the theoretical mean 9.0504. ∎

1.9 Behavior of the Large Sample

In many practical situations we encounter random variables that are results of many repeated trials. For instance, the return of a portfolio over a long period is a sum of daily returns, the total number of insurance claims is a sum of individual claims, etc. Let $X(i)$ be the random variable corresponding to the i-th observation. Assume that the random variables $X(i)$ are independent and identically distributed. For simplicity, assume that both the expectation and the variance are finite with $\mathbb{E}[X(i)] = \mu$ and $\text{Var}(X) = \sigma^2$.

Define $S(n) = \sum_{i=1}^{n} X(i)$ and study the asymptotic behavior of $S(n)$ as n goes to infinity. A simple computation leads to

$$\mathbb{E}[S(n)] = \mathbb{E}\left[\sum_{i=1}^{n} X(i)\right] = \sum_{i=1}^{n} \mathbb{E}[X(i)] = n \cdot \mu,$$

and

$$\text{Var}(S(n)) = \text{Var}\left(\sum_{i=1}^{n} X(i)\right) = \sum_{i=1}^{n} \text{Var}(X(i)) = n \cdot \sigma^2 \to \infty.$$

As n goes to infinity, the variance of the sum $S(n)$ explodes.

Consider now the (sample) average defined as

$$\bar{X}(n) = \frac{S(n)}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} X(i)$$

Keep in mind that the average itself is a random variable in contrast to the expectation which is a fixed number. When we compute the expectation of the average $\bar{X}(n)$, we get

$$\mathbb{E}[\bar{X}(n)] = \mathbb{E}\left[\frac{S(n)}{n}\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{E}[X(i)] = \mu.$$

When we compute the variance, we get

$$\text{Var}\left(\bar{X}(n)\right) = \text{Var}\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X(i)\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \text{Var}[X(i)] = \frac{\sigma^2}{n} \to 0.$$

This computation assumes that the random variables $X(i)$ have a finite variance $\mathrm{Var}(X) = \sigma^2$. The sample average has the same expectation as every individual $X(i)$, but its variance is getting smaller as n increases. In the limit, the variability of the sample average is zero, and thus it converges to a constant.

This is an important result known as the **Law of Large Numbers,** which says that the sample average converges to the theoretical expectation when the sample size converges to infinity:

$$\boxed{\lim_{n\to\infty} \bar{X}(n) = \mu.}$$

The Law of Large Numbers is valid also in the situation when the random variable X does not have a finite variance. The expectation μ can be also infinite, in which case the sample average would converge to infinity. More specifically, consider a level $\epsilon > 0$, and an interval $(\mu - \epsilon, \mu + \epsilon)$. The level ϵ can be regarded as a pre-described acceptable tolerance from the true value. The Law of Large Numbers states that for any specified tolerance level ϵ there exists some K such that $\bar{X}(n)$ will stay in the interval $(\mu - \epsilon, \mu + \epsilon)$ for all $n \geq K$. The index K clearly depends on ϵ (the smaller is the ϵ, the larger K must be), but it also depends on the random realizations of $X(1), X(2), \ldots$. The sample average may converge faster to the theoretical mean in some scenarios than in some other scenarios.

As a consequence of the Law of Large Numbers, when $\mu > 0$, the sample average must eventually cross into the positive region and stay there forever. Similarly, when $\mu < 0$, the sample average must eventually cross int the negative region and stay there forever.

1.10 Central Limit Theorem

We have seen that $\mathrm{Var}(S(n)) \to \infty$, but $\mathrm{Var}\left(\frac{S(n)}{n}\right) \to 0$. As both cases are extreme, a natural question is whether it is possible to scale $S(n)$ so that it would have a finite and positive variance. In order to achieve this goal, we need to find the corresponding random variable that keeps constant variance as a function of n. It is helpful to introduce the concept of a **standard score**, which is also known as a z-score. Given a random variable Y that has an expected value $\mu = \mathbb{E}[Y]$ and variance $\sigma^2 = \mathrm{Var}(Y)$, we can define a new random variable X as

$$X = \frac{Y - \mu}{\sigma}.$$

This represents a number that measures how many standard deviations σ is the random variable Y from its center μ. It is straightforward to compute

$$\mathbb{E}[X] = \mathbb{E}\left[\frac{Y-\mu}{\sigma}\right] = \frac{1}{\sigma} \cdot \mathbb{E}[Y-\mu] = 0$$

and

$$\mathrm{Var}(X) = \mathbb{E}\left[\frac{Y-\mu}{\sigma}\right]^2 = \frac{1}{\sigma^2} \cdot \mathbb{E}[Y-\mu]^2 = \frac{\mathrm{Var}(Y)}{\sigma^2} = 1.$$

Applying the standard score transformation on the random variable $S(n)$, we get

$$\frac{S(n) - \mathbb{E}[S(n)]}{\sqrt{\mathrm{Var}(S(n))}} = \frac{S(n) - n\mu}{\sqrt{n}\sigma},$$

which is a random variable with zero mean and unit variance regardless of the number of observations n. Scaling the variable $S(n)$ minus its expectation by a factor $\sqrt{n}$ results in a random variable with a non-trivial distribution.

The above result holds even in the situation when $n \to \infty$. The resulting random variable will have a zero mean and a unit variance. The following result shows that the limiting random variable will have a standard normal distribution $N(0,1)$. This result is called
Central Limit Theorem:

$$\mathbb{P}\left(\frac{S(n) - n \cdot \mu}{\sqrt{n} \cdot \sigma} \leq x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}.$$

The random variable with the density

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

is known as the **standard normal random variable.** We will denote it by $N(0,1)$. The standard normal random variable plays a major role in probability and statistics, and thus we will denote its density by $\phi(x)$ and its cumulative distribution function by $\Phi(x)$. Figure 1.11 shows the density function of the standard normal random variable, and the corresponding Python code to generate it is

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

x = np.linspace(sp.norm.ppf(0.001),sp.norm.ppf(0.999), 100)
plt.plot(x,sp.norm.pdf(x))
```

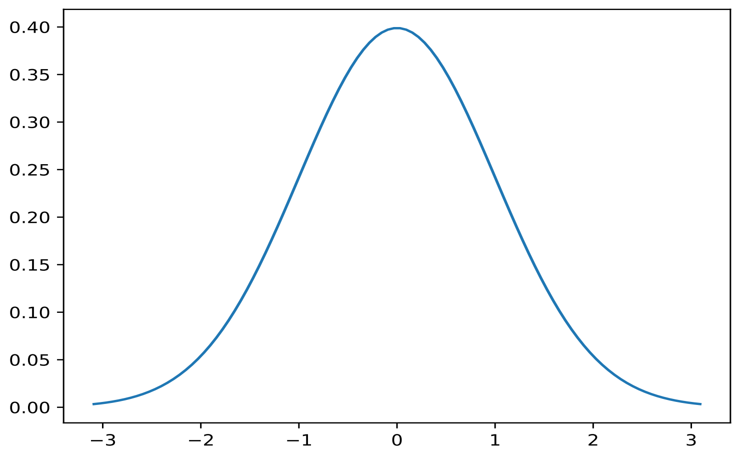

FIGURE 1.11: The density function of the standard normal variable $N(0, 1)$.

More generally, a normal random variable $N(\mu, \sigma)$ that has a mean μ and a standard deviation σ has the density

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

Let Y be such a random variable $N(\mu, \sigma)$. We can always transform it to the standard normal random variable $N(0, 1)$ by shifting it by μ and scaling it by σ. In other words, the random variable

$$X = \frac{Y - \mu}{\sigma}$$

has $N(0, 1)$ distribution. Obviously we can also get the random variable $N(\mu, \sigma)$ from $N(0, 1)$ by inverting the previous relationship

$$Y = \sigma \cdot X + \mu.$$

The Central Limit Theorem states that

$$\frac{S(n) - n \cdot \mu}{\sqrt{n} \cdot \sigma} \approx N(0, 1).$$

If we use the above transformation, we can write that

$$S(n) \approx N(n \cdot \mu, \sigma\sqrt{n}).$$

This is a slight abuse of the notation as the above expression does not converge when $n \to \infty$ in contrast to the original statement of the Central Limit Theorem. But since we always work with finite n, we will apply the above expression in our applications.

Example 1.22

Consider tossing a fair coin 1000 times and count the number of heads. The number of heads $S(n)$ has a binomial distribution $Bin(1000, \frac{1}{2})$. According to the Central Limit Theorem,

$$Bin(n,p) \approx N(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}).$$

Keep in mind that $\mu = p$ and $\sigma^2 = p \cdot (1-p)$ are the mean and the variance of the Bernoulli random variable. Therefore $Bin(1000, \frac{1}{2})$ is approximately $N(500, \sqrt{250})$.

In order to use the approximation to compute the probabilities, we need to make an extra step as the binomial distribution is discrete, while the normal distribution is continuous. For the binomial distribution, we may write for a given integer k

$$\mathbb{P}(S(n) = k) = \mathbb{P}(S(n) \leq k + 0.5) - \mathbb{P}(S(n) \leq k - 0.5)$$

as there is no probability mass outside the integers. But the Central Limit Theorem states that

$$\mathbb{P}(S(n) \leq k) \approx F(k),$$

where $F(x)$ is the cumulative distribution function of a normal variable $N(n\mu, \sigma\sqrt{n})$. Thus

$$\mathbb{P}(S(n) = k) \approx F(k + 0.5) - F(k - 0.5)$$

using the corresponding cumulative distribution function of the normal distribution.

Figure 1.12 illustrates the comparison of the probability mass function from $Bin(1000, \frac{1}{2})$, the density of $N(500, \sqrt{250})$ and the probabilities computed from the difference of the two cumulative distribution functions from the normal distribution. The probability mass function of the binomial distribution and the normal density has a very good fit and the resulting probabilities are impossible to indistinguish as the difference between the values is negligent. In order to show the exact difference of the exact and approximative probability, we plot Figure 1.13.

The Python code that generates these two plots is the following:

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

n=1000
p=1/2
x = np.arange(sp.binom.ppf(0.001, n, p), sp.binom.ppf(0.999, n, p))

normapprox =[sp.norm.cdf(i+0.5,n*p,np.sqrt(n*p*(1-p)))-
             sp.norm.cdf(i-0.5,n*p,np.sqrt(n*p*(1-p))) for i in x]

plt.figure(1)
plt.plot(x, sp.binom.pmf(x, n, p), 'ro', ms=4)
plt.plot(x,normapprox,'go', ms=2)
plt.plot(x,sp.norm.pdf(x,n*p,np.sqrt(n*p*(1-p))))

plt.figure(2)
plt.plot(x,sp.binom.pmf(x, n, p)- normapprox,'bo', ms=3)

alpha = 0.95
sp.binom.interval(alpha, n, p)
```

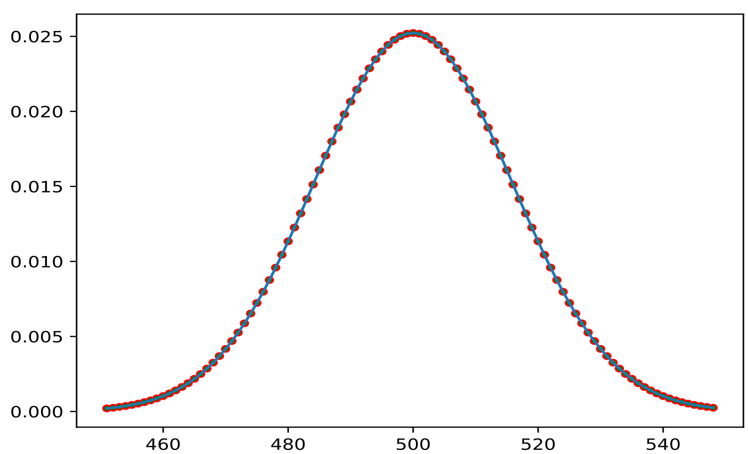

FIGURE 1.12: Comparison of the probability mass function $Bin(1000, \frac{1}{2})$ and the density of $N(500, \sqrt{250})$. Any difference is barely visible.

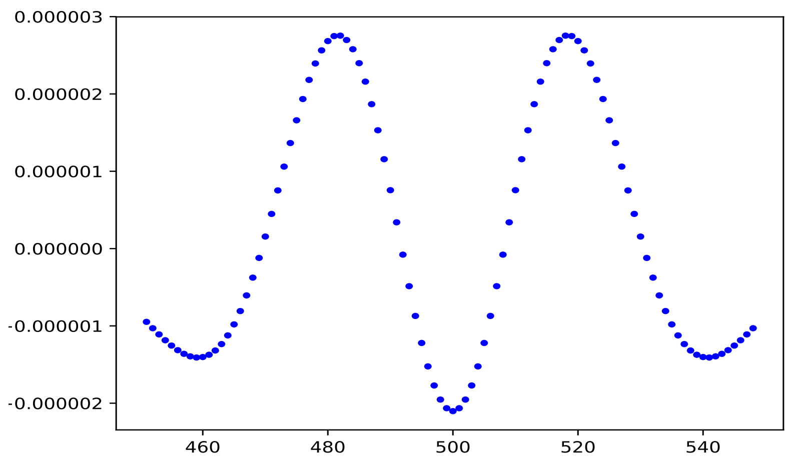

FIGURE 1.13: The difference of probabilities from $Bin(1000, \frac{1}{2})$ and from $N(500, \sqrt{250})$.

$\square$

1.11 Typical Values of the Random Variable

One of the main applications of probability is an identification of the typical values of a given random variable. A closely connected problem is to identify the likely range of the parameters of a given distribution based on the observed data. Typical values of the random variable are characterized by a **confidence interval.** Given a value α, we want to find an interval $[a, b]$, such that the random variable X falls in that interval with at least probability of α:

$$\mathbb{P}(X \in [a, b]) = \alpha.$$

Typical choices of α are 0.95 (1 in 20 chance that the random variable falls outside the corresponding confidence interval) and 0.99 (1 in 100 chance that the random variable falls outside the corresponding confidence interval).

Finding the values of a and b is related to a concept of a **quantile.** The quantile is defined as the value q_α for which

$$\mathbb{P}(X \leq q_\alpha) = \alpha.$$

This is a well defined number when the random variable is continuous and strictly monotonic like in the case of the standard normal distribution. The probability distribution has α fraction of the mass to the left of the quantile q_α and $1-\alpha$ fraction of the mass to the right of the quantile q_α. The quantile corresponding to $\alpha = 0.5$ is also known as the median of the distribution (it splits 0.5 mass of the distribution to the left and to the right). Table 1.3 gives a list of some representative quantiles of the standard normal distribution.

TABLE 1.3: Quantiles of the standard normal distribution

α	0.005	0.01	0.025	0.05	0.5	0.95	0.975	0.99	0.995
q_α	-2.576	-2.326	-1.960	-1.645	0	1.645	1.960	2.326	2.576

The interval $[q_{\frac{1-\alpha}{2}}, q_{\frac{1+\alpha}{2}}]$ is an α confidence interval as

$$\mathbb{P}(q_{\frac{1-\alpha}{2}} \leq X \leq q_{\frac{1+\alpha}{2}}) = \alpha.$$

This is known as a **two sided confidence interval** as the probability is split in the pieces to the left and to the right of the confidence interval: $\frac{1-\alpha}{2}$ mass is to the left of the quantile $q_{\frac{1-\alpha}{2}}$ and $\frac{1-\alpha}{2}$ mass is to the right of the quantile $q_{\frac{1+\alpha}{2}}$.

Example 1.23
The 95% two sided confidence interval for the standard normal variable X is given by $[q_{0.025}, q_{0.975}] = [-1.96, 1.96]$. There is 2.5% probability that the standard normal random variable will be lower than -1.96 and 2.5% probability that the standard normal random variable will be higher than 1.96. Similar argument leads to the confidence interval of a normal random variable $Y \sim N(\mu, \sigma)$. Then Y is related to X by $Y = \sigma X + \mu$. The confidence interval for Y is obtained in the following way:

$$\begin{aligned}\alpha = \mathbb{P}(q_{\frac{1-\alpha}{2}} \leq X \leq q_{\frac{1+\alpha}{2}}) &= \mathbb{P}\left(q_{\frac{1-\alpha}{2}} \leq \tfrac{Y-\mu}{\sigma} \leq q_{\frac{1+\alpha}{2}}\right) \\ &= \mathbb{P}\left(\mu + \sigma \cdot q_{\frac{1-\alpha}{2}} \leq Y \leq \mu + \sigma \cdot q_{\frac{1+\alpha}{2}}\right),\end{aligned}$$

where $q_{\frac{1-\alpha}{2}}$ and $q_{\frac{1+\alpha}{2}}$ are the quantiles of the standard normal distribution. For instance, 95% two sided confidence interval for $Y \sim N(\mu, \sigma)$ is given by $[\mu - 1.96\sigma, \mu + 1.96\sigma]$. In this sense, about 95% of the values of the normal distribution lie within 2 standard deviations away from the mean μ.

Consider the example of 1000 coin tosses of a fair toss. The number of heads follows a binomial distribution $Z \sim Bin(1000, \frac{1}{2})$, which is approximately

$N(500, \sqrt{250})$ according to the Central Limit Theorem. Using this approximation, the 95% confidence interval is given by $500 \pm 1.96 \cdot \sqrt{250} = [469, 531]$. If we compute the probabilities exactly from the cumulative function of the binomial distribution, we get

$$\mathbb{P}(Z \leq 468) = \mathbb{P}(Z \geq 532) = 0.0231456,$$

$$\mathbb{P}(Z \leq 469) = \mathbb{P}(Z \geq 531) = 0.0268389,$$

and

$$\mathbb{P}(469 \leq Z \leq 531) = 0.953709,$$

so indeed the normal approximation found the correct 95% confidence interval. $\square$

The above example showed that in discrete distributions, there may not exist a point that would split the probability mass to α to the left and $1-\alpha$ to the right as required for a quantile. For instance if we wanted to find a quantile $q_{0.025}$, it would have to be somewhere between 468 and 469 as

$$\mathbb{P}(Z \leq 468) = 0.0231456, \qquad \mathbb{P}(Z \leq 469) = 0.0268389.$$

A more general definition of the q_α quantile picks the smallest x such that $\mathbb{P}(Z \leq x) \geq \alpha$. More formally,

$$q_\alpha = \min\{x : \mathbb{P}(Z \leq x) \geq \alpha\}.$$

According to this definition, $q_{0.025} = 469$, and $q_{0.975} = 531$.

In some cases, we may be interested to exclude unlikely cases in one direction only. For instance, when studying the distribution of the returns, it is more important to address unusually large negative returns. Unusually large positive returns are good for the portfolio management, and thus one is generally not concerned about this case. This leads to the concept of **one sided confidence interval.** We want to find a or b such that

$$\mathbb{P}\left(X \in [a, \infty)\right) = \alpha,$$

or

$$\mathbb{P}\left(X \in (-\infty, b]\right) = \alpha.$$

When the random variable is continuous and strictly monotonic, we have

$$\mathbb{P}(q_{1-\alpha} \leq X) = \alpha,$$

and

$$\mathbb{P}(X \leq q_\alpha) = \alpha,$$

and thus the corresponding confidence intervals are given by

$$[q_{1-\alpha}, \infty), \qquad \text{and} \qquad (-\infty, q_\alpha].$$

Example 1.24 Time of the first Poisson jump:
Stochastic process is a sequence of random variables indexed by time. One prominent example is a Poisson process that counts a number of certain events within a given time interval. For instance the number of defaults in a given portfolio or a number of goals in football follows Poisson distribution $N(t)$, where

$$\mathbb{P}(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

It is a Poisson random variable with the parameter λt. We have seen that $\mathbb{E}[N(t)] = \lambda t$, so the expectation is linearly increasing in time. The process starts at $N(0) = 0$ and it jumps up by one at each time of the jump of the process $N(t)$. Let τ be the time of the first jump defined as

$$\tau = \min\{t \geq 0 : N(t) = 1\}.$$

We can compute the cumulative distribution function of τ using

$$F(t) = \mathbb{P}(\tau \leq t) = \mathbb{P}(N(t) \geq 1) = 1 - \mathbb{P}(N(t) = 0) = 1 - e^{-\lambda t}.$$

We have used the fact that if the jump happened before time t, the level of $N(t)$ must be one or higher. The distribution of τ has a density given by

$$f(t) = F'(t) = \lambda e^{-\lambda t}, \qquad t \geq 0.$$

This is known as the **exponential distribution**. The expected value of τ is given by

$$\mathbb{E}[\tau] = \int_0^\infty t\lambda e^{-\lambda t}dt = \frac{1}{\lambda}.$$

This is perfectly intuitive, the larger parameter λ leads to the larger expectation of $N(t)$, and thus the time between the jumps must be smaller and inversely proportional to the intensity λ.

Consider the case of scoring in a football game. Assume that the total number of goals scored by both teams is $\lambda = 3$. It is helpful to introduce the game time

$$T = 1 \text{ game time} = 90 \text{ minutes}.$$

In this case we have

$$N(1 \text{ game time}) = 3,$$

and

$$\mathbb{E}[\tau] = \frac{1}{3} \text{ game time} = 30 \text{ minutes}.$$

Therefore the expected time between two goals is 30 minutes. Let us determine the one sided 95% confidence interval for τ in the form $[0, t_{0.95}]$. We want to find $t_{0.95}$ (95% quantile of the exponential distribution) such that $F(t_{0.95}) = \mathbb{P}(\tau \leq t_{0.95}) = 0.95$. Since $F(t) = 1 - e^{-\lambda t}$, we can solve for $t_{0.95}$ to get

$$t_{0.95} = -\frac{\ln(0.05)}{\lambda} = 0.998577 \text{ game time.}$$

Thus the one sided 95% confidence interval is approximately $[0, 1]$, meaning that the first goal is expected to happen in 19 out of 20 cases. Only in 5% scenarios (1 in 20) the game would end scoreless under the assumption $\lambda = 3$. $\square$

1.12 Roulette Case Study: Central Limit Theorem, Law of Large Numbers

Consider two different betting strategies in roulette, one that bets on the red and one that bets on a single number (say on 13). We consider the roulette with 37 outcomes numbered $0, 1, \ldots, 36$, 18 of them are red. The the random variable that describes profit and loss in the first bet has the following distribution.

$$X_1 = \begin{cases} +1 & \text{if red, with probability } \frac{18}{37}, \\ -1 & \text{if black or zero, with probability } \frac{19}{37}. \end{cases}$$

Similarly, the random variable that describes profit and loss in the bet on a single number is given by

$$X_2 = \begin{cases} +35 & \text{if 13, with probability } \frac{1}{37}, \\ -1 & \text{if not 13, with probability } \frac{36}{37}. \end{cases}$$

It is not difficult to see that

$$\mathbb{E}[X_1] = 1 \cdot \tfrac{18}{37} - 1 \cdot \tfrac{19}{37} = -\tfrac{1}{37},$$

and

$$\mathbb{E}[X_2] = 35 \cdot \tfrac{1}{37} - 1 \cdot \tfrac{36}{37} = -\tfrac{1}{37},$$

and thus the two betting approaches have the same mean. From the risk management point of view, the expected value is negative and thus an investor should not take it.

$$\mathrm{Var}(X_1) = \mathbb{E}[(X_1)^2] - (\mathbb{E}[X_1])^2 = 1^2 \cdot \tfrac{18}{37} + (-1)^2 \cdot \tfrac{19}{37} - (-\tfrac{1}{37})^2 = \tfrac{1368}{1369} \approx 1.$$

$$\begin{aligned}\operatorname{Var}(X_2) &= \mathbb{E}[(X_2)^2] - (\mathbb{E}[X_2])^2 \\ &= 35^2 \cdot \tfrac{18}{37} + (-1)^2 \cdot \tfrac{19}{37} - (-\tfrac{1}{37})^2 = \tfrac{46656}{1369} \approx 34.0804.\end{aligned}$$

Let $S_1(n) = \sum_{i=1}^n X_1(i)$ be the cumulative win when betting on red and $S_2(n) = \sum_{i=1}^n X_2(i)$ be the cumulative win when betting on 13. The random variables $S_1(n)$ and $S_2(n)$ are related to the binomial distribution in the following way. If we introduce the random variables that take the value 1 when the bet is successful and the value 0 when the bet is not successful by

$$Y_1 = \begin{cases} 1 & \text{if red, with probability } \frac{18}{37}, \\ 0 & \text{if black or zero, with probability } \frac{19}{37}, \end{cases}$$

and

$$Y_2 = \begin{cases} 1 & \text{if 13, with probability } \frac{1}{37}, \\ 0 & \text{if not 13, with probability } \frac{36}{37}, \end{cases}$$

we can write $X_1 = 2 \cdot Y_1 - 1$, and $X_2 = 36 \cdot Y_2 - 1$. Thus we have

$$S_1(n) = \sum_{i=1}^n X_1(i) = \sum_{i=1}^n (2 \cdot Y_1(i) - 1) = 2 \cdot \sum_{i=1}^n Y_1(i) - n,$$

$$S_2(n) = \sum_{i=1}^n X_2(i) = \sum_{i=1}^n (36 \cdot Y_2(i) - 1) = 36 \cdot \sum_{i=1}^n Y_2(i) - n.$$

It is not clear what of the two strategies to follow if the player (gambler in this case) decides to play. Both strategies have the same negative expectation so from this respect they equally unappealing. Let us consider the case that the player wants to have a large $\mathbb{P}(S(n) \geq 0)$, the probability of not losing after n turns. Let us compute this probability. Note that

$$\begin{aligned}\mathbb{P}(S_1(n) \geq 0) &= \mathbb{P}\left(2 \cdot \sum_{i=1}^n Y_1(i) - n \geq 0\right) \\ &= \mathbb{P}\left(\sum_{i=1}^n Y_1(i) \geq \frac{n}{2}\right) = 1 - \mathbb{P}\left(\sum_{i=1}^n Y_1(i) < \frac{n}{2}\right) \\ &= 1 - \mathbb{P}\left(\sum_{i=1}^n Y_1(i) \leq \frac{n-1}{2}\right) = 1 - F_{Bin(n,\frac{18}{37})}\left(\frac{n-1}{2}\right),\end{aligned}$$

where $F_{Bin(n,\frac{18}{37})}$ is the cumulative distribution function of $Bin(n, \frac{18}{37})$ random variable. Similarly,

$$\begin{aligned}\mathbb{P}(S_2(n) \geq 0) &= \mathbb{P}\left(36 \cdot \sum_{i=1}^{n} Y_2(i) - n \geq 0\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} Y_2(i) \geq \frac{n}{36}\right) = 1 - \mathbb{P}\left(\sum_{i=1}^{n} Y_2(i) < \frac{n}{36}\right) \\ &= 1 - \mathbb{P}\left(\sum_{i=1}^{n} Y_2(i) \leq \frac{n-1}{36}\right) = 1 - F_{Bin(n,\frac{1}{37})}\left(\frac{n-1}{36}\right),\end{aligned}$$

where $F_{Bin(n,\frac{1}{37})}$ is the cumulative distribution function of $Bin(n, \frac{1}{37})$ random variable.

Before we present the results, note that

$$\begin{aligned}&\mathbb{P}(S_1(1) \geq 0) = \mathbb{P}(S_1(1) > 0) = \tfrac{18}{37} \approx 0.486486\\ &\mathbb{P}(S_1(2) \geq 0) = 1 - \mathbb{P}(S_1(2) < 0) = 1 - (\tfrac{19}{37})^2 = \tfrac{1008}{1369} \approx 0.736304,\\ &\mathbb{P}(S_1(2) > 0) = (\tfrac{18}{37})^2 = \tfrac{324}{1369} \approx 0.236669\\ &\mathbb{P}(S_1(3) \geq 0) = \mathbb{P}(S_1(3) > 0) \approx 0.479735.\end{aligned}$$

As one can see, the $\mathbb{P}(S_1(n) \geq 0)$ and $\mathbb{P}(S_1(n) > 0)$ widely oscilate for these relatively small n. This is influenced by the probability that one hits zero exactly $\mathbb{P}(S_1(n) = 0)$, which can happen in even number of steps. This probability is relatively big for small even numbers n. To make the result smoother and more intuitive, consider the quantity

$$\frac{1}{2} \cdot \Big(\mathbb{P}(S_1(n) \geq 0) + \mathbb{P}(S_1(n) > 0)\Big) = \mathbb{P}(S_1(n) > 0) + \frac{1}{2} \cdot \mathbb{P}(S_1(n) = 0).$$

One can regard that when hitting zero, it compares only as one half the weight of being positive. This quantity will not oscilate when considering of betting on red. Figure 1.14 shows $\mathbb{P}(S(n) > 0) + \frac{1}{2} \cdot \mathbb{P}(S(n) = 0)$ as computed directly from the corresponding cumulative distribution functions of the binomial distribution in both cases when betting on red and when betting on 13. The corresponding Python code is

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

n=np.arange(1,1001)

def above_zero(n):
    return sp.binom.sf(n/2,n,18/37) + 0.5*sp.binom.pmf(n/2,n,18/37)
```

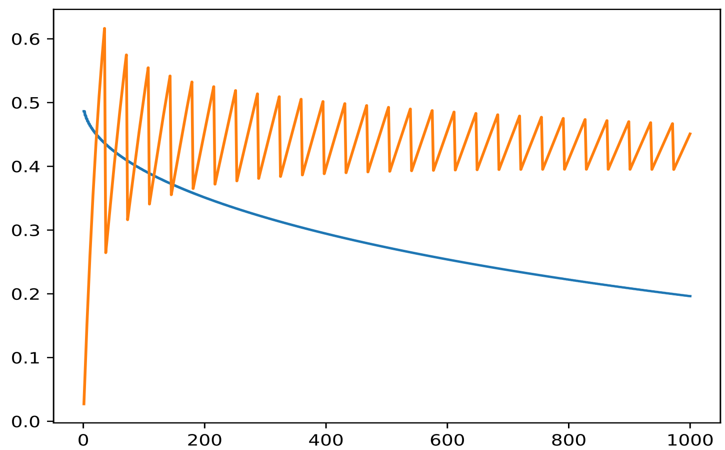

FIGURE 1.14: Probabilities of being positive and half of the probabilities of being at zero ($\mathbb{P}(S(n) > 0) + \frac{1}{2} \cdot \mathbb{P}(S(n) = 0)$) when betting on red and on 13.

```
def above_zero_single(n):
    return sp.binom.sf(n/36,n,1/37) + 0.5*sp.binom.pmf(n/36,n,1/37)

plt.plot(n,above_zero(n))
plt.plot(n,above_zero_single(n))
```

Figure 1.14 requires some additional explanation. It is not surprising that when betting on red, the probability of being positive starts slightly above 0.5 for $n = 1$ (around 0.486486), and it gradually decreases. This is expected. The probability of being above zero after 1000 bets is about 1 in 5 (around 0.196383), which is still reasonable given the fact that betting in roulette has a negative drift. What is more surprising is the probability of winning when betting on a single number. This probability starts at small numbers when n is small, but it quickly peaks and it even gets higher than 0.5 which may seem surprising. The graph resembles a chainsaw and it is trending down only very slowly in the first 1000 bets. The probability of being above zero after one thousand bets on a single number is relatively high, it is equal to 0.451149.

If we take a more detailed look, this is also expected. Recall that

$$\mathbb{P}(S_2(n) \geq 0) = \mathbb{P}\left(\sum_{i=1}^{n} Y_2(i) \geq \frac{n}{36}\right),$$

so in order to be nonnegative in the first 36 bets, it is enough to have one win, in order to be nonnegative in the first 72 bets, it is enough to have two wins, and so on. This probability may get relatively high, for instance

$$\mathbb{P}(S_2(36) \geq 0) = \mathbb{P}\left(\sum_{i=1}^{n} Y_2(i) \geq 1\right)$$
$$= 1 - \mathbb{P}\left(\sum_{i=1}^{n} Y_2(i) = 0\right) = 1 - \left(\frac{36}{37}\right)^{36} \approx 0.627069.$$

Advances in the computing technology allow us these days to work directly with distributions of the type $Bin(n, \frac{18}{37})$ and $Bin(n, \frac{1}{37})$ for relatively large n, but it still poses a challenge to work with them directly, for instance when computing the confidence intervals. Thus it is still useful to use the normal approximation to simplify our computations. From the Central Limit Theorem, we have

$$S_1(n) \approx N(n \cdot \mu_1, \sigma_1\sqrt{n}) = N\left(-\tfrac{n}{37}, \sqrt{\tfrac{1368}{1369} \cdot n}\right).$$

and

$$S_2(n) \approx N(n \cdot \mu_2, \sigma_2\sqrt{n}) = N\left(-\tfrac{n}{37}, \sqrt{\tfrac{46656}{1369} \cdot n}\right).$$

For instance, 95% confidence interval for $S_1(n)$ is approximately given by

$$\left(-\tfrac{n}{37} - 1.96 \cdot \sqrt{\tfrac{1368}{1369} \cdot n}, -\tfrac{n}{37} + 1.96 \cdot \sqrt{\tfrac{1368}{1369} \cdot n}\right),$$

95% confidence interval for $S_2(n)$ is approximately given by

$$\left(-\tfrac{n}{37} - 1.96 \cdot \sqrt{\tfrac{46656}{1369} \cdot n}, -\tfrac{n}{37} + 1.96 \cdot \sqrt{\tfrac{46656}{1369} \cdot n}\right).$$

We can simulate the evolution of the profit loss from the two betting strategies in roulette using the following Python code.

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

kmax= 300000

np.random.seed(seed=12345)

a= np.random.choice(np.arange(0,37) ,kmax)
red = [1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36]
drift = -np.arange(1,kmax+1)/37
zeros=np.zeros(kmax)
```

```
mean=np.zeros(kmax); mean.fill(-1/37)

# Profit/Loss in Red Strategy + Quantiles
profit1 = np.where(np.in1d(a, red),1,-1)
total1 = np.cumsum(profit1)
average1 = np.cumsum(profit1)/np.arange(1,kmax+1)
sigma_sq1 = (1 - (-1/37))**2*18/37 + (-1 - (-1/37))**2*19/37
upper_quantile1=drift+sp.norm.ppf(0.975)*np.sqrt(sigma_sq1)*np.sqrt(np.arange(1,kmax+1))
lower_quantile1=drift+sp.norm.ppf(0.025)*np.sqrt(sigma_sq1)*np.sqrt(np.arange(1,kmax+1))

# Quantiles for Average
upper_q1=-1/37+sp.norm.ppf(0.975)*np.sqrt(sigma_sq1)/np.sqrt(np.arange(1,kmax+1))
lower_q1=-1/37+sp.norm.ppf(0.025)*np.sqrt(sigma_sq1)/np.sqrt(np.arange(1,kmax+1))


# Profit/Loss in 13 Strategy + Quantiles
profit2 = np.where(a == 13,35,-1)
total2 = np.cumsum(profit2)
average2 = np.cumsum(profit2)/np.arange(1,kmax+1)
sigma_sq2 = (35 - (-1/37))**2*1/37 + (-1 - (-1/37))**2*36/37
upper_quantile2=drift+sp.norm.ppf(0.975)*np.sqrt(sigma_sq2)*np.sqrt(np.arange(1,kmax+1))
lower_quantile2=drift+sp.norm.ppf(0.025)*np.sqrt(sigma_sq2)*np.sqrt(np.arange(1,kmax+1))

# Quantiles for Average
upper_q2=-1/37+sp.norm.ppf(0.975)*np.sqrt(sigma_sq2)/np.sqrt(np.arange(1,kmax+1))
lower_q2=-1/37+sp.norm.ppf(0.025)*np.sqrt(sigma_sq2)/np.sqrt(np.arange(1,kmax+1))

plt.figure(1)
plt.plot(total1,'b')
plt.plot(total2,'b')
plt.plot(drift,'r')
plt.plot(upper_quantile1,'m')
plt.plot(lower_quantile1,'m')
plt.plot(upper_quantile2,'m')
plt.plot(lower_quantile2,'m')
plt.plot(zeros,'g')
# unmark the following commands to obtain zoomed graph
#axes = plt.gca()
#axes.set_xlim([0,5000])
#axes.set_ylim([-1000,1000])



plt.figure(2)
plt.plot(average1,'b')
plt.plot(average2,'b')
plt.plot(mean,'r')
plt.plot(upper_q1,'m')
plt.plot(lower_q1,'m')
plt.plot(upper_q2,'m')
plt.plot(lower_q2,'m')
plt.plot(zeros,'g')
axes = plt.gca()
axes.set_ylim([-0.15,0.15])
```

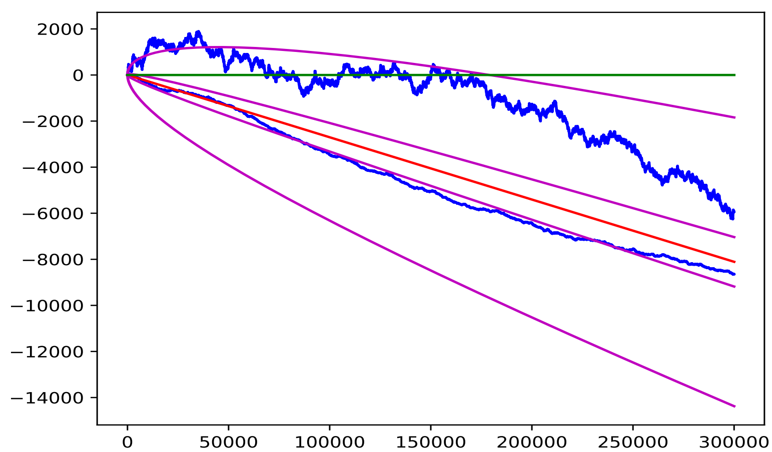

FIGURE 1.15: A simulated wealth evolution of the two trading strategies, one that bets on red and one that bets on 13 in the roulette together with the expected trend and the confidence intervals.

The result of the code is the following. Figure 1.15 shows a simulated evolution of the player's wealth $S_1(n)$ and $S_2(n)$ together with expectation $-\frac{n}{37}$ and the corresponding 95% two sided confidence intervals. We chose again the random seed 12345 for reproducibility of the code. We see some exits of the wealth processes from these confidence intervals, but such discrepancies are not substantial and can happen from time to time. As seen from the graph, the single number betting strategy is much more volatile and this volatility is indeed the main driving force of the profit loss process that can beat the negative drift, even on a longer term horizon. On the other hand, the betting strategy on red has only a small volatility and as such it tracks the drift more closely. However, this is undesirable given that the drift is negative.

As a part of our analysis of the profit loss performace, we can extract some important statistics. Namely, the historical maximum, when the maximum happened and the last point when the profit loss was at zero for the two strategies. The corresponding Python commands are

```
np.max(total1)
np.argmax(total1)
np.argwhere(total1==0)[-1][0]

np.max(total2)
np.argmax(total2)
```

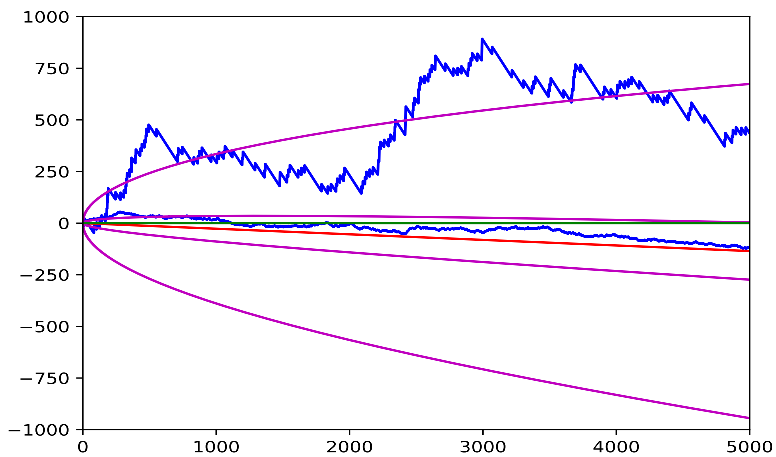

FIGURE 1.16: Same graph as Figure 1.15, but zoomed in to see the evolution on a smaller time horizon. The single number strategy seems to overperform the typical range of the values, but obviously this discrepancy lasts only for a finite time before getting back to the 95% confidence interval.

```
np.argwhere(total2==0)[-1][0]
```

The command np.max extracts the maximum value. The command np.argmax extracts the index when the maximum happened. The command np.argwhere gives the list of all points where the profit loss was exactly at 0. As we are interested only in the last value, we take the last value from the array by calling [-1]. As the resulting object is still an array (of one element) but not the value, we get the value by calling additional index [0]. This gives maximum value of the profit loss for the red strategy as 55. It happened at time 276. The last time when the profit loss was at zero as 2013. This is better seen from Figure 1.16. For the single number strategy, we get the maximum level 1905, it happened at time 35138 and the last time when the profit loss was at zero is 170603. These values are relatively optimistic as seen from the corresponding plots as the wealth evolutions of both wealth processes are on the top of the confidence intervals (at least in a short run for the betting strategy on red), but these values are still fully possible as seen from the following analysis.

A natural question is how large n is needed so that the confidence interval does not include zero anymore. The confidence interval for the normal distribution $N(n \cdot \mu, \sigma\sqrt{n})$ is given by

$$(n \cdot \mu + \sqrt{n} \cdot q_{\frac{\alpha}{2}} \cdot \sigma,\ n \cdot \mu + \sqrt{n} \cdot q_{1-\frac{\alpha}{2}} \cdot \sigma).$$

In the case when μ is negative, we want that the right end point

$$n \cdot \mu + \sqrt{n} \cdot q_{1-\frac{\alpha}{2}} \cdot \sigma$$

is also negative, or in the boundary case it should be equal to zero. Thus we have

$$n \cdot \mu + \sqrt{n} \cdot q_{1-\frac{\alpha}{2}} \sigma = 0,$$

or

$$n = \left(\frac{q_{1-\frac{\alpha}{2}} \sigma}{\mu} \right)^2 .$$

In our case when we considered $q_{0.975} = 1.96$ (only 2.5% probability of exceeding the right end point of the confidence interval), we get that $n = 5,255$ when betting on red and $n = 179,234$ when betting on 13. In particular, when betting on 13, it takes almost 180 thousand bets and one still has 2.5% probability of being ahead of the casino. I find this result rather amazing as this make it very difficult to detect that someone is playing in a casino in contrast to making a serious investment.

Figure 1.17 is an illustration of the Law of Large Numbers where we study the behavior of the average wealth $\bar{X}_1(n) = \frac{S(n)}{n}$. The standard deviation of the average is $\frac{\sigma}{\sqrt{n}}$, so the confidence interval is given by

$$\left(-\tfrac{n}{37} - 1.96 \cdot \sqrt{\tfrac{1368}{1369 \cdot n}}, -\tfrac{n}{37} + 1.96 \cdot \sqrt{\tfrac{1368}{1369 \cdot n}} \right)$$

This interval is shrinking with increasing n, and so the average win should converge to the expected win as seen in the figure. The same figure shows the evolution of the average win when betting on 13. The confidence interval is larger due to the larger σ, but it is also shrinking with increasing n:

$$\left(-\tfrac{n}{37} - 1.96 \cdot \sqrt{\tfrac{46656}{1369 \cdot n}}, -\tfrac{n}{37} + 1.96 \cdot \sqrt{\tfrac{46656}{1369 \cdot n}} \right).$$

Example 1.25 Monte Carlo Simulation for Roulette Statistics:
The previous example considered only one possible scenario of the profit loss evolution. One scenario gives only one value for some statistics of interest, such as the maximum reached or the last time the profit loss was at zero. This does not give us enough information about the full distribution of these statistics. In order to have a better idea about these values, we need to simulate many possible scenarios.

Let us consider three statistics of interest:

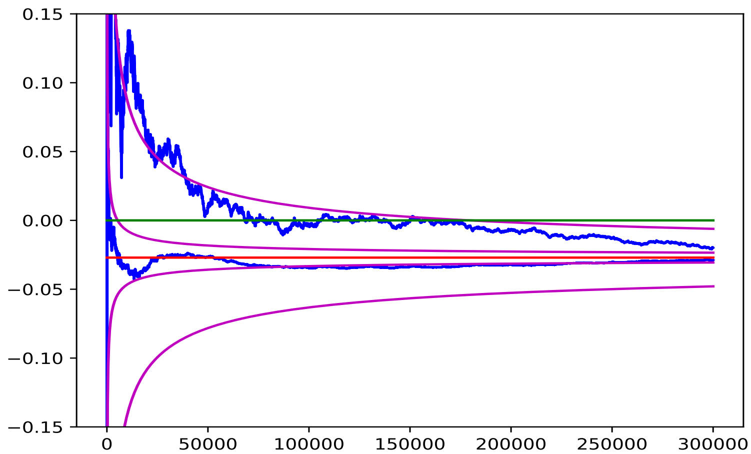

FIGURE 1.17: The averages corresponding to the trading strategies from Figure 1.15 with the corresponding expectation and the confidence intervals. Law of Large Numbers states that the averages will converge to the expectation $-\frac{1}{37}$.

1. the maximum reached,

2. number of times when profit loss is equal to zero,

3. time of the last visit of zero.

In the situation of betting on red, the probability distributions of these values can be obtained analytically, so we can confirm that our simulations fit them. Regarding the maximum reached denoted by M, it can be computed from Markov chain theory that

$$\mathbb{P}(M = 0) = \frac{1}{19},$$

which means that $\frac{1}{19}$ fraction of all possible scenarios results in a profit loss evolution that never becomes positive. The player may revisit zero at later times, but the wealth will never move to positive values. Interestingly enough, since $\frac{1}{19} > 5\%$, it means that this possibility is statistically significant. We can obtain probability that $\mathbb{P}(M = 1)$ with the following argument. First, $\mathbb{P}(M \geq 1) = 1 - \mathbb{P}(M = 0) = \frac{18}{19}$. In order that the maximum is exactly one, the maximum cannot increase after reaching level one. This happens again with probability $\frac{1}{19}$, which represents the situation when the maximum is not increased at later times. In conclusion,

$$\mathbb{P}(M=1)=\mathbb{P}(M\geq 1)\cdot\mathbb{P}(M=0)=\frac{18}{19}\cdot\frac{1}{19}.$$

Repeating this argument, we can see that the next higher level is reached with probability $\frac{18}{19}$ and the next higher level is not reached with probability $\frac{1}{19}$. Level k requires k steps up, followed by evolution that does not increase the maximum. The probability is

$$\mathbb{P}(M=k)=\frac{1}{19}\cdot\left(\frac{18}{19}\right)^k.$$

This is geometric distribution with one exception that the values start at $k=0$, while the standard geometric distribution starts at 1. This is not a big problem, the only thing is that all statistics of interest are shifted by a constant 1. For instance, $\mathbb{E}[M]=18$, the median is 12, and the 95% quantile is 55.

In comparison, probability of never reaching a positive value in a single number strategy is $\frac{1}{36}$ and the expected value of the maximum is $630=18\cdot 35$. This is rather interesting that a strategy of betting one dollar on a single number in roulette does produce a maximal wealth of 630 in expectation. The author does not know an analytical proof of these values, only very precise numerical computations suggest that this is indeed the case.

Getting back to the red strategy, we can also derive the number of visits to zero denoted by V. Let us compute the probability that the profit loss will never get back to zero from the zero value. Two things must happen: the wealth process must go down in the next step, which happens with probability $\frac{19}{37}$ and once there, the process will not go higher, meaning that the maximum will not be reset to a higher value. This happens with probability $\frac{1}{19}$. Thus the probability of the last visit is equal to

$$\frac{19}{37}\cdot\frac{1}{19}=\frac{1}{37}.$$

This means that the process will make an additional visit of zero with probability $\frac{36}{37}$ and will go down for good with probability $\frac{1}{37}$. Thus the number of visits to zero is also geometric with parameter $p=\frac{1}{37}$, so

$$\mathbb{P}(V=k)=\frac{1}{37}\cdot\left(\frac{36}{37}\right)^{k-1}.$$

The wealth process starts at zero, which we count as a visit, so V has the minimal value 1.

As for the time T of the last visit to zero. In order that T is the last visit, the wealth process has to be exactly zero and the process will go down after that. A visit to zero can happen only at even number of times n, when half of the outcomes $\frac{n}{2}$ are wins and the other half are losses. This has binomial probability $Bin(n, p)$ for $p = \frac{18}{37}$. The process will go down with probability $\frac{1}{37}$. Thus

$$\mathbb{P}(T = n) = \frac{1}{37}\binom{n}{\frac{n}{2}}\left(\frac{18}{37}\right)^{\left(\frac{n}{2}\right)}\left(\frac{19}{37}\right)^{\left(\frac{n}{2}\right)}.$$

Now we can run the simulations and check the fit to the theoretical distributions. The idea is rather straightforward given that we have already simulated one scenario. We will simply generalize this idea and simulate a number of scenarios in a 'for' loop. We just need to initialize and keep the statistics of interest, namely 'max_reached', 'visits_to_zero' and 'last_visit'. The simulation of the bet outcome is done from a transformation of the uniform random variable as described in the section of introduction to Python.

```
import numpy as np
import scipy.stats as sp
import scipy.special
import matplotlib.pyplot as plt

def pl(x):
    if x<= 18/37:
        a = 1
    else:
        a= -1
    return a

pl = np.vectorize(pl)

np.random.seed(seed=12345)
simulations=100000
runs = 15000

# initialize the empty arrays
max_reached = np.empty(simulations, dtype=int)
visits_to_zero = np.empty(simulations, dtype=int)
last_visit = np.empty(simulations, dtype=int)


for i in range(simulations):
    random_vector = sp.uniform.rvs(size = runs)
    roulette_pl = pl(random_vector)
    profit = np.cumsum(roulette_pl)
    profit = np.append(0,profit)
    max_reached[i]=np.max(profit)
```

```
    zero_times = np.argwhere(profit==0)
    visits_to_zero[i] = len(zero_times)
    last_visit[i] = np.max(zero_times)
    if i%1000==0: print(i)
```

The statistics related to the maximum reached is analyzed by the following code.

```
x=np.arange(-0.5,80.5,1)
plt.figure(1)
plt.hist(max_reached, normed = True, bins=x)
x=np.arange(0,80)
plt.plot(sp.geom.pmf(x,1/19,loc = -1),'r',ms=3)


print(np.mean(max_reached))
print(np.median(max_reached))
print(np.percentile(max_reached,95))
print(np.max(max_reached))
```

The average value is 17.99131, empirical median is 12, empirical 95% quantile is 55, and the highest maximum reached is 209. This is fully in line with the theoretical distribution. The comparison of the histogram and the theoretical distribution can be seen in Figure 1.18.

Statistics related to the visits of zero are obtained from:

```
plt.figure(2)
x=np.arange(0.5,210.5,1)
plt.hist(visits_to_zero,normed=True, bins=x)
x=np.arange(1,210)
plt.plot(sp.geom.pmf(x,1/37),'r',ms=3)

print(np.mean(visits_to_zero))
print(np.median(visits_to_zero))
print(np.percentile(visits_to_zero,95))
print(np.max(visits_to_zero))
```

We get the average 36.9589 (expectation is 37), median 26 (fits the median of the geometric distribution), 95% quantile 109 (theoretical quantile is 110), and the maximal number of visits in these simulations was 470. The fit to the distribution is seen in Figure 1.19.

The final stastistics of interest, the time of the last visit, is obtained from

```
def last(x):
    return (1/37)*scipy.special.binom(x,x/2)*(18/37)**(x/2)*(19/37)**(x/2)
```

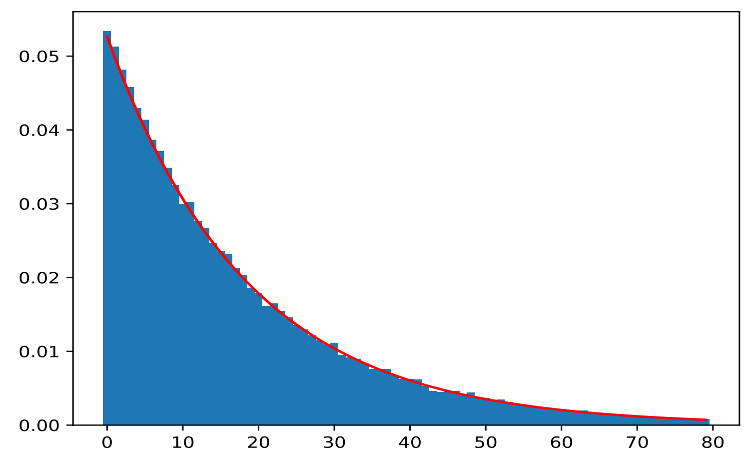

FIGURE 1.18: Comparison of the histogram of the maximum reached with the theoretical geometric distribution.

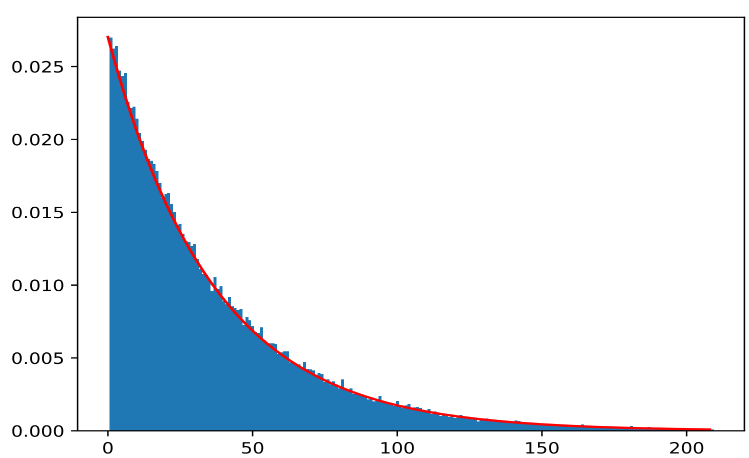

FIGURE 1.19: Comparison of the histogram of the visits to zero with the theoretical geometric distribution.

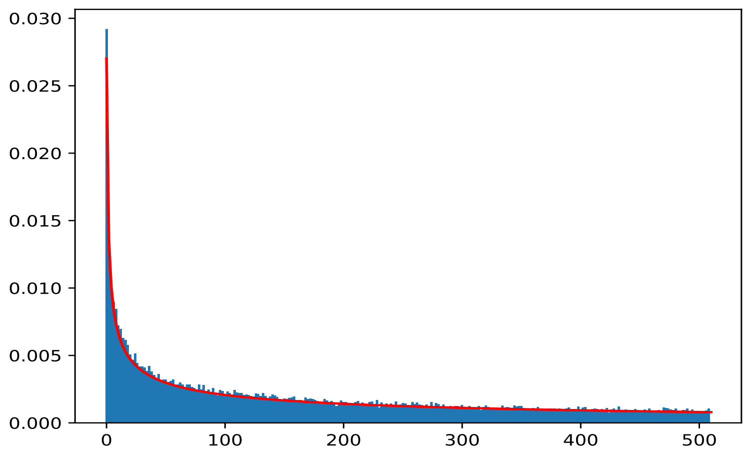

FIGURE 1.20: Comparison of the histogram of the last visit with the theoretical distribution.

```
plt.figure(3)
x=np.arange(-1,511,2)
plt.hist(last_visit,normed = True, bins=x)
x=np.arange(0,511,2)
plt.plot(x,last(x),'r')

print(np.mean(last_visit))
print(np.median(last_visit))
print(np.percentile(last_visit,95))
print(np.max(last_visit))
```

We get the average of 1367.9101 (expectation is 1368), median is 618 (theoretical is 622), 95% quantile is 5296 (theoretical is 5256) and the maximum obtained is 15000, which is at the very end of the simulation range. This means that some scenarios reached positive values even beyond our horizon, but the fraction of such scenarios is indeed very small. The fit of the simuated values to the theoretical distribution of the last exit time is in Figure 1.20.

$\square$

1.13 Monte Carlo Simulation for Asset Prices

Consider a situation when we need to simulate an evolution of asset prices. We motivate such approach with a pricing example at the end of the section. Shortly speaking, the asset prices are typically continuous, although they may exhibit jumps. However, the direction of the price moves must be unpredictable, otherwise one could construct a profitable trading strategy. The finance theory assumes that the market does not allow for a sure profit (no arbitrage condition), which in turn means that the price moves must be unpredictable (more specifically, the price processes must be so called martingales – processes that keep the expected value). Mathematically, the price process should have no derivative, which is in the situation of continuity of the price a rather interesting restriction as the mathematics taught at calculus studies only differentiable function. Still, the resulting price processes, are now well understood with stochastic calculus. The processes that are continuous and have no derivative are known as diffisions.

The non-differentiability of the price process can be a starting point of the simulation of the asset prices. Even when the derivative of the price process does not exist in the infinitesimal sense, we can still get a price differential in the discrete sense. Such price differential is called as return in finance and we formally define it as a log return:

$$\log\left(\frac{S_{t+1}}{S_t}\right) = \log(S_{t+1}) - \log(S_t),$$

which is technically a discrete derivative of the log price. As the return is according to the previous discussion unpredictable, we can simulate it via some symmetric random variable, such as normal distribution.

Let us write a code that generates these random variables and combines them in a stock price with exponentiation of the returns. We need to choose some parameters, such as sigma that represents the volatility (annualized standard deviation of the returns), time horizon T, number of sampling points and the actual time grid with a specific time step dt.

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

sigma =0.3
T = 3
sampling_points = 250*T
dt = T/sampling_points
```

```
time = np.arange(0,T + dt,dt)

np.random.seed(seed=12345)

noise = sp.norm.rvs(size = sampling_points)*np.sqrt(dt)
brownian_motion = np.cumsum(noise)
brownian_motion = np.append(0,brownian_motion)

stock =np.empty(sampling_points+1, dtype=float)
stock[0] = 100
stock = stock[0]*np.exp(sigma*brownian_motion - 0.5*sigma**2*time)
```

The code then generates properly scaled normal random variables. This collection of random variables is called a white noise (you may have seen a white noise on untuned color television). The noise represents the model of price returns (unscaled by sigma). An integrated white noise is called a Brownian motion, which is fundamental example of the diffusion process. The procedure finishes with exponentiation of the Brownian motion, which simulates a stock price. To visualize these processes, run

```
plt.figure(1)
plt.plot(time[1:],noise)

plt.figure(2)
plt.plot(time,brownian_motion)

plt.figure(3)
plt.plot(time,stock)
```

which gives Figure 1.21 that represents the noise, Figure 1.22 that shows the corresponding Brownian motion and Figure 1.23 that shows a simulation of the stock price.

We can repeat this procedure many times to get a representative distribution of the stock price evolutions. All that is needed is to run a 'for' loop. For instance, if we run

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

sigma =0.3
T = 3
sampling_points = 250*T
dt = T/sampling_points
time = np.arange(0,T + dt,dt)
```

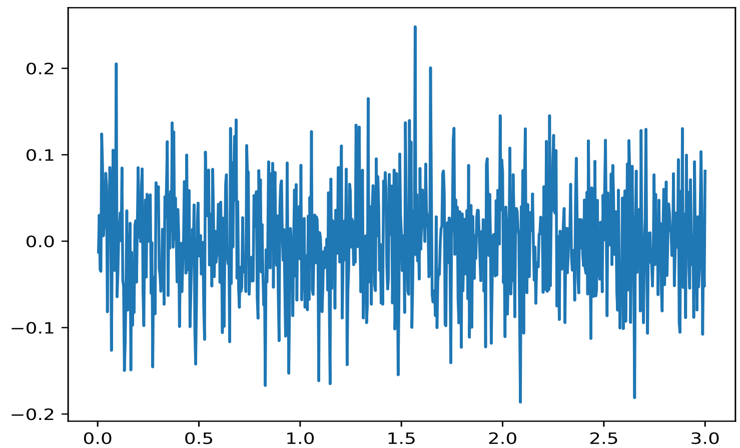

FIGURE 1.21: Simulated noise.

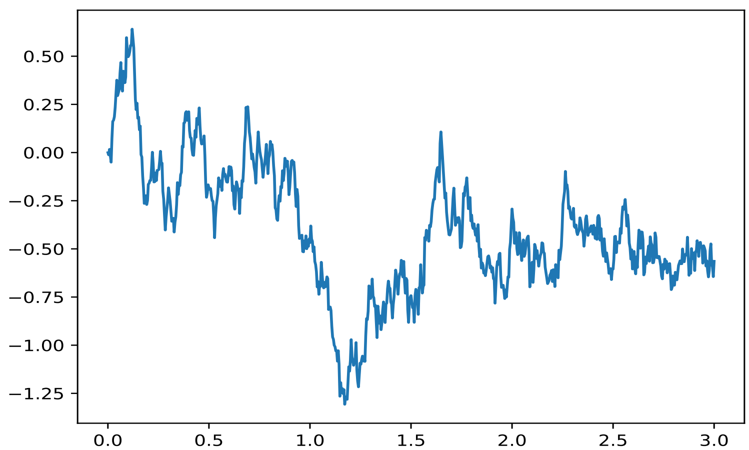

FIGURE 1.22: Corresponding Brownian motion. Note that this process has no derivative.

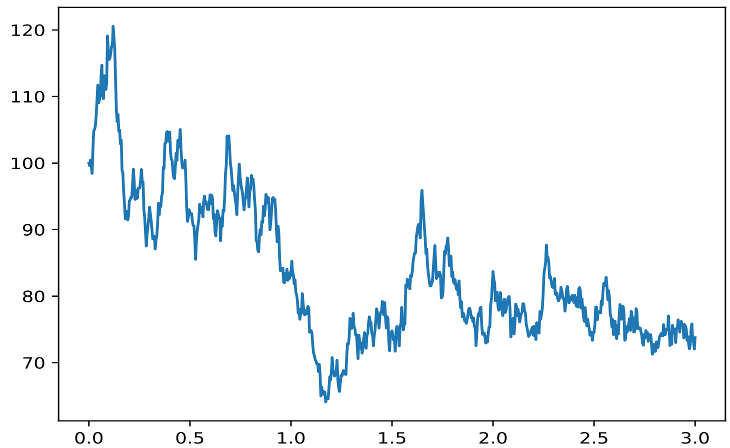

FIGURE 1.23: Simulated stock price.

```
np.random.seed(seed=12345)

for i in range(1,9):
    noise = sp.norm.rvs(size = sampling_points)*np.sqrt(dt)
    brownian_motion = np.cumsum(noise)
    brownian_motion = np.append(0,brownian_motion)

    stock =np.empty(sampling_points+1, dtype=float)
    stock[0] = 100
    stock = stock[0]*np.exp(sigma*brownian_motion - 0.5*sigma**2*time)

    plt.plot(time,stock)
```

we produce 8 different stock price evolutions shown in Figure 1.24.

Given that we have now the capacity to run asset price simulations, we can price some complex financial contract. Consider the following example.

Example 1.26 Pricing a structured product:
Recently, banks offer the following product. There is an underlying asset S which is typically a specific stock. The contract has maturity $T = 3$ years and 6 monitoring semianual points. At each monitoring point, the client receives a 2.5% coupon subject if these conditions are met:

1. The contract is terminated at the first monitoring point when the stock

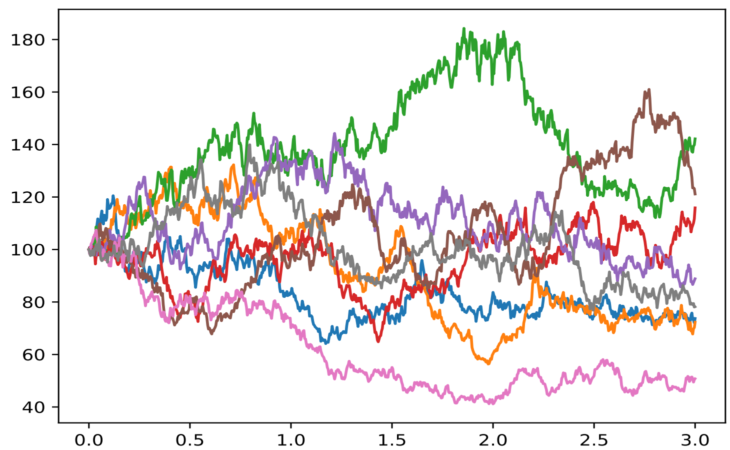

FIGURE 1.24: Eight simulated evolutions of the stock price.

price exceeds the initial value (100%). The coupon 2.5% is paid.

2. The client receives the 2.5% coupon for each monitoring point when the stock exceeds 80% of its initial value. The coupon is also received ex post if the stock reaches this region at a later monitoring time.

3. The client receives 100% of the face value if the terminal stock price exceeds 60% of its initial value. If the stock drops below 60%, the client gets the corresponding percentage.

To compute the value of this contract, we can simulate the stock prices and compute the payoff for each scenario. This ingredient was already covered above. The price of such a contract should be the expected value over all possible payoffs, but since we have no analytical distribution of the payoffs, we need to resort to simulations, where the average value will serve as an estimate of the expectation. We just need to code the contractual condition that determines the payoff, which is just a nested 'if' statement which sorts out the early termination of the contract from the rest of the scenarios and the case when the stock is above or below 60% at the maturity.

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

sigma =0.3
T = 3
```

```
sampling_points = 2*T
dt = T/sampling_points
time = np.arange(0,T + dt,dt)

simulations = 100000
payoff =np.empty(simulations, dtype=float)

np.random.seed(seed=12345)

for i in range(simulations):
    stock =np.empty(sampling_points+1, dtype=float)
    stock[0] = 100
    noise = sp.norm.rvs(size = sampling_points)*np.sqrt(dt)
    brownian_motion = np.cumsum(noise)
    brownian_motion = np.append(0,brownian_motion)
    stock = stock[0]*np.exp(sigma*brownian_motion - 0.5*sigma**2*time)

    if np.argmax(stock>100) != 0:
            payoff[i] = 100+2.5*np.argmax(stock>100)
    else:
        if stock[-1] < 60:
                    payoff[i] = stock[-1]+2.5*np.max(np.where(stock>80))
        else:
                    payoff[i] = 100+ 2.5*np.max(np.where(stock>80))
```

We can extract the average and the 1.96 times the standard deviation using

```
np.mean(payoff)
1.96*np.std(payoff)/np.sqrt(simulations)
```

which gives $95.901655226805971 \pm 0.13585753764556999$ as an estimate of the fair price of this contract. However, if we change the volatility from 0.3 to 0.2, the price changes to $101.8687531588042 \pm 0.086522402146608293$, suggesting that the contract has a negative dependence on the volatility. Indeed, if the stock is less volatile, it is less likely that the contract will be terminated early, plus it is more likely that the coupon will be collected and the stock will not drop below 60% at the maturity of the contract. In the extreme case of zero volatility, the stock will stay constant and the client will collect all 6 coupons, which will result in a value of 115.

We can also print the histogram of the payoff distribution that appears in Figure 1.25. We have plotted the payoff distribution for both choices of sigma = 0.2 and sigma = 0.3. Notice that the smaller volatility results in higher payoffs. The corresponding code is

```
x=np.arange(25,117,2.5/2)
plt.hist(payoff, alpha=0.5,normed=True, bins=x)
```

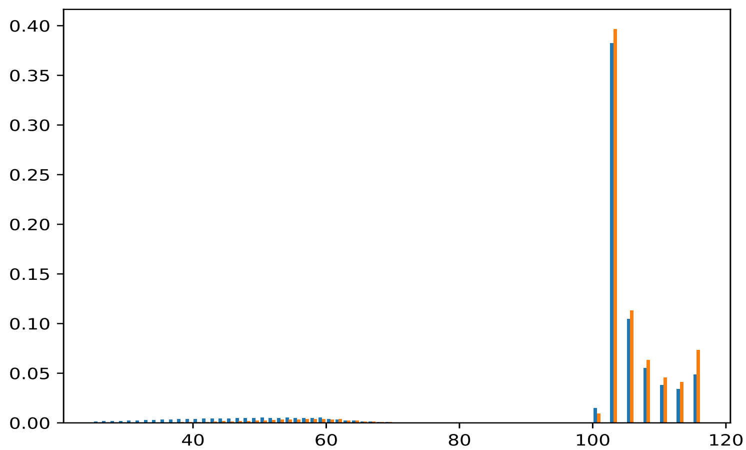

FIGURE 1.25: Histogram of the payoffs for two different choices of sigma. Blue corresponds to sigma = 0.3, yellow corresponds to sigma = 0.2.

□

Chapter 2

Statistics

We have studied idealized models in the chapter on probability. Probability assumes that we know from which distribution we generate random variables. It is also assumed that the parameters of these distributions are known. Using this knowledge, we can determine various characteristics of the generated random variables such as the mean, the variance, the confidence intervals, etc. For instance, a die roll assumes that the probability of each number is equal to $\frac{1}{6}$. But how do we know that a given die indeed produces equally likely outcomes? This is a question for statistics. Statistics assumes a certain probabilistic model for the data, possibly without specifying the numerical parameters (such as the mean and the variance of the normal distribution), observes the data, and finds the set of parameters that explain the observed data in some optimal way. Alternatively, one may confirm or reject that the assumed probability model is supported by the observed data, such as in the case of checking for a fairness of the die roll.

2.1 Point Estimation of the Mean and the Variance

Let $X(1), X(2), \ldots, X(n)$ be n independent observations that come from the same distribution. Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$ be the mean and the variance of this distribution. What is our best guess about the values of μ and σ^2 if we do not know them? Our guess must be based on the observed data, so we need to consider some function f of the data

$$f(X(1), X(2), \ldots, X(n)).$$

The function of data f is called an **estimator.** Estimator is a rule for calculating an estimate of a given quantity based on observed data: thus the rule and its result, the **estimate,** are distinguished. Point estimation means that the data give a single value for the estimated parameter as opposed to a possible range of values which corresponds to interval estimation.

Example 2.1
Estimator

$$f(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^n x_i}{n}$$

is called the **average.** Different estimators are possible, for instance the first observation

$$f(x_1, x_2, \ldots, x_n) = x_1$$

or the largest observation

$$f(x_1, x_2, \ldots, x_n) = \max(x_1, x_2, \ldots, x_n)$$

are legitimate estimators. $\square$

Estimates should have some desirable properties. One property is **unbiasedness.** An estimator of a parameter θ is called unbiased if

$$\mathbb{E}[f(X(1), X(2), \ldots, X(n))] = \theta.$$

Example 2.2 Estimators of the mean:
Consider an estimator in the form

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} a(i) \cdot x_i,$$

which is a weighted linear combination of the observations $X(i)$. The estimators with

$$\sum_{i=1}^{n} a(i) = 1,$$

represent unbiased estimators of the mean μ since

$$\mathbb{E}[f(X(1), X(2), \ldots, X(n))] = \mathbb{E}\left[\sum_{i=1}^{n} a(i) \cdot X(i)\right] = \sum_{i=1}^{n} a(i) \cdot \mu = \mu.$$

For instance, the average (also denoted by $\bar{X}(n)$) is an unbiased estimator of the mean μ ($a(1) = \cdots = a(n) = \frac{1}{n}$), but also the first observation ($a(1) = 1, a(2) = \cdots = a(n) = 0$) is another unbiased estimator of the mean. Or any other individual observation.

However, using a single observation for estimating the mean (or any other parameter) is not optimal as it is wasteful on data. Comparing the variances of these estimators, one can see that

$$\mathrm{Var}(X(1)) = \sigma^2,$$

but

$$\operatorname{Var}(\bar{X}(n)) = \frac{\sigma^2}{n}.$$

Having a smaller variance is desirable since the estimate would be closer to the true value.

The unbiased estimators in the form $f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^n a(i) \cdot x_i$, $\sum_{i=1}^n a(i) = 1$ have variance

$$\operatorname{Var}\left(\sum_{i=1}^n a(i) \cdot X(i)\right) = \sigma^2 \cdot \sum_{i=1}^n a^2(i).$$

This quantity is minimized when $a(1) = \cdots = a(n) = \frac{1}{n}$, so the average has the smallest variance among the unbiased estimators. Such an estimator is called the **minimum variance unbiased estimator.** $\square$

Let us find unbiased estimators of the variance σ^2. When μ is known, we can use the fact that

$$\sigma^2 = \operatorname{Var}(X) = \mathbb{E}[(X-\mu)^2]$$

to conclude $\frac{1}{n}\sum_{i=1}^n (X(i)-\mu)^2$ is an unbiased estimator of σ^2:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (X(i)-\mu)^2\right] = \frac{1}{n} \cdot n \cdot \sigma^2 = \sigma^2.$$

When μ is not known, a natural idea is to replace it with its estimator $\bar{X}(n)$. However, $\frac{1}{n}\sum_{i=1}^n (X(i)-\bar{X}(n))^2$ is not an unbiased estimator as we can see from the following computation

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n (X(i)-\bar{X}(n))^2\right] &= \mathbb{E}\left[\sum_{i=1}^n ((X(i)-\mu)-(\bar{X}(n)-\mu))^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n (X(i)-\mu)^2\right] + \mathbb{E}\left[\sum_{i=1}^n (\bar{X}(n)-\mu)^2\right] \\
&\quad -2 \cdot \mathbb{E}\left[\sum_{i=1}^n (X(i)-\mu) \cdot (\bar{X}(n)-\mu)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n (X(i)-\mu)^2\right] - n \cdot \mathbb{E}\left[(\bar{X}(n)-\mu)^2\right] \\
&= n \cdot \sigma^2 - n \cdot \frac{\sigma^2}{n} = (n-1) \cdot \sigma^2.
\end{aligned}$$

We can see that the unbiased estimator of σ^2 is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X(i)-\bar{X}(n))^2.$$

One can think about it that in order to estimate μ, we must invest one data point (in the form of their linear combination, the average), while the remaining $n-1$ points can be used for estimation of σ^2.

2.2 Maximum Likelihood Estimation

Finding an unbiased estimator with the minimum variance in a general situation may not be easy and straightforward. The method of the maximum likelihood described in this section is more straightforward and it gives reasonable estimates, although the resulting estimator is not always unbiased. Independent and identically distributed observations $X(1), X(2), \ldots, X(n)$ given the parameter θ have the joint density function

$$\begin{aligned} f(X(1), X(2), \ldots, X(n)|\theta) &= f(X(1)|\theta) \cdot f(X(2)|\theta) \ldots f(X(n)|\theta) \\ &= \prod_{i=1}^{n} f(X(i)|\theta). \end{aligned}$$

However, the observations $X(1), X(2), \ldots, X(n)$ are already known, so the only unknown variable is the parameter θ. The **maximum likelihood estimator** chooses the parameter θ for which the quantity

$$L(\theta|X(1), X(2), \ldots, X(n)) = \prod_{i=1}^{n} f(X(i)|\theta)$$

is maximized. It means that one finds parameter(s) θ such that the observations have the largest density (or probability in the discrete distribution case) among the possible choices of all parameters. The maximum likelihood estimator explains the observed data as more likely when compared to any other choice of the parameters. In practice it is more convenient to work with the $\ln(L)$ as the maximizer is the same as for L. Therefore the maximum likelihood estimator is given by

$$\theta_{MLE} = \arg\max_{\theta} \left(\sum_{i=1}^{n} \ln(f(X(i)|\theta)) \right)$$

Example 2.3 Maximum likelihood estimator for normal distribution:
Assume that the observations $X(1), X(2), \ldots, X(n)$ come from normal distribution $N(\mu, \sigma^2)$. The goal is to find the maximum likelihood estimators of μ and σ^2. The density of $N(\mu, \sigma^2)$ random variable is equal to

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

and thus the log likelihood is given by

$$\ln(L(\mu, \sigma^2 | X(1), X(2), \ldots, X(n))) = \sum_{i=1}^{n} \left(-\tfrac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{(X(i)-\mu)^2}{2\sigma^2} \right).$$

To find the maximum, take the partial derivatives with respect to μ and σ and set them to zero. We find

$$0 = \frac{\partial \ln(L)}{\partial \mu} = \sum_{i=1}^{n} \left(\frac{X(i)-\mu}{\sigma^2} \right)$$

that leads to a maximum likelihood estimator of μ

$$\hat{\mu} = \bar{X}(n),$$

and

$$0 = \frac{\partial \ln(L)}{\partial \sigma} = \sum_{i=1}^{n} \left(-\frac{1}{\sigma} + \frac{(X(i)-\mu)^2}{\sigma^3} \right)$$

that leads to a maximum likelihood estimator of σ^2

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X(i)-\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(X(i)-\bar{X}(n))^2.$$

Note that the maximum likelihood estimator of σ^2 is biased, the unbiased estimator divides $\sum_{i=1}^{n}(X(i)-\bar{X}(n))^2$ by $\frac{1}{n-1}$, while the maximum likelihood estimator divides this sum by $\frac{1}{n}$. Both estimators are legitimate, they are optimal using two different criteria. $\square$

Example 2.4 Maximum likelihood for geometric distribution:
Let $X(1), X(2), \ldots, X(n)$ be n independent observations from geometric distribution

$$\mathbb{P}(X=k) = p \cdot (1-p)^{k-1}, \quad k = 1, 2, \ldots.$$

Let's estimate the parameter p. The likelihood function is given by

$$L(p | X(1), X(2), \ldots, X(n)) = \prod_{i=1}^{n} \left[p \cdot (1-p)^{X(i)-1} \right] = p^n \cdot (1-p)^{\sum_{i=1}^{n}(X(i)-1)}.$$

The log likelihood simplifies to

$$\ln L = n\ln(p) + \left[\sum_{i=1}^{n}(X(i)-1) \right] \cdot \ln(1-p).$$

Taking the derivative with respect to p, we get

$$\frac{\partial \ln L}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^{n}(X(i)-1)}{1-p}$$

Solving for $\frac{\partial \ln L}{\partial p} = 0$ leads to

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X(i)} = \frac{1}{\bar{X}(n)}.$$

$\square$

One major advantage of the maximum likelihood estimation is that it can be used to estimate the parameters in rather complex models where no simple formula exists for the estimators. Finding the maximum likelihood estimator is done by numerical procedures in such cases.

2.3 Interval Estimation

Since the data are random, different realizations would lead to different point estimates. In this sense, there is also some uncertainty about the values of the parameters. In order to address this uncertainty, one may give a confidence interval where the parameters can be in contrast to listing just a single value as in the point estimation.

Let us illustrate the concept of the interval estimation on the example of estimating μ when knowing σ from $N(\mu, \sigma)$ distribution. Assume that we have n observations $X(1), X(2), \ldots, X(n)$. The average $\bar{X}(n)$ is a point estimator of μ, but there is still some randomness as $\mathrm{Var}(\bar{X}(n)) = \frac{\sigma^2}{n}$ and the average itself has a normal distribution $N(\mu, \frac{\sigma}{\sqrt{n}})$. The $1-\alpha$ two sided confidence interval for $\bar{X}(n)$ is given by

$$\mathbb{P}\left(\mu + \tfrac{\sigma}{\sqrt{n}} \cdot q_{\frac{\alpha}{2}} \le \bar{X}(n) \le \mu + \tfrac{\sigma}{\sqrt{n}} \cdot q_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

This is assuming that we know μ and the average $\bar{X}(n)$ is yet to be observed. However, we can reverse this argument and assume that $\bar{X}(n)$ is already observed and μ is unknown. By rearranging the above inequalities, we get

$$\mathbb{P}\left(\bar{X}(n) - \tfrac{\sigma}{\sqrt{n}} \cdot q_{1-\frac{\alpha}{2}} \le \mu \le \bar{X}(n) - \tfrac{\sigma}{\sqrt{n}} \cdot q_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

For instance, 95% confidence interval for μ is given by

$$\left(\bar{X}(n) - 1.96 \cdot \tfrac{\sigma}{\sqrt{n}},\ \bar{X}(n) + 1.96 \cdot \tfrac{\sigma}{\sqrt{n}}\right).$$

Note that this uncertainty $(\pm 1.96 \cdot \frac{\sigma}{\sqrt{n}})$ is proportional to σ and inversely proportional to $\sqrt{n}$. Thus scaling down this interval by a factor of 2 requires

4 times more observations as $\sqrt{4} = 2$. Thus obtaining more precise estimation of the true value of μ is rather costly on the amount of data required. We have seen in the roulette case study that the 95% confidence interval for $\bar{X}(n)$ includes zero even for huge n. The last n for which the (one sided) confidence interval when betting on a single number includes zero is equal to

$$n = \left(\frac{q_{1-\alpha}\sigma}{\mu} \right)^2 = 125,486.$$

If you consider a player betting once on a single number each working day in a year (about 250), he may repeat this procedure for 500 years and still conclude that the roulette is fair in 5% scenarios as the confidence interval for μ would include zero!!! As the reader may imagine, identifying statistically positive or negative trends in financial data (meaning being able to conclude that $\mu > 0$ or $\mu < 0$) is extremely unreliable. It is a difficult task even in the situation when the "investment" is made in a roulette, a game with a known negative drift equal to $\mu = -\frac{1}{37}$.

It is not typical to know σ, so this value itself should be estimated. Before proceeding with further analysis, we need to introduce the following distributions. Let $X(1), X(2), \dots, X(n)$ be independent standard normal random variables $N(0,1)$. The distribution of

$$X(1)^2 + X(2)^2 + \cdots + X(n)^2$$

is known as χ^2_n-distribution (**chi-squared distribution with n degrees of freedom).** The density function of χ^2_n-distribution is given by

$$\frac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1}e^{-x/2}$$

where

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt.$$

Figure 2.1 shows this density function for $n = 1, 2, \dots, 9$. The corresponding Python code that generates this graph is

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

x=np.linspace(0,12,250)

for i in range(1,10):
    plt.plot(x,sp.chi2.pdf(x,i))
axes = plt.gca()
axes.set_ylim([0,0.6])
```

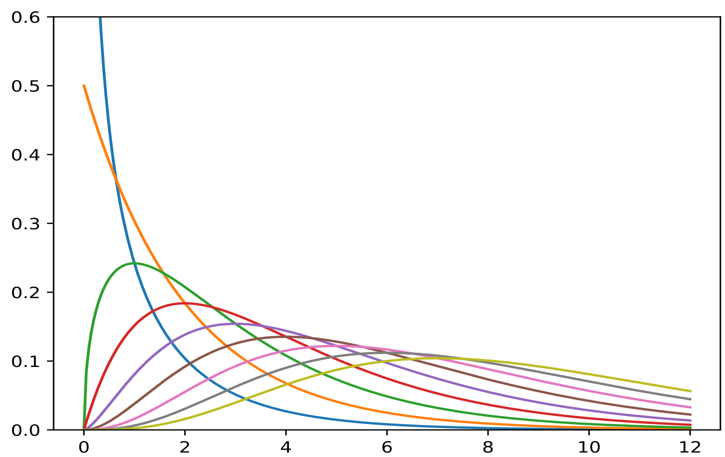

FIGURE 2.1: The density function of the χ^2_n-distribution for $n = 1, 2, \ldots, 9$.

Let $X(1), X(2), \ldots, X(n)$ be independent observations from $N(\mu, \sigma^2)$. Then the random variable

$$\sum_{i=1}^{n} \left(\frac{X(i) - \mu}{\sigma} \right)^2$$

has χ^2_n-distribution. This immediately follows from the fact that $\frac{X(i)-\mu}{\sigma}$ has a normal distribution $N(0, 1)$. This gives us $1 - \alpha$ confidence interval in the form

$$\mathbb{P}\left(\chi^2_{n,\frac{\alpha}{2}} \leq \sum_{i=1}^{n} \left(\frac{X(i) - \mu}{\sigma} \right)^2 \leq \chi^2_{n,1-\frac{\alpha}{2}} \right) = 1 - \alpha,$$

where $\chi^2_{n,\alpha}$ denotes the α quantile of the χ^2_n-distribution. This leads to an interval estimate for σ^2

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n} (X(i) - \mu)^2}{\chi^2_{n,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n} (X(i) - \mu)^2}{\chi^2_{n,\frac{\alpha}{2}}} \right) = 1 - \alpha.$$

When μ is not known, we can replace it with $\bar{X}(n)$. It can be shown that the random variable

$$\sum_{i=1}^{n} \left(\frac{X(i) - \bar{X}(n)}{\sigma} \right)^2$$

has χ^2_{n-1}-distribution, where the number of degrees of freedom is equal to $n - 1$. Following the same arguments as above, we get the $1 - \alpha$ confidence interval for σ^2 in the form

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n}\left(X(i)-\bar{X}(n)\right)^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\leq \sigma^2 \leq \frac{\sum_{i=1}^{n}\left(X(i)-\bar{X}(n)\right)^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)=1-\alpha.$$

In order to construct an interval estimate for μ when σ is not known, we need to introduce a **t-distribution** with n degrees of freedom that has a density

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Figure 2.2 shows the density of the t-distribution for the degrees of freedom $n=1,2,3,4$ together with the standard normal density. It can be shown that the density of the t-distribution converges to the standard normal density as $n->\infty$, so visually the densities quickly approach the shape of the normal distribution for higher n's. The smaller is the number of degrees of freedom, the "fatter" tail has the density function, and it peaks at a smaller level. The t-distribution is popular to use when modeling the financial returns as it better explains more extreme observations than the normal distribution. The corresponding Python code to generate the graph is

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt

x=np.linspace(-4,4,250)

for i in range(1,5):
    plt.plot(x,sp.t.pdf(x,i))

plt.plot(x,sp.norm.pdf(x))
```

One can show that the random variable

$$\frac{\bar{X}(n)-\mu}{\frac{S}{\sqrt{n}}}$$

has t-distribution with $n-1$ degrees of freedom, where

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X(i)-\bar{X}(n))^2.$$

Compare this result with the fact that the random variable

$$\frac{\bar{X}(n)-\mu}{\frac{\sigma}{\sqrt{n}}}$$

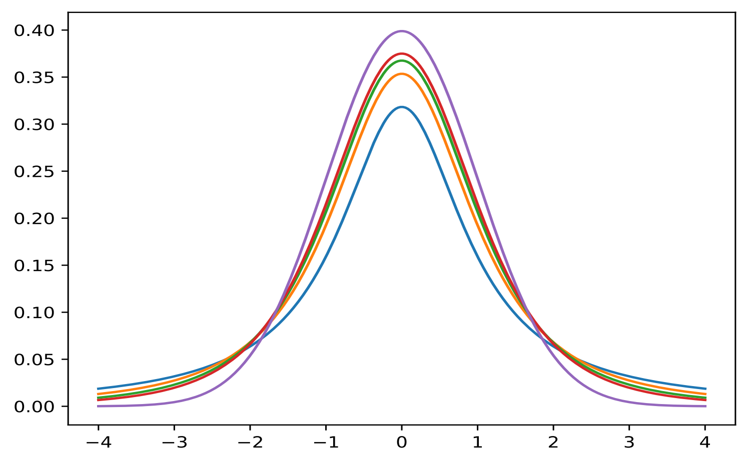

FIGURE 2.2: The density function of the t-distribution for $n = 1, 2, 3, 4$ together with the standard normal density. The smaller is the number of degrees of freedom, the fatter is the tail of the distribution.

that has $N(0,1)$ distribution. Thus we get $1-\alpha$ confidence interval in the form

$$\mathbb{P}\left(t_{n-1,\frac{\alpha}{2}} \leq \frac{\bar{X}(n)-\mu}{\frac{S}{\sqrt{n}}} \leq t_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha,$$

where $t_{n,\alpha}$ denotes the α quantile of the t-distribution with n degrees of freedom. The confidence interval for μ is thus given by

$$\mathbb{P}\left(\bar{X}(n) - \frac{S}{\sqrt{n}} \cdot t_{n-1,1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}(n) - \frac{S}{\sqrt{n}} \cdot t_{n-1,\frac{\alpha}{2}}\right) = 1-\alpha.$$

Since the t-distribution has fatter tails than the normal distribution, we have $t_{n,\alpha} < q_\alpha$ for $\alpha < \frac{1}{2}$.

2.4 Linear Regression

Suppose there are n data observations of pairs (x_i, y_i), where $i = 1, 2, \ldots, n$. The goal is to find the equation of the straight line

$$y = \alpha + \beta x,$$

which would provide a "best" fit for the data points. As the data are typically not forming a straight line, the relationship between data x and y is not expected to be exact, but rather with a small error ε:

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

The true values of α and β are unobserved parameters and may be only estimated with a certain precision from the observed data points. The standard approach called *ordinary least squares* estimation (or least-squares estimation) is to find a line $a + bx$ that has the smallest distance to the data points in the following sense: numbers a and b solve the following minimization problem:

$$\text{Find } \min_{a,\,b} L(a,b), \text{ where } L(a,b) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2.$$

The value

$$u_i = y_i - (a + bx_i),$$

the difference of the data point y_i and the line $a + bx_i$ is called **residual.** The function $L(a, b)$ is a quadratic function (paraboloid) and its minimum value can be found at the points where the partial derivative equals to zero. The partial derivative with respect to a gives

$$\frac{\partial L}{\partial a} = \sum_{i=1}^{n} (-2)(y_i - (a + bx_i)) = 0,$$

or in other words,

$$\sum_{i=1}^{n} u_i = 0.$$

Thus the optimal line satisfies a condition that the residuals sum up to zero. From the equality $\frac{\partial L}{\partial a} = 0$, we conclude that

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (a + bx_i) = a \cdot n + b \sum_{i=1}^{n} x_i.$$

This gives a condition for a after dividing by n:

$$a = \bar{y} - b\bar{x}.$$

The second partial derivative leads to

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} (-2)(y_i - (a + bx_i))x_i = 0,$$

or in other words,

$$\sum_{i=1}^{n} u_i x_i = 0.$$

This conditional essentially says that the residuals u_i and data x_i are independent, so whatever is left unexplained after the estimation (the residuals u_i) has no further knowledge about the data x_i. This condition can be further expressed as

$$\mathrm{Cov}(u, x) = \tfrac{1}{n}\sum_{i=1}^{n}(u_i - \bar{u})(x_i - \bar{x}) = 0,$$

meaning that the sample covariance of u and x is equal to zero. The condition $\frac{\partial L}{\partial b} = 0$ gives

$$\sum_{i=1}^{n}((y_i - \bar{y}) - b(x_i - \bar{x}))x_i = 0$$

after substituting for $a = \bar{y} - b\bar{x}$. Also note that $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$ and $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$, so we can freely multiply these terms and add or subtract them without changing the values. In particular, we can subtract $-\bar{x}$ multiple to get

$$\sum_{i=1}^{n}((y_i - \bar{y}) - b(x_i - \bar{x}))(x_i - \bar{x}) = 0.$$

This gives a condition for b:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\mathrm{Cov}[x, y]}{\mathrm{Var}[x]}.$$

The coefficients a and b are called an *intercept* and a *slope* respectively.

This procedure describes the best line that fits the data, but it has not answered the question about the confidence interval for the true unobserved parameters α and β. As both the intercept a and the slope b are functions of data, they are random (like the average taken from a random sample is random), and thus we need to find the exact distribution of these estimates in order to find the confidence interval for the estimated parameters. As the slope β is more important for determining the relationship of x and y, let us focus our attention on this parameter. From the formula

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

it may not be obvious where is the randomness as this if fully determined by the observed data. The randomness is in the relationship

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

Let us make an additional assumption that ε_i are independent identically distributed random variables with distribution $N(0, \sigma)$. After some simple algebra, we get

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\alpha + \beta x_i + \varepsilon_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} =$$
$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta(x_i - \bar{x}) + \varepsilon_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta + \frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Thus

$$b = \beta + \frac{\sum_{i=1}^n \varepsilon_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

meaning that the estimated value b is equal to the true parameter β plus some error term determined by the collection of ε_i. As the sum of normal variables are normal, the error term itself has a normal distribution. In particular, since $\mathbb{E}[\varepsilon_i] = 0$, the error term has zero expectation and

$$\beta = \mathbb{E}[b],$$

which makes b an unbiased estimator of β. Thus the distribution of b is

$$b \sim N(\beta, \sigma(b)).$$

Let us determine the variance (or equivalently, the standard deviation) of b. We have

$$\mathrm{Var}(b) = \mathrm{Var}\left(\sum_{i=1}^n \varepsilon_i (x_i - \bar{x})\right) = \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2 \sum_{i=1}^n \mathrm{Var}(\varepsilon_i (x_i - \bar{x})) =$$
$$= \left(\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2 \sum_{i=1}^n \sigma^2 (x_i - \bar{x}))^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Thus

$$b \sim N\left(\beta, \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}\right).$$

In particular,

$$\frac{b - \beta}{\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim N(0, 1).$$

Note that

$$\mathrm{Var}(b) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{\sigma^2}{n}}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\frac{\sigma^2}{n}}{\mathrm{Var}(x)},$$

the variance of the slope b is smaller when the sample variance of x, $\mathrm{Var}(x)$, is larger. At the same time, the variance of the slope is smaller with more observation points n as the numerator in the last fraction converges to zero, while the denominator of the same fraction converges to a constant (the true variance of x).

The two sided α confidence interval for b is thus

$$\beta + q_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \leq b \leq \beta + q_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

where q are quantiles from the standard normal distribution $N(0,1)$. This in turn gives an α confidence interval for β in the form

$$b - q_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \leq \beta \leq b - q_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

The problem is that we typically do not know the exact value of σ, which in turn has to be estimated from the existing data. This is done similarly to estimation of the variance in the situation where an estimator of σ^2, $(X-\mu)^2$, had to be replaced with $(X-\bar{X})^2$. The unbiased estimator of σ^2 were given by $\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ when μ was known, but by $\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ when μ was unknown. The reason is that the estimator $\bar{X}$ of μ is positively correlated to the data, which shinks the variance of $\sum_{i=1}^n (X_i - \bar{X})$ to $(n-1)\sigma^2$. The same impact is observed in the regression estimation. The estimate of σ^2 is

$$\varepsilon_i^2 = (y_i - (\alpha + \beta x_i))^2$$

as $\mathbb{E}[\varepsilon_i^2] = \sigma^2$. Since α, β and ε_i are unobserved, we need to use the next best proxy, namely the squared residual u_i^2:

$$u_i^2 = (y_i - (a + bx_i))^2.$$

Since a and b are correlated to the data, it can be shown that

$$\mathbb{E}\left[\sum_{i=1}^n u_i^2\right] = (n-2)\sigma^2.$$

Thus

$$S^2 = \frac{1}{n-2}\sum_{i=1}^n u_i^2$$

is an unbiased estimator of σ^2. Recall that we have

$$\frac{b-\beta}{\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim N(0,1).$$

When we substitute the estimator S for σ, we also change the distribution of the ratio to t_{n-2}. Thus we have

$$\frac{b-\beta}{\frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2}.$$

In particular, for testing hypothesis $\beta = 0$, we should check the ratio

$$T = \frac{b}{\sigma(b)} = \frac{b}{\frac{\sqrt{\frac{1}{n-2}\sum_{i=1}^{n} u_i^2}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}$$

and determine its likelihood with respect to the t_{n-2} statistics. This is done via a number called a P-value, which gives the probability of the observed T being larger or smaller than a randomly generated value t_{n-2}. Formally, this is given by

$$\mathbb{P}(|T| > t_{n-2}).$$

Example 2.5

Consider that we observe the following 4 points (x_i, y_i):

x_i	y_i
1	2
2	1
3	1
4	0

Compute the regression estimates. Verify that $\sum_{i=1}^{n} u_i$ and $\sum_{i=1}^{n} u_i x_i$ sum up to zero. In order to solve the problem, let us expand the above table for several extra columns: $(x_i - \bar{x})$, $(y_i - \bar{y})$, $(x_i - \bar{x})(y_i - \bar{y})$, and $(x_i - \bar{x})^2$. The last two mentioned columns will determine the slope b. Note that $\bar{x} = \frac{5}{2}$ and $\bar{y} = 1$. For the sake of completness, let us add two additional columns, namely the list of residuals u_i and the product of residuals with data $u_i \cdot x_i$. Thus

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	u_i	$u_i \cdot x_i$
1	2	$-\frac{3}{2}$	1	$-\frac{3}{2}$	2.25	0.1	0.1
2	1	$-\frac{1}{2}$	0	0	0.25	-0.3	-0.6
3	1	$\frac{1}{2}$	0	0	0.25	0.3	0.9
4	0	$\frac{3}{2}$	- 1	$-\frac{3}{2}$	2.25	-0.1	-0.4

Thus

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{-\frac{3}{2} + 0 + 0 + (-\frac{3}{2})}{2.25 + 0.25 + 0.25 + 2.25} = -0.6,$$

and

$$a = \bar{y} - b\bar{x} = 1 - (-0.6) \cdot 2.5 = 2.5.$$

Thus the regression line is given by

$$y = 2.5 - 0.6 \cdot x.$$

The residuals are simply

$$u_i = y_i - (2.5 - 0.6 \cdot x_i),$$

which gives the residuals listed in the corresponding column in the above table. Note that both $\sum_{i=1}^{n} u_i$ and $\sum_{i=1}^{n} u_i x_i$ are equal to zero.

2.4.1 Multiple Regression and Matrix Notation

In our previous discussion, we had n equations in the form

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

This can be written more succintly in a matrix form as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

We can define

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \beta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

allowing us to write the linear regression problem in a matrix form as

$$y = X\beta + \varepsilon.$$

This notation allows for a fairly straightforward generalization to multiple regression problem, where we have more than 1 explanatory vectors x:

$$y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

in the same matrix form

$$\boxed{y = X\beta + \varepsilon.}$$

The only difference is the form of the matrix X:

$$X = \begin{pmatrix} 1 & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{2n} & \cdots & x_{kn} \end{pmatrix}.$$

The regression estimator is based on the same idea of minimizing the sqared residuals. Using the matrix notation, the residual vector is

$$u = y - Xb.$$

Using ' for transpose, we can write

$$\sum_{i=1}^{n} u_i^2 = u'u = (y - Xb)'(y - Xb) = y'y - y'Xb - b'X'y + b'X'Xb.$$

Taking the derivative with respect to b leads to

$$\frac{\partial S}{\partial b} = -2X'y + 2X'Xb.$$

Setting the derivative to zero, we obtain

$$X'Xb = X'y.$$

Solving for b (using the matrix inversion), we get

$$\boxed{b = (X'X)^{-1}X'y.}$$

This is indeed a default approach how Python implements the regression, meaning that one has to create the full X matrix that appears in the formulation of the problem. For instance, in our previous example, we had $x = (1, 2, 3, 4)'$, but we need to create a full matrix

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}.$$

in order to include the intercept, which is represented by the first column of the X matrix. The whole regression analysis can be done using the library 'statsmodels'. The code which includes the addition of the constant column is the following:

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt
import statsmodels.formula.api as sm
import statsmodels.tools as tools

x=[1,2,3,4]
x= tools.add_constant(x)
y=[2,1,1,0]
```

```
model = sm.OLS(y,x)

coefs = model.fit().params
residuals = model.fit().resid

print(sm.OLS(y, x).fit().summary())

plt.plot(x[:,1],y,'bo',ms=6)

xx=np.linspace(np.min(x[:,1])-1,np.max(x[:,1])+1,100)
plt.plot(xx,coefs[0]+coefs[1]*xx,'r')
```

This prints the following regression summary table:

```
OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.900
Model:                            OLS   Adj. R-squared:                  0.850
Method:                 Least Squares   F-statistic:                     18.00
Date:                Thu, 26 Oct 2017   Prob (F-statistic):             0.0513
Time:                        10:19:39   Log-Likelihood:                0.31571
No. Observations:                   4   AIC:                             3.369
Df Residuals:                       2   BIC:                             2.141
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.5000      0.387      6.455      0.023       0.834       4.166
x1            -0.6000      0.141     -4.243      0.051      -1.208       0.008
==============================================================================
Omnibus:                          nan   Durbin-Watson:                   3.400
Prob(Omnibus):                    nan   Jarque-Bera (JB):                0.308
Skew:                           0.000   Prob(JB):                        0.857
Kurtosis:                       1.640   Cond. No.                         7.47
==============================================================================
```

and illustrates the regression line in Figure 2.3.

□

2.4.2 Regression Analysis of Financial Data

We can apply the regression analysis on financial data. Let us take exchange rate data that is made available by the European Central Bank in a file 'eurofxref-hist.csv'. This file contains historical exchange rate data of 41 currencies that include all major currencies. Let us start with uploading the data to Python, computing the returns and plotting the return statistics. Processing .csv files is done by a library called 'pandas'. The code starts with importing the libraries, followed by uploading the data to a data frame object called 'df'. Make sure that the file appears in the Python working directory.

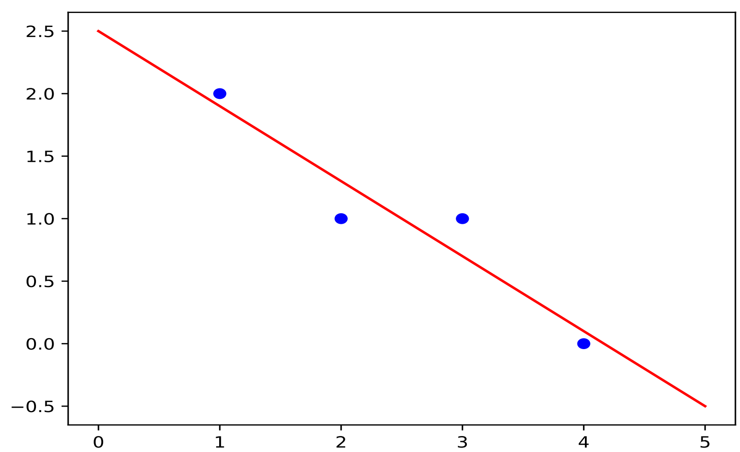

FIGURE 2.3: Points (x, y) and the corresponding regression line.

```
import numpy as np
import scipy.stats as sp
import matplotlib.pyplot as plt
import statsmodels.formula.api as sm
import statsmodels.tools as tools
import pandas as pd

df = pd.read_csv('eurofxref-hist.csv')

EURUSD = df['USD']
Dates = pd.to_datetime(df['Date'], format='%Y-%m-%d')
plt.figure(1)
plt.plot(Dates,EURUSD)

LogReturnsEURUSD = np.diff(np.log(EURUSD))
plt.figure(2)
plt.plot(Dates[1:], LogReturnsEURUSD)
plt.figure(3)
x=np.linspace(-0.045,0.045,120)
plt.hist(LogReturnsEURUSD, normed = True, bins=x)
plt.plot(x,sp.norm.pdf(x,np.mean(LogReturnsEURUSD),np.std(LogReturnsEURUSD)),'r')
```

The code then extracts the EURUSD exchange rate data, specifies the date format and plots the exchange rate evolution in Figure 2.4. Next, the code computes the log returns and plots them in Figure 2.5 The histogram of the

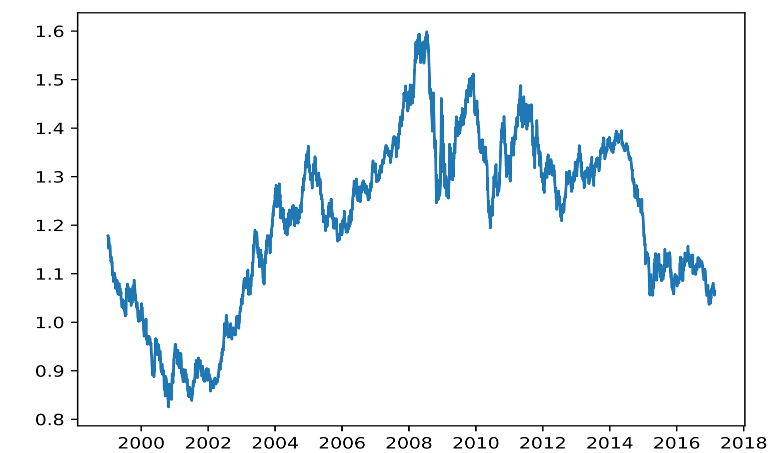

FIGURE 2.4: Evolution of the EURUSD exchange rate.

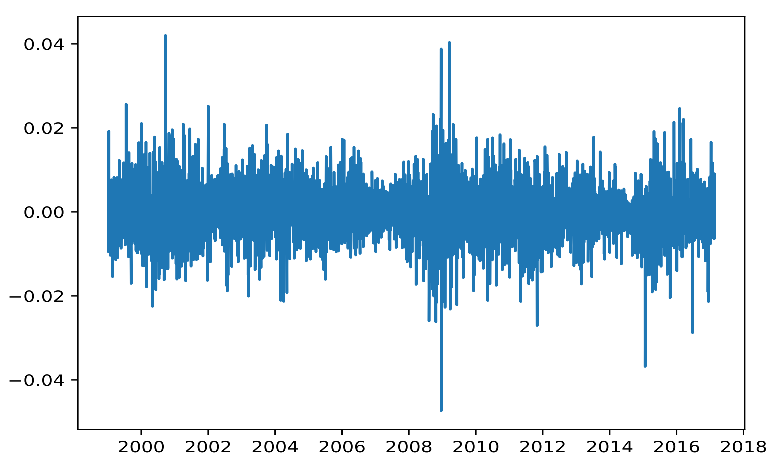

FIGURE 2.5: Log returns of the EURUSD exchange rate.

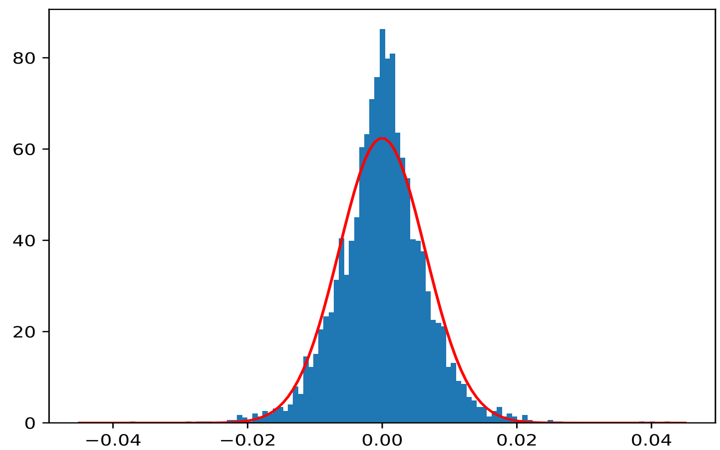

FIGURE 2.6: Histogram of the log returns of the EURUSD exchange rate.

log returns is plot in Figure 2.6. The fitted denstity of the normal distribution is plotted on the same graph. Note that the fit is not perfect - the reason is that the financial data tend to have fat tails that are extremely unlikely to happen if sampled from the normal distribution. Thus the normal distribution fit somewhat pushes to the tail while leaving some mass discrepancy in the values around zero. Other choice of the distribution, such as t, would lead to a better fit, but it would require to code a numerical routine for such a fit. Note that we can compute statistics called Value at Risk at this point. Value at Risk corresponds to a quantile of the loss distribution, typically taken at levels 95% or 99%. The loss corresponds to the minus profit, and thus we can take the bottom quantile (say 5%) with a minus sign to obtain Value at Risk at 95%. The empirical 5% quantile from the historical data is obtained from

```
np.percentile(LogReturnsEURUSD,5)
```

where we get -0.010354466319819351, so the 95% Value at Risk is 0.010354466319819351. This means that we should see only 5% scenarios that result in a daily loss greater than 0.010354466319819351. We can also obtain a parametric estimate using the quantile from the fitted normal distribution as

```
sp.norm.ppf(0.05,np.mean(LogReturnsEURUSD),np.std(LogReturnsEURUSD))
```

which gives -0.010532691093332634. The two values are remarkably close even in the situation when the visual fit of the normal distribution to the data is far from perfect.

Let us proceed with fitting a CAPM type model that links the returns of two assets. Let us take EURJPY exchange rate as the second asset. The model is

$$r_{EURUSD}(i) = \alpha + \beta \cdot r_{EURJPY}(i) + \varepsilon(i).$$

The following code prepares the returns of the EURJPY exchange rate, performs the linear regression, prints the results and plots the data with the corresponding regression line.

```
EURJPY = df['JPY']
LogReturnsEURJPY= np.diff(np.log(EURJPY))

plt.figure(4)
plt.plot(LogReturnsEURJPY,LogReturnsEURUSD,'.', ms=2)
model1 = sm.OLS(LogReturnsEURUSD,tools.add_constant(LogReturnsEURJPY))
print(model1.fit().summary())

xx=np.linspace(np.min(LogReturnsEURJPY),np.max(LogReturnsEURJPY),100)
coefs = model1.fit().params
plt.plot(xx,coefs[0]+coefs[1]*xx,'r')
```

We get the following output. As expected, the constant coefficient is not statistically significant, it looks as a statistical zero. In finance, we should not see two returns to be significantly shifted by a constant, otherwise the discrepancy between the returns would allow for generating profits. On the other hand, the slope is significant, roughly in the scale of 0.5. This means that a percentage shift in the EURJPY exchange rate results in a half percent shift in the EURUSD exchange rate. The corresponding graph is in Figure 2.7.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.354
Model:                            OLS   Adj. R-squared:                  0.354
Method:                 Least Squares   F-statistic:                     2549.
Date:                Wed, 01 Nov 2017   Prob (F-statistic):               0.00
Time:                        12:33:44   Log-Likelihood:                 17893.
No. Observations:                4644   AIC:                        -3.578e+04
Df Residuals:                    4642   BIC:                        -3.577e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -1.06e-05   7.54e-05     -0.141      0.888      -0.000       0.000
x1             0.4863      0.010     50.487      0.000       0.467       0.505
==============================================================================
Omnibus:                      268.581   Durbin-Watson:                   2.024
```

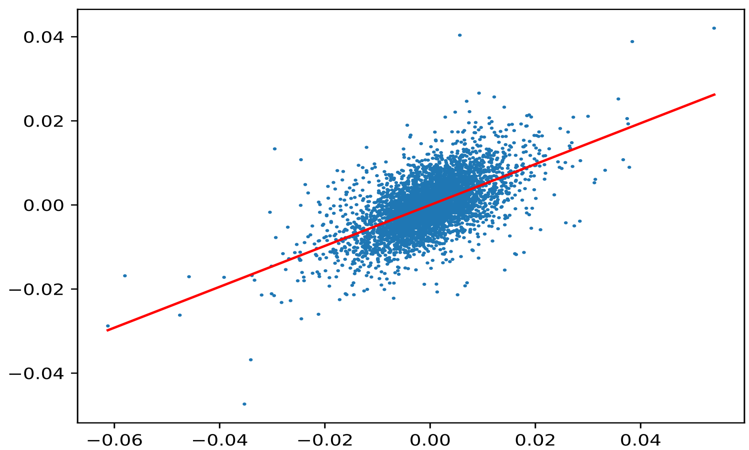

FIGURE 2.7: CAPM type model linking the returns of EURJPY with EURUSD.

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1040.812
Skew:                           0.136   Prob(JB):                    9.79e-227
Kurtosis:                       5.303   Cond. No.                         128.
==============================================================================
```

A natural question is whether this knowledge of the dependency of returns could lead to some profitable trading strategies. The problem with the above model is that it links two simultaneous returns, so if we see one return say going up, the other moves with it simultaneously, leaving no room for trading on that information. What is needed is a time lag, meaning whether an observation from the past could predict the return in the future. Let us limit our study to the model of EURUSD and consider a model

$$r_{EURUSD}(i+1) = \alpha + r_{EURUSD}(i) + \varepsilon(i).$$

This addresses the question whether the return from yesterday could predict the return today. If yes, we could trade on this information. The corresponding piece of the Python code is

```
LR_EURUSD_T = LogReturnsEURUSD[1:]
LR_EURUSD_Y = LogReturnsEURUSD[:len(LogReturnsEURUSD)-1]
plt.figure(5)
plt.plot(LR_EURUSD_Y,LR_EURUSD_T,'.', ms=2)

model = sm.OLS(LR_EURUSD_T,tools.add_constant(LR_EURUSD_Y))
```

```
coefs = model.fit().params

print(sm.OLS(LR_EURUSD_T, tools.add_constant(LR_EURUSD_Y)).fit().summary())

xx=np.linspace(np.min(LR_EURUSD_Y),np.max(LR_EURUSD_Y),100)
plt.plot(xx,coefs[0]+coefs[1]*xx,'r')
```

This outputs

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.2218
Date:                Wed, 01 Nov 2017   Prob (F-statistic):              0.638
Time:                        12:46:45   Log-Likelihood:                 16873.
No. Observations:                4643   AIC:                        -3.374e+04
Df Residuals:                    4641   BIC:                        -3.373e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -2.205e-05   9.38e-05     -0.235      0.814      -0.000       0.000
x1            -0.0069      0.015     -0.471      0.638      -0.036       0.022
==============================================================================
Omnibus:                      306.351   Durbin-Watson:                   2.000
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1475.801
Skew:                           0.009   Prob(JB):                         0.00
Kurtosis:                       5.762   Cond. No.                         156.
==============================================================================
```

together with the graph in Figure 2.8. As seen from both the model summary and the graph, the regression coefficients are insignificant, indicating that the future retuns are independent from the past returns and there is no such measurable dependency that could be extracted for a profitable trading strategy based on this data.

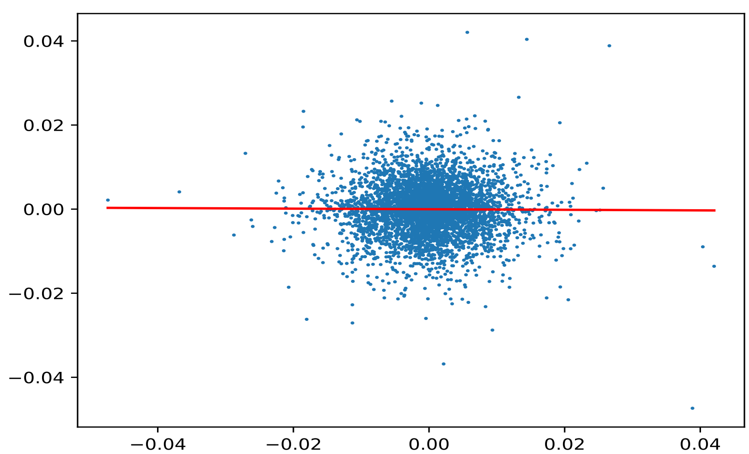

FIGURE 2.8: Relationship between yesterday's and today's returns of EURUSD

Quick Summary

Basic Concepts:

Probability Mass Function: $\mathbb{P}(X = k)$ for discrete distribution.

Probability Density Function: $f(x)$ for continuous distribution.

Cumulative Distribution Function: $F(x) = \mathbb{P}(X \leq x)$.
$F(x) = \sum_{y \leq x} \mathbb{P}(X = y)$ (discrete), $F(x) = \int_{-\infty}^{x} f(y)dy$ (continuous).
Note that $\mathbb{P}(X \in (a, b]) = F(b) - F(a) = \int_a^b f(x)dx$.

Expectation:

$$\mathbb{E}[X] = \sum_k k \cdot \mathbb{P}(X = k) \text{ (discrete)}, \qquad \mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)dx \text{ (continuous)}.$$

Variance:

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

$$\mathrm{Var}(X) = \sum_k (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k) \text{ (discrete)},$$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)dx \text{ (continuous)}.$$

Standard Deviation:

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

Covariance:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])] = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

$$\mathrm{Cov}(X, Y) = \sum_k \sum_l (k - \mathbb{E}[X]) \cdot (l - \mathbb{E}[Y]) \cdot \mathbb{P}(X = k, Y = l) \text{ (discrete)},$$

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X]) \cdot (y - \mathbb{E}[X]) f(x, y)dxdy \text{ (continuous)}.$$

Correlation:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Selected Univariate Distributions:

Binomial: $X \sim Bin(n,p) : \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \dots n$, $\mathbb{E}[X] = n \cdot p$, $\mathrm{Var}(X) = np(1-p)$.

Bernoulli: $X \sim Bin(1,p) : \mathbb{P}(X = 1) = p$, $\mathbb{E}[X] = p$, $\mathrm{Var}(X) = p(1-p)$.

Poisson: $X \sim Po(\lambda) : \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \dots,$ $\mathbb{E}[X] = \lambda$, $\mathrm{Var}(X) = \lambda$.

Geometric: $X \sim Geo(p) :, \mathbb{P}(X = k) = p(1-p)^{k-1}$, $k = 1, 2, \dots,$ $\mathbb{E}[X] = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

Normal: $X \sim N(\mu, \sigma)$: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$.

Exponential: $X \sim Exp(\lambda)$: $f(x) = \lambda e^{-\lambda x}, x > 0$, $\mathbb{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

Multivariate Distributions Concepts:

Joint Distribution: $\mathbb{P}(X = x, Y = y)$ (discrete), $f(x, y)$ (continuous).

Marginal Distribution: $\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y)$, $\mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y)$ (discrete), $f(x) = \int_{-\infty}^{\infty} f(x, y) dy$, $f(y) = \int_{-\infty}^{\infty} f(x, y) dx$ (continuous).

Conditional Distribution: $\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$ (discrete), $f(x|y) = \frac{f(x,y)}{f(y)}$ (continuous).

Independence: for every (x, y), marginal distribution = conditional distribution: $\mathbb{P}(X = x | Y = y) = \mathbb{P}(X = x)$, or $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$ (discrete), $f(x|y) = f(x)$, or $f(x, y) = f(x) \cdot f(y)$ (continuous).

Scaling Properties:

Expectation:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Variance:

$$\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + 2ab \, \mathrm{Cov}(X, Y) + b^2 \mathrm{Var}(Y).$$

Standard Deviation:

$$\sigma(aX) = |a| \cdot \sigma(X).$$

Covariance:

$$\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y).$$

Correlation:

$$\rho(aX, bY) = \rho(X, Y), \qquad a, b > 0.$$

Standard Score (Statistical Normalization, z-score): Random variable

$$Y = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}} = \frac{X - \mu}{\sigma}$$

has $\mathbb{E}[Y] = 0$, $\mathrm{Var}(Y) = 1$.

Large Sample:

Let X_i, $i = 1, \ldots, n$ be independent identically distributed (IID) random variables, $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1)$. Define

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

Then

$$\mathbb{E}[S_n] = n\mu, \qquad \mathrm{Var}(S_n) = n\sigma^2, \qquad \mathbb{E}[\bar{X}_n] = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \tfrac{\sigma^2}{n}.$$

Quantiles:

Given $\alpha \in (0, 1)$, quantile q_α is the smallest value such that

$$F(q_\alpha) = \mathbb{P}(X \le q_\alpha) \ge \alpha.$$

Alternatively,

$$q_\alpha = F^{-1}(\alpha).$$

Confidence Intervals:

$[a, b]$ that satisfies $\mathbb{P}(X \in [a, b]) = \alpha$ for a given $\alpha \in (0, 1)$ is called an α confidence interval.

One Sided Confidence Intervals: $(-\infty, q_\alpha]$, or $[q_{1-\alpha}, +\infty)$.

Two Sided Confidence Interval: $[q_{\frac{1-\alpha}{2}}, q_{\frac{1+\alpha}{2}}]$.

Law of Large Numbers:

Sample average converges to the expectation:

$$\bar{X}_n \to \mu.$$

Central Limit Theorem (CLT):

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1).$$

Confidence Intervals Based on CLT:

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \in [a,b]\right) \approx \Phi(b) - \Phi(a), \quad \text{for} \quad \Phi(x) = \int_{-\infty}^{x} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy.$$

Two sided α confidence interval:

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \in [q_{\frac{1-\alpha}{2}}, q_{\frac{1+\alpha}{2}}]\right) \approx \Phi(q_{\frac{1+\alpha}{2}}) - \Phi(q_{\frac{1-\alpha}{2}}) = \tfrac{1+\alpha}{2} - \tfrac{1-\alpha}{2} = \alpha.$$

$$\mathbb{P}(S_n \in [n\mu + q_{\frac{1-\alpha}{2}}\sigma\sqrt{n}, n\mu + q_{\frac{1+\alpha}{2}}\sigma\sqrt{n}]) \approx \alpha.$$

$$\mathbb{P}(\bar{X}_n \in [\mu + q_{\frac{1-\alpha}{2}}\tfrac{\sigma}{\sqrt{n}}, \mu + q_{\frac{1+\alpha}{2}}\tfrac{\sigma}{\sqrt{n}}]) \approx \alpha.$$

One sided α confidence intervals:

$$\mathbb{P}(S_n \in (-\infty, n\mu + q_\alpha\sigma\sqrt{n}]) \approx \alpha.$$

$$\mathbb{P}(S_n \in [n\mu + q_{1-\alpha}\sigma\sqrt{n}, +\infty)) \approx \alpha.$$

$$\mathbb{P}(\bar{X}_n \in (-\infty, \mu + q_\alpha\tfrac{\sigma}{\sqrt{n}}]) \approx \alpha.$$

$$\mathbb{P}(\bar{X}_n \in [\mu + q_{1-\alpha}\tfrac{\sigma}{\sqrt{n}}, +\infty)) \approx \alpha.$$

Specific choice of confidence interval $\alpha = 0.95$ leads to quantiles $q_{\frac{1-\alpha}{2}} = q_{0.025} \approx -1.96$ and $q_{\frac{1+\alpha}{2}} = q_{0.975} \approx 1.96$ for two sided confidence intervals and to quantiles $q_{1-\alpha} = q_{0.05} \approx -1.64$, $q_\alpha = q_{0.95} \approx 1.64$ for one sided intervals. Thus

$$\mathbb{P}(S_n \in [n\mu \pm 1.96\sigma\sqrt{n}]) \approx 95\%,$$

$$\mathbb{P}(\bar{X}_n \in [\mu \pm 1.96\tfrac{\sigma}{\sqrt{n}}]) \approx 95\%,$$

and

$$\mathbb{P}(S_n \in (-\infty, n\mu + 1.64\sigma\sqrt{n}]) \approx 95\%.$$

$$\mathbb{P}(S_n \in [n\mu - 1.64\sigma\sqrt{n}, +\infty)) \approx 95\%.$$

$$\mathbb{P}(\bar{X}_n \in (-\infty, \mu + 1.64\tfrac{\sigma}{\sqrt{n}}]) \approx 95\%.$$

$$\mathbb{P}(\bar{X}_n \in [\mu - 1.64\tfrac{\sigma}{\sqrt{n}}, +\infty)) \approx 95\%.$$

Linear Regression:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$a = \bar{y} - b\bar{x}.$$

Exercises

Problem 1. Consider the following data points sample:

$$S = \{9, 5, 6, -8, 7, -4\}.$$

Calculate the mean and the median of the sample.

Solution:
Mean:

$$\frac{9 + 5 + 6 + (-8) + 7 + (-4)}{6} = \frac{15}{6} = 2.5$$

Median: Order the data in an increasing order:

$$-8, -4, 5, 6, 7, 9$$

Median of a finite sample is defined as the $\frac{n+1}{2}$ value of an ordered data set when n is odd, and as an average of the $\frac{n}{2}$ and $\frac{n}{2} + 1$ values when n is even. In our case, $n = 6$, so the median is the average of the 3rd and 4th values (5 and 6):

$$\frac{5 + 6}{2} = 5.5.$$

Problem 2.
(a) Let X be a normally distributed random variable with mean 2, and standard deviation 4. Determine $\mathbb{P}(X \leq 0)$.
(b) Assume that one generates a sample of 100 independent random variables $X(i)$, $i = 1, \ldots 100$ from the above normal distribution $N(1, 2)$. Denote by $\bar{X}$ the average defined as

$$\bar{X}(n) = \frac{1}{100} \sum_{i=1}^{100} X(i).$$

Compute the mean and the standard deviation of $\bar{X}(n)$. Give a 95% confidence interval for $\bar{X}(n)$.

Solution:
(a) This problem can be transformed to the standard normal variable using

the fact that when $X \sim N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$:

$$\begin{aligned}\mathbb{P}(X \leq 0) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq -\frac{\mu}{\sigma}\right)\\ &= \mathbb{P}\left(Z \leq -\frac{2}{4}\right)\\ &= \Phi(-\tfrac{1}{2}) \approx 0.308538.\end{aligned}$$

The random variable Z is standard normal $N(0,1)$ and Φ is the cumulative distribution function of the standard normal variable.
(b) When $X \sim N(\mu, \sigma)$, the average $\bar{X}_n$ has $N(\mu, \frac{\sigma}{\sqrt{n}})$ distribution. The variance of the average is n times smaller than the variance of an individual observation (and the standard deviation is $\sqrt{n}$ smaller than the standard deviation of an individual observation). The mean is computed as

$$\mathbb{E}[\bar{X}(n)] = \mathbb{E}\left[\frac{1}{100}\sum_{i=1}^{100} X(i)\right] = \frac{1}{100}\sum_{i=1}^{100}\mathbb{E}[X(i)] = \frac{1}{100}\cdot 100 \cdot \mu = \mu = 1,$$

the variance is given by

$$\begin{aligned}\mathrm{Var}(\bar{X}(n)) = \mathrm{Var}\left(\frac{1}{100}\sum_{i=1}^{100} X(i)\right) &= \left(\frac{1}{100}\right)^2 \sum_{i=1}^{100}\mathrm{Var}(X(i))\\ &= \left(\frac{1}{100}\right)^2 \cdot 100 \cdot \sigma^2 = \frac{\sigma^2}{100} = \frac{4}{100} = \frac{1}{25}.\end{aligned}$$

We have used linearity of the expectation and the properties of the variance (listed in the main text). The variance of the sum is the sum of the variances when the random variables are independent.
The two sided 95% confidence interval for a normal variable $N(\mu, \sigma)$ is given by

$$(\mu - 1.96 \cdot \sigma, \mu + 1.96 \cdot \sigma)$$

(approximately 95% of the values lie within the two standard deviations away from the mean). Keep in mind that the standard deviation of the average is $\frac{\sigma}{\sqrt{n}}$, in our case the value is $\sqrt{\frac{1}{25}} = \frac{1}{5}$. Thus the 95% confidence interval for the average is

$$\left(1 - 1.96 \cdot \frac{1}{5}, 1 + 1.96 \cdot \frac{1}{5}\right) = (0.608, 1.392).$$

Problem 3. Consider random variables X and Y with the following observed values:

X	1	2	3	4
Y	3	2	5	4

(a) Compute correlation coefficient between X and Y.
(b) Compute the regression parameters of $Y = a_0 + a_1 \cdot X$.

Solution: (a) First, compute $\bar{X}$, $\bar{Y}$ and $\overline{X \cdot Y}$ (from the sample). We get

$$\bar{X} = \frac{1+2+3+4}{4} = \frac{5}{2}, \quad \bar{Y} = \frac{3+2+5+4}{4} = \frac{7}{2},$$

and

$$\overline{X \cdot Y} = \frac{1 \cdot 3 + 2 \cdot 2 + 3 \cdot 5 + 4 \cdot 4}{4} = \frac{19}{2}.$$

Thus

$$\mathrm{Cov}(X, Y) = \overline{X \cdot Y} - \bar{X} \cdot \bar{Y} = \frac{19}{2} - \frac{5}{2} \cdot \frac{7}{2} = \frac{3}{4}.$$

The sample variances are

$$s_x^2 = \frac{1}{4} \cdot \left((1 - \tfrac{5}{2})^2 + (2 - \tfrac{5}{2})^2 + (3 - \tfrac{5}{2})^2 + (4 - \tfrac{5}{2})^2 \right) = \frac{5}{4},$$

$$s_y^2 = \frac{1}{4} \cdot \left((3 - \tfrac{7}{2})^2 + (2 - \tfrac{7}{2})^2 + (5 - \tfrac{7}{2})^2 + (4 - \tfrac{7}{2})^2 \right) = \frac{5}{4},$$

and thus the correlation coefficient equals to

$$r_{xy} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{\frac{3}{4}}{\sqrt{\frac{5}{4} \cdot \frac{5}{4}}} = \frac{3}{5}.$$

(b)

$$a_1 = r_{xy} \cdot \frac{s_y}{s_x} = \frac{3}{5} \cdot \frac{\sqrt{\frac{5}{4}}}{\sqrt{\frac{5}{4}}} = \frac{3}{5},$$

$$a_0 = \bar{Y} - a_1 \cdot \bar{X} = \frac{7}{2} - \frac{3}{5} \cdot \frac{5}{2} = 2.$$

Problem 4. Binary digits (zeros and ones) are sent down a noisy communication channel. They are received as sent with probability 0.9 but errors occur with probability 0.1. Assuming that 0's and 1's are equally likely, what is probability that 1 was sent given that 1 was received?

Solution:
Let A be the event $\{1 \text{ is sent}\}$ and B the event $\{1 \text{ is received}\}$. We want to compute $\mathbb{P}(A|B)$. We have $\mathbb{P}(A) = 0.5$ and we know $\mathbb{P}(B|A) = 0.9$. Now

$$\begin{aligned}\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} &= \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)} \\ &= \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)} = \frac{0.9 \cdot 0.5}{0.9 \cdot 0.5 + 0.1 \cdot 0.5} = 0.9.\end{aligned}$$

Problem 5.
(a) Suppose we draw 13 cards out of a deck of 52 (bridge hand). What is the probability that you get no ace?
(b) What is the expected number of different card denominations in the bridge hand? (Follows from (a)).
(c) Are the events of getting no ace and no king dependent or independent?
(d) One of the players in bridge (your partner) has one card from each denomination, 13 denomination in total. What is the probability of this event?
(e) How the answers from (a) and (b) change for your hand given that you have this information from (d)?

Solution:
(a)

$$\frac{\binom{48}{13}}{\binom{52}{13}}.$$

(b) For each denomination i (there are 13 different ones) define a random variable $Y(i)$ as

$$Y(i) = \begin{cases} 1 \text{ if the } i\text{th denomination appears in the bridge hand,} \\ 0 \text{ if the } i\text{th denomination does not appear in the bridge hand.} \end{cases}$$

The number of different denominations in the hand is then given by a variable

$$S = \sum_{i=1}^{13} Y(i).$$

In order to compute the expectation of S, we use

$$\begin{aligned}\mathbb{E}[S] = \mathbb{E}\left[\sum_{i=1}^{13} Y(i)\right] &= \sum_{i=1}^{13} \mathbb{E}[Y(i)] = \sum_{i=1}^{13} [1 \cdot \mathbb{P}(Y(i) = 1) + 0 \cdot \mathbb{P}(Y(i) = 0)] \\ = \sum_{i=1}^{13} \mathbb{P}(Y(i) = 1) &= \sum_{i=1}^{13} (1 - \mathbb{P}(Y(i) = 0)) = 13 \cdot \left(1 - \frac{\binom{48}{13}}{\binom{52}{13}}\right) \approx 9.05.\end{aligned}$$

(c) Let A be the event of having no ace in the bridge hand and B the event of having no king in the bridge hand. We have seen that

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{\binom{48}{13}}{\binom{52}{13}}.$$

Independence requires

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

The probability of the joint event $\mathbb{P}(A \cap B)$ is

$$\frac{\binom{44}{13}}{\binom{52}{13}}$$

(remove all 4 aces and 4 kings from the deck of 52 cards). But

$$\frac{\binom{44}{13}}{\binom{52}{13}} \neq \frac{\binom{48}{13}}{\binom{52}{13}} \cdot \frac{\binom{48}{13}}{\binom{52}{13}},$$

so the events A and B are dependent.
(d)

$$\frac{4^{13}}{\binom{52}{13}} \approx 0.0001.$$

Each of the 13 denominations can be chosen 4 times.
(e) If one player has all 13 denominations, it leaves 39 cards with 3 cards for each denomination. Thus probability of getting no ace in the bridge hand from the remaining cards becomes

$$\frac{\binom{36}{13}}{\binom{39}{13}}$$

and the expected number of denominations in a given bridge hand changes to

$$13\left(1 - \frac{\binom{36}{13}}{\binom{39}{13}}\right) \approx 9.30.$$

Problem 6. An insurance company has 1,000,000 clients. The probability that a client has an insurance claim in a given year is $\frac{1}{10}$.
(a) Find the two sided 95% confidence interval for X, the number of insurance claims in a given year. In other words, find a and b such that

$$\mathbb{P}(a \leq X) = 0.025, \qquad \mathbb{P}(X \leq b) = 0.025.$$

(b) Find one sided confidence interval for X in the form $(0, c)$. In other words, find c such that

$$\mathbb{P}(X \leq c) = 0.95.$$

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$

$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
The number of insurance claims X has binomial distribution $Bin(1000000, \frac{1}{10})$. According to the Central Limit Theorem,

$$Bin(n,p) \sim N(n \cdot p, \sqrt{np(1-p)}).$$

(Recall: $np(1-p)$ is the variance of the binomial distribution.) In our case, $n = 1000000$ and $p = \frac{1}{10}$, so

$$Bin(1000000, \tfrac{1}{10}) \sim N(100000, 300).$$

Let us find a such that $\mathbb{P}(a \leq X) = 0.025$. The number a represents 0.025 quantile of the random variable X. This can be found using the quantiles of the standard normal distribution. When $X \sim N(\mu, \sigma)$, the transformed variable $\frac{X-\mu}{\sigma} \sim N(0,1)$. Thus

$$\mathbb{P}(a \leq X) = \mathbb{P}(\tfrac{a-\mu}{\sigma} \leq \tfrac{X-\mu}{\sigma}) = \mathbb{P}(\tfrac{a-\mu}{\sigma} \leq Z),$$

where $Z \sim N(0,1)$. But

$$\mathbb{P}(q_{0.025} \leq Z) = 0.025$$

from the definition of the quantile. Therefore we must have

$$\frac{a-\mu}{\sigma} = q_{0.025},$$

which translates to

$$a = \mu + q_{0.025} \cdot \sigma.$$

We have $\mu = 100000$, $\sigma = \sqrt{90000}$ and $q_{0.025} \approx -1,96$. Therefore

$$a = 99412.$$

Similarly,

$$b = \mu + q_{0.975} \cdot \sigma = 100588.$$

The 95% two sided confidence interval for the number of insurance claims is thus

$$(99412, 100588).$$

(b) Using the same arguments,

$$c = \mu + q_{0.95} \cdot \sigma = 100000 + 1.64 \cdot \sqrt{90000} = 100492.$$

Problem 7. The probability of dying in an airplane crash is about 1:10,000,000 (one in ten millions flights).
(a) The largest world airline (Delta Airlines) operates about one million flights per year. What is the probability that in a given year it has no fatal crash? What is the probability that it has exactly one fatal crash? What is the probability that it has two or more fatal crashes? (Hint: Use Poisson approximation.)
(b) Quantas Airlines operates about 150,000 flights per year (judging from the size of its fleet). What is the probability that it has no fatal crash in a given year? Assuming that this probability is constant for every year, which is a reasonable assumption (in the past there were less flights but higher probability of fatalities), what is the probability that this airline has no fatal crash in 60 years? Would you say this is unusual? (Quantas has no fatal crash since 1951).
(c) A pilot flies around 10,000 flights in his career. What is the probability that he would make it safely to his retirement (= no fatal airplane crash in his career)? How does this probability compare with the probability of not dying in a car crash on the way to his work? (Fatality incidence in a car crash is 1 per 300,000,000 km traveled. Assume that the pilot has to drive 30 km on average to the airport – 300,000 km in the lifetime.)

Solution:
(a) Let X be the number of fatal crashes. The exact distribution of the random variable X is $Bin(n, p)$, where $n = 1,000,000$, $p = \frac{1}{10,000,000}$. Using Poisson approximation, $Bin(n, p) \sim Po(np)$. Thus

$$\mathbb{P}(X = 0) \approx e^{-np} = e^{-0.1} = 0.904837,$$
$$\mathbb{P}(X = 1) \approx np \cdot e^{-np} = 0.1 \cdot e^{-0.1} = 0.090483,$$
$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \approx 0.00468.$$

(b) Quantas flew a (fatal crash risk) equivalent of

$$n = 150,000 \cdot 60 = 9,000,000$$

flights over the last 60 years. Probability of no crash is

$$\mathbb{P}(X = 0) \approx e^{-np} = e^{-0.9} = 0.40657.$$

The probability that Quantas did not experience any fatal crash just by a random coincidence over the last 60 years is rather high.

(c) Probability of no fatal crash for a professional pilot is approximately

$$\mathbb{P}(X=0) = e^{-np} = e^{-\frac{10,000}{10,000,000}} = e^{-0.001} = 0.999.$$

Probability of no fatal car crash is

$$\mathbb{P}(X=0) = e^{-\frac{300,000}{300,000,000}} = e^{-0.001} = 0.999,$$

which is the same. Driving 30 kilometers has the same risk as taking one flight.

Problem 8. Let T have a geometric distribution

$$\mathbb{P}(T=k) = p \cdot (1-p)^{k-1}, \quad k = 1, 2, \ldots.$$

Find one sided $(1-\alpha)$-confidence interval for T in the form $[1, k_{1-\alpha}]$. In other words, find the smallest integer $k_{1-\alpha}$ such that

$$\mathbb{P}(T \leq k_{1-\alpha}) \geq 1-\alpha.$$

(Hint: Use the cumulative distribution function $F(k) = \mathbb{P}(T \leq k)$ for the geometric distribution. Find k such that $F(k) = 1-\alpha$. Find the corresponding nearest integer that satisfies the inequality.)
List $k_{0.95}$ for $p = \frac{1}{2}$ (coin toss), $p = \frac{1}{6}$ (die roll), $p = \frac{1}{37}$ (roulette), and $p = \frac{1}{139,838,160}$ (jackpot in the German Lotto).

Solution: This is a question for the quantile of the geometric distribution. The cumulative distribution function of the geometric distribution is given by

$$F(k) = \mathbb{P}(T \leq k) = 1 - \mathbb{P}(T > k) = 1 - (1-p)^k.$$

We want to find a value k such that

$$F(k) = 1 - \alpha.$$

Solving for

$$1 - (1-p)^k = 1 - \alpha$$

gives

$$k = \frac{\ln(\alpha)}{\ln(1-p)}.$$

The value of k may not be an integer, in which case $k_{1-\alpha}$ must be the nearest higher integer (also known as the ceiling of k).
When $\alpha = 0.05$, we get the following values for $k_{1-\alpha}$: Note that for small p,

$$\ln(1-p) \approx -p.$$

We also have $\ln(0.05) \approx -3$. Thus

$$k_{0.95} \approx \frac{3}{p}.$$

p	$k_{0.95}$
$\frac{1}{2}$	5
$\frac{1}{6}$	17
$\frac{1}{37}$	110
$\frac{1}{139{,}838{,}160}$	418,917,687

For instance, to guarantee 95% chance of a jackpot win in the lottery, one needs to file 3 times as many (random) tickets as there are combinations. Similar result holds for win in a roulette ($3 \cdot 37 = 111$ approximates 110) or for a waiting for number 6 in a die roll ($3 \cdot 6 = 18$ approximates 17).

Problem 9. **(a)** Let $x_1, x_2, \ldots, x_n$ be the observations from the exponential distribution with density

$$f(x) = \lambda e^{-\lambda x}, x > 0.$$

Find the maximum likelihood estimator of λ.
(b) Let $x_1, x_2, \ldots, x_n$ be the observations from the Poisson distribution

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, x > 0.$$

Find the maximum likelihood estimator of λ.

Solution:
(a) The likelihood function is

$$L(x_1, x_2, \ldots, x_n, \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

The log likelihood function is

$$\ln(L) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

We want to find the maximum with respect to the parameter λ, which can be identified as a point where the derivative with respect to λ is zero. The derivative of the log likelihood with respect to lambda is

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i,$$

which is equal to zero for

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i},$$

the inverse of the average of x_i.
(b) The likelihood function is

$$L(x_1, x_2, \ldots, x_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}.$$

The log likelihood function is

$$\ln(L) = -n\lambda + \ln(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!).$$

The derivative with respect to the parameter λ is

$$-n + \frac{1}{\lambda} \sum_{i=1}^{n} x_i,$$

which is equal to zero for

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

the average of x_i.

Problem 10. Consider random variables X and Y with the following observed values:

X	2	3	4	5
Y	2	1	4	3

(a) Compute correlation coefficient between X and Y.
(b) Compute the regression parameters of $Y = a_0 + a_1 \cdot X$.

Solution: (a) First, compute $\bar{X}$, $\bar{Y}$ and $\overline{X \cdot Y}$ (from the sample). We get

$$\bar{X} = \frac{2+3+4+5}{4} = \frac{7}{2}, \quad \bar{Y} = \frac{2+1+4+3}{4} = \frac{5}{2},$$

and

$$\overline{X \cdot Y} = \frac{2 \cdot 2 + 3 \cdot 1 + 4 \cdot 4 + 5 \cdot 3}{4} = \frac{19}{2}.$$

Thus

$$\mathrm{Cov}(X, Y) = \overline{X \cdot Y} - \bar{X} \cdot \bar{Y} = \frac{19}{2} - \frac{7}{2} \cdot \frac{5}{2} = \frac{3}{4}.$$

The sample variances are

$$s_x^2 = \frac{1}{4} \cdot \left((2 - \tfrac{7}{2})^2 + (3 - \tfrac{7}{2})^2 + (4 - \tfrac{7}{2})^2 + (5 - \tfrac{7}{2})^2 \right) = \frac{5}{4},$$

$$s_y^2 = \frac{1}{4} \cdot \left((2 - \tfrac{5}{2})^2 + (1 - \tfrac{5}{2})^2 + (4 - \tfrac{5}{2})^2 + (3 - \tfrac{5}{2})^2 \right) = \frac{5}{4},$$

and thus the correlation coefficient equals to

$$r_{xy} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{\frac{3}{4}}{\sqrt{\frac{5}{4} \cdot \frac{5}{4}}} = \frac{3}{5}.$$

(b)

$$a_1 = r_{xy} \cdot \frac{s_y}{s_x} = \frac{3}{5} \cdot \frac{\sqrt{\frac{5}{4}}}{\sqrt{\frac{5}{4}}} = \frac{3}{5},$$

$$a_0 = \bar{Y} - a_1 \cdot \bar{X} = \frac{5}{2} - \frac{3}{5} \cdot \frac{7}{2} = \frac{2}{5}.$$

Problem 11.
(a) 5 cards are picked out of a deck of 52. What is the probability that you get exactly two hearts?
(b) A class has 6 girls and 14 boys. 4 are randomly chosen. What is the probability that exactly 2 girls and 2 boys are picked out?
(c) Euromillions lottery draws 5 main numbers out of 50 plus 2 star numbers out of 11. What is more likely, to get 4 main numbers and no star numbers (4+0), or 3 main numbers and 2 star numbers (3+2)?

Solution:
(a) Choice of 2 cards out of 13 hearts and 3 cards out of the remaining 39:

$$\frac{\binom{13}{2} \cdot \binom{39}{3}}{\binom{52}{5}} \approx 0.27428$$

(b) Choice of 2 girls out of 6 and 2 boys out of 14, total choice is 4 out of 20:

$$\frac{\binom{6}{2} \cdot \binom{14}{2}}{\binom{20}{4}} \approx 0.281734.$$

(c) Exactly 4 main numbers has probability

$$\frac{\binom{5}{4} \cdot \binom{45}{1} \cdot \binom{9}{2}}{\binom{50}{5} \cdot \binom{11}{2}} \approx 0.0000695089.$$

We are choosing good 4 main numbers out of 5 on the ticket, one bad main number out of 45 not on the ticket plus 2 out of the 9 not chosen star numbers. Combination 3+2 has probability

$$\frac{\binom{5}{3} \cdot \binom{45}{2} \cdot \binom{2}{2}}{\binom{50}{5} \cdot \binom{11}{2}} \approx 0.0000849554.$$

We are choosing 3 main numbers out of 5 on the ticket, 2 bad main numbers out of 45 not on the ticket plus 2 out of the 2 star numbers. Combination 3+2 is more likely than combination 4+0.

Problem 12. The number of goals scored during a match by an average Bundesliga football team is an integer random variable with the mean $\mu = 1.41$ and the standard deviation $\sigma = 1.27$. Let X denote the number of goals of a given team in the whole season which includes 34 matches in total.
(a) Find the two sided 95% confidence interval for X, the number of goals scored by an average team in one season. In other words, find a and b such that

$$\mathbb{P}(X \le a) = 0.025, \qquad \mathbb{P}(b \le X) = 0.025.$$

(b) Find one sided confidence interval for X in the form $(0, c)$. In other words, find c such that

$$\mathbb{P}(X \le c) = 0.95.$$

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$

$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
(a) The 95% two sided confidence interval for a normal variable $N(\mu, \sigma)$ is given by

$$(\mu - 1.96\sigma, \mu + 1.96\sigma).$$

The distribution of the total goals in $n = 34$ games is approximately normal $N(n\mu, n\sigma)$, so the 95% confidence interval is given by

$$(n\mu - 1.96\sigma\sqrt{n}, n\mu + 1.96\sigma\sqrt{n}) = (33.43, 62.45) = [34, 62]$$

as the goals take only integer values.
(b) The one sided confidence interval is

$$(-\infty, n\mu + 1.64\sigma\sqrt{n}) = (-\infty, 60.08) = [0, 60].$$

Problem 13. The probability of winning a jackpot in the German Lotto is $p = \frac{1}{139{,}838{,}160}$.
(a) Suppose that a player bets 10 random tickets for each draw for 100 years. Since there are 2 draws per week, this totals to $10 \cdot 100 \cdot 100 = 100{,}000$ tickets (assuming that a year has 50 weeks). What is the probability that the players wins the jackpot during this period?
(b) Assuming that the player buys N random independent tickets, what is the smallest number of tickets he/she needs to buy in order to have a 50%

chance of winning the jackpot? What is the smallest number of tickets he/she needs to buy in order to have a 95% chance?

Solution:
(a) The number of jackpot wins X has a binomial distribution $Bin(100000, \frac{1}{139{,}838{,}160})$, so the probability of at least one win is

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - (1-p)^n$$

for $p = \frac{1}{139{,}838{,}160}$ and $n = 100000$. The Poisson approximation states that $Bin(n, p) \approx Po(np)$, so

$$\mathbb{P}(X = 0) = (1-p)^n = e^{n \cdot \ln(1-p)} \approx e^{-np}.$$

Poisson approximation is equivalent to the statement $\ln(1-p) \approx -p$. In our case,

$$\mathbb{P}(X = 0) \approx e^{-np} = 0.999285,$$

so

$$\mathbb{P}(X \geq 1) \approx 0.000714857.$$

(b) We want to find N such that

$$\mathbb{P}(X \geq 1) = 1 - \alpha$$

for $\alpha = 0.5$ and $\alpha = 0.05$. We have seen that

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - (1-p)^N.$$

Thus we need to solve for

$$1 - (1-p)^N = 1 - \alpha,$$

which gives

$$N = \frac{\ln(\alpha)}{\ln(1-p)}.$$

If we used Poisson approximation, we would get

$$1 - e^{-Np} = 1 - \alpha,$$

which gives

$$N = -\frac{\ln(\alpha)}{p}.$$

This is the same as saying $\ln(1-p) \approx -p$. From the exact formula we get that we see a win with 50% probability for

$$N = 96{,}928{,}427,$$

and a win with 95% for

$$N = 418,917,687.$$

Problem 14. Let T have an exponential distribution with density

$$f(t) = \lambda e^{-\lambda t}, \qquad t > 0.$$

Find two sided $(1-\alpha)$-confidence interval for T in the form $\left[t_{\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}\right]$. In other words, find the number $t_{\frac{\alpha}{2}}$ such that

$$\mathbb{P}\left(T \le t_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2},$$

and the number $t_{1-\frac{\alpha}{2}}$ such that

$$\mathbb{P}\left(T \le t_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

(Hint: Use the cumulative distribution function $F(t) = \mathbb{P}(T \le t)$ for the exponential distribution.)

Solution:
The cumulative distribution function of the exponential distribution is given by

$$F(t) = \mathbb{P}(T \le t) = 1 - e^{-\lambda t}.$$

The t_α quantiles of the exponential distribution satisfy

$$F(t_\alpha) = \alpha,$$

or

$$1 - e^{-\lambda t_\alpha} = \alpha.$$

This gives

$$t_\alpha = -\frac{\ln(1-\alpha)}{\lambda}.$$

Thus the $(1-\alpha)$ confidence interval is given by

$$\left[t_{\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}\right] = \left[-\frac{\ln(1-\frac{\alpha}{2})}{\lambda}, -\frac{\ln(\frac{\alpha}{2})}{\lambda}\right].$$

Problem 15. (a) Let $x_1, x_2, \ldots, x_m$ be the observations from the geometric distribution

$$P(X = k) = p \cdot (1-p)^{k-1}, \quad k = 1, 2, \ldots.$$

Find the maximum likelihood estimator of p.

(b) Let $x_1, x_2, \ldots, x_m$ be the observations from the binomial distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Find the maximum likelihood estimator of p.

Solution:

(a) The likelihood function is equal to

$$L(x_1, x_2, \ldots, x_m, p) = \prod_{i=1}^{m} p(1-p)^{x_i - 1} = p^m (1-p)^{\sum_{i=1}^{m}(x_i - 1)}.$$

The log likelihood is given by

$$\ln(L) = m \ln(p) + \ln(1-p) \cdot \sum_{i=1}^{m} (x_i - 1).$$

The derivative with respect to p gives

$$\frac{m}{p} - \frac{\sum_{i=1}^{m}(x_i - 1)}{1-p}.$$

This is equal to zero when

$$m(1-p) = p\left(\sum_{i=1}^{m} x_i - m\right),$$

or when

$$\hat{p} = \frac{m}{\sum_{i=1}^{m} x_i},$$

the inverse of the average.

(b) The likelihood function is equal to

$$L(x_1, x_2, \ldots, x_m, p) = \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} = p^{\sum_{i=1}^{m} x_i} (1-p)^{\sum_{i=1}^{m}(n-x_i)} \prod_{i=1}^{m} \binom{n}{x_i}.$$

The log likelihood is given by

$$\ln(L) = \ln(p) \sum_{i=1}^{m} x_i + \ln(1-p) \sum_{i=1}^{m} (n - x_i) + \sum_{i=1}^{m} \ln\left(\binom{n}{x_i}\right).$$

The derivative with respect to p gives

$$\frac{1}{p} \sum_{i=1}^{m} x_i - \frac{1}{1-p} \sum_{i=1}^{m} (n - x_i).$$

This is equal to zero when

$$(1-p)\sum_{i=1}^{m} x_i = p\sum_{i=1}^{m}(n-x_i),$$

or

$$\sum_{i=1}^{m} x_i = p \cdot n \cdot m.$$

Thus

$$\hat{p} = \frac{\sum_{i=1}^{m} x_i}{nm} = \frac{\bar{x}_m}{n}.$$

Problem 16.
(a) Suppose we draw 2 cards out of a deck of 52 (bridge hand). Let X be a random variable that represents the number of hearts in the hand (so it takes only values 0, 1, 2). Determine the distribution of X (give probabilities that $X = k$ for $k = 0, 1, 2$).
(b) Compute the expectation and the variance of X.
(c) Let Y be a random variable that represents the number of spades in the hand. Determine the joint distribution of X and Y (give probabilities $\mathbb{P}(X = k, Y = j)$ for $k, j = 0, 1, 2$.
(d) Compute the covariance $\mathrm{Cov}(X, Y)$ and the correlation $\rho(X, Y)$ between the random variables X and Y.

Solution:
(a)

$$\mathbb{P}(X=0) = \frac{\binom{39}{2}}{\binom{52}{2}} = \frac{19}{34} = 0.558\ldots$$

$$\mathbb{P}(X=1) = \frac{\binom{39}{1}\cdot\binom{13}{1}}{\binom{52}{2}} = \frac{13}{34} = 0.382\ldots$$

$$\mathbb{P}(X=2) = \frac{\binom{13}{2}}{\binom{52}{2}} = \frac{1}{17} = 0.059\ldots$$

(b)

$$\mathbb{E}[X] = 0\cdot\mathbb{P}(X=0) + 1\cdot\mathbb{P}(X=1) + 2\cdot\mathbb{P}(X=2) = \frac{1}{2},$$

$$\mathbb{E}[X^2] = 0^2\cdot\mathbb{P}(X=0) + 1^2\cdot\mathbb{P}(X=1) + 2^2\cdot\mathbb{P}(X=2) = \frac{21}{34},$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{25}{68}.$$

(c)

$X\backslash Y$	0	1	2
0	$\frac{\binom{26}{2}}{\binom{52}{2}}$	$\frac{\binom{13}{1}\cdot\binom{26}{1}}{\binom{52}{2}}$	$\frac{\binom{13}{2}}{\binom{52}{2}}$
1	$\frac{\binom{13}{1}\cdot\binom{26}{1}}{\binom{52}{2}}$	$\frac{\binom{13}{1}\cdot\binom{13}{1}}{\binom{52}{2}}$	0
2	$\frac{\binom{13}{2}}{\binom{52}{2}}$	0	0

(d)

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{\binom{13}{1}\cdot\binom{13}{1}}{\binom{52}{2}} - \frac{1}{4} = -\frac{25}{204} \sim -0.122549.$$

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} = \frac{-\frac{25}{204}}{\frac{25}{68}} = -\frac{1}{3}.$$

Problem 17. Euromillions Lottery draws 5 main numbers out of 50. Thus the probability that a given main number is drawn in Euromillions lottery is $\frac{1}{10}$. Let $X(n)$ be a random variable that counts how many times a particular number (say 13) was drawn in n draws.
(a) What is the distribution of $X(n)$? Give $\mathbb{P}(X(n) = k)$, which represents the probability that number 13 was drawn k times in n draws.
(b) Use the Central Limit Theorem to find the two sided 95% confidence interval for $X(n)$. In other words, find a and b such that

$$\mathbb{P}(a \le X(n)) = 0.025, \qquad \mathbb{P}(X(n) \le b) = 0.025.$$

(c) Using the same approach, find the two sided 95% confidence interval for the variable

$$\frac{X(n)}{n}$$

that represents the frequency of the number 13.
(d) Evaluate the formulas for (b) and (c) for $n = 500$.

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$

$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
(a) $\mathrm{Bin}(n,p)$ with $p = \frac{1}{10}$.
(b)

$$n\cdot p \pm 1.96\sqrt{np(1-p)}.$$

(c)

$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}}.$$

(d)

$$50 \pm 1.96\sqrt{45} = 50 \pm 13.15 = (36.85, 63.15),$$
$$\frac{1}{10} \pm 1.96\sqrt{\frac{9}{50000}} = 0.1 \pm 0.0263 = (0.0737, 0.1263).$$

Problem 18. Consider again Euromillions lottery that chooses 5 main numbers out of 50 at each draw. Let T be a random variable that measures how many draws it takes to pick a specific number for the first time (say number 13). If 13 is drawn in the next draw, $T = 1$. If 13 is drawn after 5 draws and it is not picked earlier, $T = 5$.

- **(a)** What is the distribution of T? Give $\mathbb{P}(T = k)$.
- **(b)** What is $\mathbb{E}[T]$, the expected waiting time for the number to come out?
- **(c)** Compute the median and 95% quantile.

Solution:
(a) Geometric:

$$\mathbb{P}(T = k) = p \cdot (1-p)^{k-1}.$$

(b)

$$\mathbb{E}[T] = \frac{1}{p} = 10.$$

(c) $1 - \alpha$ quantile of the geometric distribution is given by the nearest integer above

$$\frac{\ln(\alpha)}{\ln(1-p)}$$

(see Problem 8). Therefore the median (50% quantile) is given by

$$\left\lceil \frac{\ln(0.5)}{\ln(0.9)} \right\rceil = \lceil 6.57\dots \rceil = 7,$$

the 95% quantile is given by

$$\left\lceil \frac{\ln(0.05)}{\ln(0.9)} \right\rceil = \lceil 28.43\dots \rceil = 29.$$

Problem 19. Let X be a random variable with density

$$f(x) = \frac{1}{x^2}, \qquad x \geq 1.$$

(a) Compute the expectation of X.
(b) Find the median of X.
(c) Find the 95% confidence interval in the form $[1, a]$. In other words, find a such that $\mathbb{P}(X \leq a) = 95\%$.

Solution:
(a)

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{1}^{\infty} \frac{1}{x}dx = \Big[\ln(x)\Big]_{x=1}^{x=\infty} = \infty.$$

(b) Median satisfies $F(x) = 0.5$, where F is the cumulative distribution function. We have

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{1}^{x} \frac{1}{y^2}dy = \left[-\frac{1}{y}\right]_{y=1}^{y=x} = 1 - \frac{1}{x}.$$

Solving for $F(x) = 0.5$, we get $x = 2$.
(c) 95% quantile satisfies $F(x) = 0.95$, which is satisfied for $x = 20$.

Problem 20. Consider random variables X and Y with the following observed values:

X	0	1	2
Y	2	0	1

(a) Compute correlation coefficient between X and Y.
(b) Compute the regression parameters of $Y = a_0 + a_1 \cdot X$.

Solution: (a) First, compute $\bar{X}$, $\bar{Y}$ and $\overline{X \cdot Y}$ (from the sample). We get

$$\bar{X} = \frac{0+1+2}{3} = 1, \quad \bar{Y} = \frac{2+0+1}{3} = 1,$$

and

$$\overline{X \cdot Y} = \frac{0 \cdot 2 + 1 \cdot 0 + 2 \cdot 1}{3} = \frac{2}{3}.$$

Thus

$$\mathrm{Cov}(X, Y) = \overline{X \cdot Y} - \bar{X} \cdot \bar{Y} = \frac{2}{3} - 1 \cdot 1 = -\frac{1}{3}.$$

The sample variances are

$$s_x^2 = \frac{1}{3} \cdot \Big((0-1)^2 + (1-1)^2 + (2-1)^2\Big) = \frac{2}{3},$$

$$s_y^2 = \frac{1}{4} \cdot \Big((2-1)^2 + (0-1)^2 + (1-1)^2\Big) = \frac{2}{3},$$

and thus the correlation coefficient equals to

$$r_{xy} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{-\frac{1}{3}}{\sqrt{\frac{2}{3} \cdot \frac{2}{3}}} = -\frac{1}{2}.$$

(b)

$$a_1 = r_{xy} \cdot \frac{s_y}{s_x} = -\frac{1}{2} \cdot \frac{\sqrt{\frac{2}{3}}}{\sqrt{\frac{2}{3}}} = -\frac{1}{2},$$
$$a_0 = \bar{Y} - a_1 \cdot \bar{X} = 1 - \left(-\frac{1}{2}\right) \cdot 1 = \frac{3}{2}.$$

Problem 21. Suppose we draw two cards out of a deck of 52 (bridge hand). Let X be a random variable that counts the number of aces in the first card (so it takes only values 0 and 1) and let Y be a random variable that represents the number of aces in the second card (it also takes only values 0 and 1).
(a) Determine $\mathbb{E}[X]$, $\mathrm{Var}(X)$, $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$.
(b) Determine the joint distribution of the random variables (X, Y), namely list the following probabilities:

$\mathbb{P}(X = 0, Y = 0)$	$\mathbb{P}(X = 0, Y = 1)$
$\mathbb{P}(X = 1, Y = 0)$	$\mathbb{P}(X = 1, Y = 1)$

(c) Are the random variables X and Y dependent or independent? Show.
(d) Compute the covariance $\mathrm{Cov}(X, Y)$ and the correlation $\rho(X, Y)$ between the random variables X and Y.

Solution:
(a) The random variable X takes values

$$X = \begin{cases} 1, & p = \frac{4}{52} = \frac{1}{13} \\ 0, & 1 - p = \frac{48}{52} = \frac{12}{13}. \end{cases}$$

The random variable Y has the same distribution. Therefore

$$\mathbb{E}[X] = \mathbb{E}[Y] = 1 \cdot p + 0 \cdot (1 - p) = p = \tfrac{1}{13},$$
$$\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1^2 \cdot p + 0^2 \cdot (1 - p) = p = \tfrac{1}{13},$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p) = \tfrac{12}{169}.$$

(b)

$\frac{48}{52} \cdot \frac{47}{51}$	$\frac{48}{52} \cdot \frac{4}{51}$
$\frac{4}{52} \cdot \frac{48}{51}$	$\frac{4}{52} \cdot \frac{3}{51}$

(c)

$$\mathbb{P}(Y = 1|X = 1) = \tfrac{3}{51} \neq \mathbb{P}(Y = 1) = \tfrac{4}{52}.$$

The variables X and Y are dependent.
(d)

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}[X \cdot Y] - E[X] \cdot E[Y] \\ &= \tfrac{4}{52} \cdot \tfrac{3}{51} - \left(\tfrac{1}{13}\right)^2 \\ &= -\tfrac{1}{13} \cdot \tfrac{4}{17 \cdot 13}.\end{aligned}$$

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = \frac{-\frac{1}{13} \cdot \frac{4}{17 \cdot 13}}{\frac{12}{169}} = -\frac{1}{51}.$$

Problem 22. An exceptional negative return that corresponds to the Value at Risk at 95% is expected once in 20 trading days. Let X denote a random variable that measures the number of such exceptional returns in one year (250 trading days).

(a) What is the distribution of the random variable X? Give a formula for

$$\mathbb{P}(X = k).$$

Hint: $X = \sum_{i=1}^{250} Y_i$, where

$$Y_i = \begin{cases} 1 & p = 0.05 \\ 0 & 1 - p = 0.95 \end{cases}$$

is the random variable that counts the excess negative return for each given day i.
(b) Determine the expectation μ and variance σ^2 of Y_i. What is the expecation and variance of X?
(c) Using the Central Limit Theorem, give the (one-sided) confidence intervals for X at the 95% and 99% levels. In other words, find the numbers a and b such that

$$\mathbb{P}(X \leq a) = 0.95$$

and

$$\mathbb{P}(X \leq b) = 0.99.$$

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$
$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
(a) The distribution of X is $Bin(250, \frac{1}{20})$:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{250}{k} \left(\frac{1}{20}\right)^k \left(\frac{19}{20}\right)^{250-k}.$$

(b)

$$Y = \begin{cases} 1, & p = 0.05, \\ 0, & 1 - p = 0.95, \end{cases}$$
$$\mu = \mathbb{E}[Y] = \tfrac{1}{20}, \qquad \sigma^2 = \mathrm{Var}(Y) = p \cdot (1-p) = \tfrac{19}{20^2},$$
$$\mathbb{E}[X] = n \cdot \mathbb{E}[Y] = \tfrac{250}{20} = 12.5,$$
$$\mathrm{Var}(X) = n \cdot \mathrm{Var}(Y) = \tfrac{250 \cdot 19}{20^2} = 11.875.$$

(c)

$$\begin{aligned} a &= n\mu + 1.64\sigma\sqrt{n} \\ &= 12.5 + 1.64 \cdot \tfrac{\sqrt{19}}{20} \cdot \sqrt{250} \\ &= 18.15\ldots \end{aligned}$$

The number of such exceptions X is 18 or less with 95% probability.

$$\begin{aligned} b &= n\mu + 2.33\sigma\sqrt{n} \\ &= 12.5 + 2.33 \cdot \tfrac{\sqrt{19}}{20} \cdot \sqrt{250} \\ &= 20.53\ldots \end{aligned}$$

The number of such exceptions X is 20 or less with 99% probability.

Problem 23. Consider Mega Millions lottery that chooses 5 main numbers out of 75 and 1 number out of 15.

(a) Compute the probability of winning the jackpot p (correctly guessing 5 numbers out of 75 and 1 number out of 15).

(b) A draw with a large jackpot attracts about 70,000,000 random tickets filed in a given draw. Use Poisson approximation to compute the probability that

- nobody wins the jackpot,
- exactly one wins the jackpot,
- two or more win the jackpot.

(c) How many random tickets are needed to be filed in order to have 95% probability that somebody wins the jackpot? How does that translate to the number of draws (assuming 70 million tickets per draw)?

Solution:
(a)

$$\binom{75}{5} \cdot \binom{15}{1} = 258,890,850 = \frac{1}{p}.$$

$$p = \frac{1}{258,890,850}.$$

(b)

$$\begin{aligned}
&\mathbb{P}(X = 0) \approx e^{-np} = 0.763\ldots, \\
&\mathbb{P}(X = 1) \approx (np) \cdot e^{-np} = 0.206\ldots, \\
&\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \approx 0.031\ldots
\end{aligned}$$

(c) We want to find n such that

$$\mathbb{P}(X = 0) \approx e^{-np} = 0.05.$$

Thus

$$n \approx \frac{-\ln(0.05)}{p} \approx 3p = 776,672,550.$$

This would take approximately

$$\frac{776,672,550}{70,000,000} = 11.09\ldots$$

draws to happen.

Problem 24. Let X be a random variable with a cumulative distribution function

$$F(x) = \frac{1}{1 + e^{-x}}.$$

(a) Check that $F(x)$ is indeed a cumulative distribution function (nondecreasing function with limits $F(-\infty) = 0$ and $F(\infty) = 1$). Checking for monotonicity requires computing derivative (nondecreasing function has a derivative greater or equal to zero). The derivative of the cumulative distribution function is the density function. Thus give the density function of this random

variable.
Hint: For computing $F'(x)$, one can use

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

with $f(x) = 1$ and $g(x) = 1 + e^{-x}$. (Note that the result in (a) is not needed for the remaining questions (b),(c) and (d)).
(b) Give a formula for the quantile q_α for this random variable X, $0 < \alpha < 1$.
Hint: $F(q_\alpha) = \alpha$.
(c) Find the median $m = q_{0.5}$.
(d) Find the 2.5% quantile $q_{0.025}$, the 97.5% quantile $q_{0.975}$ and give the two sided 95% confidence interval for the random variable X in the form $(q_{0.025}, q_{0.975})$.

Solution:
(a)

$$F(x) = \frac{1}{1 + e^{-x}},$$

$$F(-\infty) = 0, \qquad F(+\infty) = 1$$

and

$$f(x) = F'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \geq 0,$$

meaning the derivative of the cumulative distribution function (density) is nonnegative and thus the cumulative distribution function itself is nondecreasing.
(b) We want to find q_α such that $F(q_\alpha) = \alpha$ for $0 < \alpha < 1$. Equality

$$\frac{1}{1 + e^{-q_\alpha}} = \alpha$$

leads to

$$\frac{1}{\alpha} = 1 + e^{-q_\alpha}$$

and

$$e^{-q_\alpha} = \frac{1}{\alpha} - 1 = \frac{1 - \alpha}{\alpha}.$$

Thus

$$q_\alpha = \ln\left(\frac{\alpha}{1 - \alpha}\right).$$

(c)

$$m = q_{0.5} = \ln\left(\frac{0.5}{0.5}\right) = 0.$$

(d)

$$q_{0.025} = \ln\left(\frac{0.025}{0.975}\right) = -3.66\ldots,$$

$$q_{0.975} = \ln\left(\frac{0.975}{0.025}\right) = 3.66\ldots,$$

and thus the two sided 95% confidence interval $(q_{0.025}, q_{0.975})$ is

$$(-3.66\ldots, 3.66\ldots).$$

Problem 25. Consider random variables X and Y with the following observed values:

X	-1	0	1	2
Y	0	2	0	1

(a) Compute the estimates a and b of the regression parameters α and β from the linear regression model

$$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i.$$

(b) List the residuals $u_i = Y_i - (a + b \cdot X_i)$ and compute $\sum_{i=1}^{4} u_i$.
(c) Compute $\sum_{i=1}^{4} u_i \cdot x_i$

Solution:
(a)

$$b = \frac{\overline{X \cdot Y} - \bar{X}\bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{\frac{1}{2} - \frac{1}{2} \cdot \frac{3}{4}}{\frac{3}{2} - \left(\frac{1}{2}\right)^2} = \frac{1}{10},$$

$$a = \bar{Y} - b \cdot \bar{X} = \frac{3}{4} - \frac{1}{10} \cdot \frac{1}{2} = \frac{7}{10}.$$

(b) The residuals are given by $u_i = Y_i - (a + b \cdot X_i)$. Thus

$$u_1 = -\frac{6}{10}, \quad u_2 = \frac{13}{10}, \quad u_3 = -\frac{8}{10}, \quad u_4 = \frac{1}{10}.$$

This gives $\sum_{i=1}^{4} u_i = 0$ (as expected).
(c)

$$u_1 \cdot x_1 = \frac{6}{10}, \quad u_2 \cdot x_2 = 0, \quad u_3 \cdot x_3 = -\frac{8}{10}, \quad u_4 \cdot x_4 = \frac{2}{10}.$$

This gives $\sum_{i=1}^{4} u_i \cdot x_i = 0$ (as expected).

Problem 26. A die is rolled 600 times. Let random variable S_{600} counts the number of 6's you get in this sample.
(a) What is the distribution of the random variable S_{600}? Give a formula for

$$\mathbb{P}(S_{600} = k).$$

Hint: $S_{600} = \sum_{i=1}^{600} Y_i$, where

$$Y_i = \begin{cases} 1 & p = \frac{1}{6} \\ 0 & 1-p = \frac{5}{6} \end{cases}$$

is the random variable that counts the number of 6's for each roll i.
(b) Determine the expectation μ and variance σ^2 of Y_i. What is the expecation and variance of S_{600}?
(c) Using the Central Limit Theorem, give the two sided confidence interval for S_{600} at the 95% level. In other words, find the numbers a and b such that

$$\mathbb{P}(a \leq X \leq b) = 0.95.$$

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$
$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
(a) The distribution of S_{600} is $Bin(600, \frac{1}{6})$:

$$\mathbb{P}(S_{600} = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{600}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{600-k}.$$

(b)

$$Y = \begin{cases} 1, & p = \frac{1}{6}, \\ 0, & 1-p = \frac{5}{6}, \end{cases}$$
$$\mu = \mathbb{E}[Y] = \tfrac{1}{6}, \qquad \sigma^2 = \mathrm{Var}(Y) = p \cdot (1-p) = \tfrac{5}{36},$$
$$\mathbb{E}[X] = n \cdot \mathbb{E}[Y] = \tfrac{600}{6} = 100,$$
$$\mathrm{Var}(X) = n \cdot \mathrm{Var}(Y) = 600 \cdot \tfrac{5}{36} = \tfrac{500}{6}.$$

(c) The two sided 95% confidence interval is given by

$$n\mu \pm 1.96\sigma\sqrt{n} = 100 \pm 1.96 \cdot \tfrac{\sqrt{500}}{6} = 100 \pm 17.89 = (82.10, 117.89).$$

Problem 27. A pair of random variables (X, Y) takes 3 possible values, namely

X	Y	probability
0	0	$\frac{1}{2}$
0	1	$\frac{1}{4}$
1	0	$\frac{1}{4}$

(a) Determine $\mathbb{E}[X]$, $\mathrm{Var}(X)$, $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$.
(b) Are the random variables X and Y dependent or independent? Show.
(c) Compute the covariance $\mathrm{Cov}(X,Y)$ and the correlation $\rho(X,Y)$ between the random variables X and Y.

Solution:
Note that $X = X^2$, $Y = Y^2$ and $X \cdot Y = 0$.
(a) $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{1}{4}$, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{4} - (\frac{1}{4})^2 = \frac{3}{16}$. $\mathrm{Var}(Y) = \frac{1}{4}$.
(b) Dependent. Independence requires $\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k) \cdot \mathbb{P}(Y = l)$ for all possible k and l. However, $\frac{1}{2} = \mathbb{P}(X = 0, Y = 0) \neq \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 0) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$.
(c)

$$\mathrm{Cov}(X,Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = -\tfrac{1}{16}.$$

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = -\frac{1}{3}.$$

Problem 28. Consider Euro Jackpot lottery that chooses 5 main numbers out of 50 and 2 extra numbers out of 10.

(a) Compute the probability of winning the jackpot p (correctly guessing 5 numbers out of 50 and 2 numbers out of 10).

(b) A draw with a large jackpot attracts about $N = 20,000,000$ single lines filed in a given draw filed at random (with no specific pattern). Use Poisson approximation to compute the probability that

- nobody wins the jackpot,
- exactly one wins the jackpot,
- two or more win the jackpot.

(c) How many random tickets are needed to be filed in order to have 95% probability that somebody wins the jackpot? How does that translate to the number of draws (assuming 20 million tickets per draw)?

Solution:
(a)

$$p = \frac{1}{\binom{50}{5}\binom{10}{2}} = \frac{1}{95,344,200}.$$

(b) Let X be a number of jackpot wins. The random variable X has $Bin(N, p)$ distribution, which is approximately $Po(N \cdot p)$. Thus

$$\begin{aligned}
&\mathbb{P}(X = 0) \approx \exp(-N \cdot p) = 0.810774 \\
&\mathbb{P}(X = 1) \approx (N \cdot p) \cdot \exp(-N \cdot p) = 0.170073 \\
&\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \approx 0.0191533
\end{aligned}$$

(c) We want to find the smallest N such that $\mathbb{P}(X = 0) \leq 0.05$. This gives

$$\mathbb{P}(X = 0) \approx \exp(-N \cdot p) = 0.05,$$

and thus

$$N \approx -\frac{\ln(0.05)}{p} \approx \frac{3}{p} = 286,032,600.$$

One has to file about 286,032,600 tickets to guarantee 95% win of the jackpot.

Problem 29. This is a follow up question after Problem 2 (but it can be solved independently). Consider a specific Euro Jackpot draw that attracts N single lines filed at random.
(a) The lowest winning prize in the Euro Jackpot lottery is to guess correctly 2 main numbers and 1 extra number. Compute the probability p_{12} of winning this prize. (The subscript 12 indicates that this is the 12th prize as there are 11 higher ranked prizes).
(b) Denote by X the number of lines that win the prize from question (a). What is the expectation of X? Give a formula in terms of N and p_{12}.
(c) Using the Central Limit Theorem, give the two sided 95% confidence interval for X so that $\mathbb{P}(X \in (a, b)) = 95\%$. More specifically, find a and b such that

$$\mathbb{P}(X \leq a) = 0.025$$

and

$$\mathbb{P}(X \leq b) = 0.975.$$

First, give a confidence interval as a formula in terms of N and p_{12}. Second, evaluate that formula for $N = 20,000,000$.

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$

$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
(a)

$$p_{12} = \frac{\binom{5}{2}\binom{45}{3}\binom{2}{1}\binom{8}{1}}{\binom{50}{5}\binom{10}{2}} = \frac{2,270,400}{95,344,200} = 0.02381\ldots.$$

(b)

$$\mathbb{E}[X] = N \cdot p_{12}.$$

(c) The 95% confidence interval is given by

$$N \cdot p_{12} \pm 1.96\sqrt{N \cdot p_{12} \cdot (1 - p_{12})} = 476253 \pm 1336 = [474917, 477590].$$

Problem 30. Let X be a random variable with a cumulative distribution function

$$F(x) = \begin{cases} \frac{x^2}{x^2+1}, & x \geq 0, \\ 0 & x < 0. \end{cases}$$

(a) Check that $F(x)$ is indeed a cumulative distribution function (nondecreasing function with limits $F(-\infty) = 0$ and $F(\infty) = 1$). Checking for monotonicity requires computing derivative (nondecreasing function has a derivative greater or equal to zero). The derivative of the cumulative distribution function is the density function. Thus give the density function of this random variable.
Hint: For computing $F'(x)$, one can use

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

with $f(x) = x^2$ and $g(x) = x^2 + 1$ for $x \geq 0$. (Note that the result in (a) is not needed for the remaining questions (b),(c) and (d)).
(b) Give a formula for the quantile q_α for this random variable X, $0 < \alpha < 1$.
Hint: $F(q_\alpha) = \alpha$.
(c) Find the median $m = q_{0.5}$.
(d) Find the 95% quantile $q_{0.95}$ give the one sided 95% confidence interval for the random variable X in the form $[0, q_{0.95}]$.

Solution:
(a)

$$f(x) = F'(x) = \frac{2x}{(x^2+1)^2} > 0, \qquad x \geq 0.$$

(b) $F(q_\alpha) = \alpha$ leads to

$$\frac{q_\alpha^2}{q_\alpha^2 + 1} = \alpha,$$

and

$$q_\alpha = \sqrt{\frac{\alpha}{1-\alpha}}.$$

(c)

$$q_{0.5} = \sqrt{\frac{0.5}{0.5}} = 1.$$

(d)

$$q_{0.95} = \sqrt{\frac{0.95}{0.05}} = \sqrt{19}.$$

Problem 31.

Consider random variables X and Y with the following observed values:

X	0	0	1
Y	0	1	0

(a) Compute the estimates a and b of the regression parameters α and β from the linear regression model

$$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i.$$

(b) List the residuals $u_i = Y_i - (a + b \cdot X_i)$ and compute $\sum_{i=1}^{3} u_i$.
(c) Compute $\sum_{i=1}^{3} u_i \cdot x_i$.

Solution:
(a)

$$a = \tfrac{1}{2}, \qquad b = -\tfrac{1}{2}.$$

(b)

$$u_1 = -\tfrac{1}{2}, \qquad u_2 = \tfrac{1}{2}, \qquad u_3 = 0.$$

$$\sum_{i=1}^{3} u_i = 0.$$

(c)

$$\sum_{i=1}^{3} u_i \cdot x_i = -\tfrac{1}{2} \cdot 0 + \tfrac{1}{2} \cdot 0 + 1 \cdot 0 = 0.$$

Problem 32. A random variable X takes values $1, 2$, the random variable takes values $1, 2, 3$ and their joint distribution is given by the following table

	$\mathbb{P}(X,Y)$	Y 1	2	3
X	1	$\frac{1}{4}$	$\frac{1}{4}$	0
	2	0	$\frac{1}{4}$	$\frac{1}{4}$

(a) Determine $\mathbb{E}[X]$, $\mathrm{Var}(X)$, $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$.
(b) Are the random variables X and Y dependent or independent? Show.
(c) Compute the covariance $\mathrm{Cov}(X,Y)$ and the correlation $\rho(X,Y)$ between the random variables X and Y.

Solution:
(a)

$$\mathbb{E}[X] = 1.5, \qquad \mathrm{Var}(X) = 0.25, \qquad \mathbb{E}[Y] = 2, \qquad \mathrm{Var}(Y) = 0.5.$$

(b) Dependent. Enough to show that that for some pair (x,y)

$$\mathbb{P}(X = x, Y = y) \neq \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y).$$

Take for instance $(x,y) = (1,3)$. Then

$$0 = \mathbb{P}(X = 1, Y = 3) \neq \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 3) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}.$$

(c)

$$\mathrm{Cov}(X,Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = \frac{13}{4} - 3 = \frac{1}{4},$$
$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = \frac{\frac{1}{4}}{\sqrt{\frac{1}{4} \cdot \frac{1}{2}}} = \frac{1}{\sqrt{2}}.$$

Problem 33. A job candidate has probability of $p = \frac{1}{20}$ to get the job offer after an interview for the specific position. He/she applies to a new job opening if he/she fails to get the offer from the previous interview (with the same probability p of success), otherwise the job offer is accepted. Let T count the number of such job interviews.
(a) What is the distribution of the random variable T (list both the formula for $\mathbb{P}(T = k)$ and the name of the distribution).
(b) What is $\mathbb{E}[T]$?
(c) Compute the median (50% quantile) and the 95% quantile for the random variable T. (OK just to use the formula if you know it, otherwise go over the computation).

Solution:
(a) Geometric distribution:

$$\mathbb{P}(T = k) = p(1-p)^{k-1}, \qquad k = 1, 2, \ldots$$

with $p = \frac{1}{20}$.
(b)

$$\mathbb{E}[T] = \frac{1}{p} = 20.$$

(c)

$$q_\alpha = \left\lceil \frac{\log(1-\alpha)}{\log(1-p)} \right\rceil .$$

Median corresponds to $\alpha = 0.5$, the formula gives

$$q_{0.5} = \left\lceil \frac{\log(1-0.5)}{\log(1-0.05)} \right\rceil = 14,$$

the 95% quantile gives

$$q_{0.95} = \left\lceil \frac{\log(1-0.95)}{\log(1-0.05)} \right\rceil = 59.$$

Problem 34. The index SP500 has a (logarithmic) daily mean return X_i with mean $\mu = 0.00043$ and standard deviation $\sigma = 0.0084$. (These parameters are never fully observed, but for the sake of this problem, assume that μ and σ are known quantities.) Assume that one year has $n = 250$ trading days.
(a) Use the Central Limit Theorem to give the two sided 95% confidence interval for the cumulative return over n days:

$$S_n = \sum_{i=1}^{n} X_i.$$

The confidence interval should depend on n. In particular, evaluate this confidence interval for one year period ($n = 250$) and for five years period ($n = 1250$).
(b) Give the two sided 95% confidence interval for the (daily) average return over n days:

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

In particular, evaluate this confidence interval for one year period ($n = 250$) and for five years period ($n = 1250$).
(c) Give the 2.5% quantile $Q_{0.025}(n)$ and the 5% quantile $Q_{0.05}(n)$ for S_n as a function of n. (The 2.5% quantile should agree with the lower bound from part (a).) Find the smallest n such that $Q_{0.025}(n) \geq 0$. Similarly, find the smallest n such that $Q_{0.05}(n) \geq 0$. (Such n correspond to an investment horizon – the holding period for which a negative return has a negligent probability).

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$

$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution:
(a) 95% two sided confidence interval for the sum is

$$\mu n \pm 1.96\sigma\sqrt{n}.$$

For one year $n = 250$, we get

$$0.00043 \cdot 250 \pm 1.96 \cdot 0.0084 \cdot \sqrt{250} = 0.1075 \pm 0.2603 = [-0.1528, 0.3678].$$

For 5 years ($n = 1250$), we obtain

$$0.00043 \cdot 1250 \pm 1.96 \cdot 0.0084 \cdot \sqrt{1250} = 0.5375 \pm 0.5821 = [-0.0446, 1.1196].$$

(b) 95% two sided confidence interval for the average is

$$\mu \pm 1.96\frac{\sigma}{\sqrt{n}}.$$

For one year $n = 250$, we get

$$0.00043 \pm 1.96 \cdot \frac{0.0084}{\sqrt{250}} = 0.00043 \pm 0.00104 = [-0.00061, 0.00147].$$

For 5 years ($n = 1250$), we obtain

$$0.00043 \cdot 1250 \pm 1.96 \cdot \frac{0.0084}{\sqrt{1250}} = 0.00043 \pm 0.00047 = [-0.00004, 0.00090].$$

(c) The quantiles are given by the formulas

$$Q_{0.025}(n) = \mu n - 1.96\sigma\sqrt{n}$$

and

$$Q_{0.05}(n) = \mu n - 1.64\sigma\sqrt{n}$$

respectively. Solving for n for which $Q_{0.025}(n) = 0$, we get

$$\mu n - 1.96\sigma\sqrt{n} = 0,$$

which gives

$$n = 1.96^2 \left(\frac{\sigma}{\mu}\right)^2 = 1466.$$

Similarly, solving for n for which $Q_{0.05}(n) = 0$ gives

$$\mu n - 1.64\sigma\sqrt{n} = 0,$$

leading to

$$n = 1.64^2 \left(\frac{\sigma}{\mu}\right)^2 = 1027.$$

Problem 35. A test is graded on the scale 0 to 1, with 0.55 needed to pass. Student scores are modeled by the following density:

$$f(x) = \begin{cases} 4x & 0 \le x \le \frac{1}{2} \\ 4 - 4x & \frac{1}{2} \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the probability that a random student passes the exam?
(b) What score is the 90 percentile (quantile $q_{0.9}$) of the distribution?

Solution:
(a) We need to calculate

$$\mathbb{P}(X \ge 0.55) = 1 - \mathbb{P}(X \le 0.55) = 1 - F(x),$$

where $F(x)$ is the cumulative distribution function of the random variable X which describes the distibution of the grade scores. From the relationship of F to the density function given by

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(y)dy,$$

we get

$$F(x) = \begin{cases} 2x^2 & 0 \le x \le \frac{1}{2}, \\ -1 + 4x - 2x^2 & \frac{1}{2} \le x \le 1. \end{cases}$$

When $x \in [0, \frac{1}{2}]$, $F(x)$ is simply

$$F(x) = \int_0^x 4y dy = 2x^2.$$

When $x \in [\frac{1}{2}, 1]$, we have

$$F(x) = \int_0^{1/2} f(y)dy + \int_{1/2}^{x} f(y)dy = \frac{1}{2} + \int_{1/2}^{x} (4 - 4y)dy = \frac{1}{2} + [4y - 2y^2]_{y=\frac{1}{2}}^{y=x}$$
$$= \frac{1}{2} + 4x - 2x^2 - 4 \cdot \frac{1}{2} + 2 \cdot \left(\frac{1}{2}\right)^2 = -1 + 4x - 2x^2.$$

Now $F(0.55) = 0.595$, so $\mathbb{P}(X \geq 0.55) = 1 - F(0.55) = 0.405$.
(b) For 90% quantile $q_{0.9}$, one must solve for

$$F(q_{0.9}) = 0.9,$$

which means

$$-1 + 4q_{0.9} - 2q_{0.9}^2 = 0.9.$$

This is valid only on the interval $[\frac{1}{2}, 1]$ as $F(x) = 1$ for $x \geq 1$ (no quadratic formula there). We have a quadratic equation with roots

$$x_{1,2} = 1 \pm \sqrt{0.05} \approx 1 \pm 0.223607 = \begin{cases} 0.77639, \\ 1.22361. \end{cases}$$

Only the smaller root is relevant (the quantile must be smaller than 1), so

$$q_{0.9} = 1 - \sqrt{0.05} \approx 1 - 0.223607 = 0.77639.$$

Problem 36. Consider random variables X and Y with the following observed values:

X	Y
1	2
2	1
3	1
4	0

(a) Compute the estimates a and b of the regression parameters α and β from the linear regression model

$$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i.$$

(b) List the residuals $u_i = Y_i - (a + b \cdot X_i)$ and compute $\sum_{i=1}^4 u_i$.
(c) Compute $\sum_{i=1}^4 u_i \cdot x_i$

Solution: (a)

$$a = 2.5, \qquad b = -0.6.$$

(b) Residuals:

$$\begin{aligned} u_1 &= Y_1 - (a + b \cdot X_1) = 2 - 1.9 = 0.1, \\ u_2 &= Y_2 - (a + b \cdot X_2) = 1 - 1.3 = -0.3, \\ u_3 &= Y_3 - (a + b \cdot X_3) = 1 - 0.7 = 0.3, \\ u_4 &= Y_4 - (a + b \cdot X_4) = 0 - 0.1 = -0.1. \end{aligned}$$

$$\sum_{i=1}^{4} u_i = 0.1 - 0.3 + 0.3 + 0.1 = 0.$$

(c)

$$\sum_{i=1}^{4} u_i \cdot x_i = 0.1 \cdot 1 + (-0.3) \cdot 2 + 0.3 \cdot 3 + (-0.1) \cdot 4 = 0.$$

Problem 37. A random variable X takes values $1, 2$, the random variable takes values $1, 2, 3$ and their joint distribution is given by the following table

	$\mathbb{P}(X,Y)$	Y 1	2	3
X	1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(a) Determine marginal distributions $\mathbb{P}(X = x)$ and $\mathbb{P}(Y = y)$.
(b) Determine conditional distributions $\mathbb{P}(X = x|Y = y)$ for $Y = 2$ and $\mathbb{P}(Y = y|X = x)$ for $X = 1$.
(c) Determine $\mathbb{E}[X]$, $\mathrm{Var}(X)$, $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$.
(d) Give the joint distribution of (X, Y) that is independent and has the same marginal distributions. Based on this, determine whether the random variables (X, Y) with the original joint distribution given in the table are dependent or independent.
(e) Compute the covariance $\mathrm{Cov}(X, Y)$ and the correlation $\rho(X, Y)$ between the random variables X and Y.

Solution: (a) Marginals are

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \frac{1}{2},$$

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 2) = \mathbb{P}(Y = 3) = \frac{1}{3}.$$

(b)

$$\mathbb{P}(X = 1|Y = 2) = \frac{\mathbb{P}(X = 1, Y = 2)}{\mathbb{P}(Y = 2)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}.$$

$$\mathbb{P}(X = 2|Y = 2) = \frac{\mathbb{P}(X = 2, Y = 2)}{\mathbb{P}(Y = 2)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}.$$

$$\mathbb{P}(Y = 1|X = 1) = \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(X = 1)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

$$\mathbb{P}(Y=2|X=1) = \frac{\mathbb{P}(X=1, Y=2)}{\mathbb{P}(X=1)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

$$\mathbb{P}(Y=3|X=1) = \frac{\mathbb{P}(X=1, Y=3)}{\mathbb{P}(X=1)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

(c)

$$\mathbb{E}[X] = 1.5, \qquad \mathrm{Var}(X) = 0.25.$$

$$\mathbb{E}[Y] = 2, \qquad \mathrm{Var}(Y) = \frac{2}{3}.$$

(d)

	$\mathbb{P}(X,Y)$	Y 1	2	3
X	1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The random variables X and Y are independent.

(e)

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 3 - 1.5 \cdot 2 = 0,$$

$$\rho(X,Y) = 0.$$

Problem 38. Consider a betting strategy in roulette that bets one dollar on red. It has been shown in the lecture that the random variable X that represents the number of times when the profit/loss is equal to zero has geometric distribution with parameter $p = \frac{1}{37}$.
(a) What is $\mathbb{E}[T]$? (Enough to list the value).
(b) Compute the median (50% quantile) and the 95% quantile for the random variable T. (Fine just to use the formula if you know it, otherwise go over the computation).

Solution: (a)

$$\mathbb{E}[T] = \frac{1}{p} = 37.$$

(b)

$$q_\alpha = \left\lceil \frac{\log(1-\alpha)}{\log(1-p)} \right\rceil.$$

$$q_{0.5} = \left\lceil \frac{\log(1-0.5)}{\log(1-\frac{1}{37})} \right\rceil = 26,$$

$$q_{0.95} = \left\lceil \frac{\log(1-0.95)}{\log(1-\frac{1}{37})} \right\rceil = 110,$$

Problem 39. Consider that we are performing a Monte Carlo simulation that generates n different scenarios for the profit/loss distribution of the betting strategy from Problem 38. In particular, we generate n random variables X_i representing the number of visits to zero that has geometric distribution with parameter $p = \frac{1}{37}$.

(a) Using the Central Limit Theorem, give a two sided 95% confidence interval for the average

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

This should be a function of n. Evaluate this confidence interval for the choice of $n = 250,000$.

(b) Imagine the situation when a prescribed precision is needed for the above confidence interval. The confidence interval has a form $[a(n), b(n)]$ (both the left end and the right end points are functions of n), and we want that

$$b(n) - a(n) = C,$$

where C is a given constant. Note that the term $b(n) - a(n)$ represents the exact length of the confidence interval, and it is a decreasing function of n. Compute n that satisfies the above equality. In particular, give the exact n for the choice of $C = 0.2$, i.e., when the confidence interval has length 0.2.

You may use some of the following quantiles of the normal distribution:

$$q_{0.01} \approx -2.33, \qquad q_{0.025} \approx -1.96, \qquad q_{0.05} = -1.64,$$

$$q_{0.99} \approx 2.33, \qquad q_{0.975} \approx 1.96, \qquad q_{0.95} = 1.64.$$

Solution: (a) 95% two sided confidence interval for $\bar{X}$:

$$\mu \pm 1.96 \frac{\sigma}{\sqrt{n}}.$$

Since $X \sim Geo(\frac{1}{37})$, we have $\mu = 37$ and $\sigma^2 = \frac{1-p}{p^2} = 36 \cdot 37 = 1332$. For $n = 250,000$, we have

$$\mu \pm 1.96 \frac{\sigma}{\sqrt{n}} = 37 \pm 1.96 \sqrt{\frac{1332}{250000}} = 37 \pm 0.14306657471261411.$$

(b) The length of a confidence interval equals to

$$b(n) - a(n) = \left(\mu + 1.96 \frac{\sigma}{\sqrt{n}}\right) - \left(\mu - 1.96 \frac{\sigma}{\sqrt{n}}\right) = 3.92 \frac{\sigma}{\sqrt{n}}.$$

We want to solve for n in

$$3.92 \frac{\sigma}{\sqrt{n}} = C.$$

This gives

$$n = 3.92^2 \cdot \frac{\sigma^2}{C^2}.$$

For our numbers,

$$n = 3.92^2 \cdot \frac{1332}{0.2^2} = 3.92^2 \cdot 1332 \cdot 25 = 511,701.$$

Problem 40. Consider a random variable with a cumulative distribution function

$$F(x) = 1 - \exp\left(-\tfrac{x^2}{2}\right), \qquad x \geq 0.$$

(a) Give the corresponding density $f(x)$ of this distribution.
(b) Compute α quantiles for any $\alpha \in (0, 1)$.
(c) Give the median (50% quantile $q_{0.5}$) and the 95% quantile $q_{0.95}$.

Solution: (a)

$$f(x) = F'(x) = x \exp\left(-\tfrac{x^2}{2}\right), \qquad x \geq 0.$$

(b) Solve

$$F(q_\alpha) = \alpha.$$

This gives

$$1 - \exp\left(-\tfrac{q_\alpha^2}{2}\right) = \alpha,$$

or

$$q_\alpha = \sqrt{-2\ln(1-\alpha)}.$$

(c)

$$q_{0.5} = \sqrt{-2\ln(1-0.5)} = 1.1774100225154747.$$

$$q_{0.95} = \sqrt{-2\ln(1-0.95)} = 2.4477468306808161.$$

Problem 41. Consider random variables X and Y with the following observed values:

X	Y
1	0
2	2
3	2
4	4

(a) Compute the estimates a and b of the regression parameters α and β from the linear regression model

$$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i.$$

(b) List the residuals $u_i = Y_i - (a + b \cdot X_i)$ and compute $\sum_{i=1}^4 u_i$.
(c) Compute $\sum_{i=1}^4 u_i \cdot x_i$.

Solution: (a)

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	u_i	$u_i \cdot x_i$
1	0	$-\frac{3}{2}$	-2	3	2.25	-0.2	-0.2
2	2	$-\frac{1}{2}$	0	0	0.25	0.6	1.2
3	2	$\frac{1}{2}$	0	0	0.25	-0.6	-1.8
4	4	$\frac{3}{2}$	2	3	2.25	0.2	0.8

$$b = \frac{\sum_{i=1}^4 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^4 (x_i - \bar{x})^2} = \frac{6}{5} = 1.2,$$
$$a = \bar{y} - b\bar{x} = 2 - 1.2 \cdot 2.5 = -1.$$

(b)

$$u = \begin{pmatrix} -0.2 \\ +0.6 \\ -0.6 \\ +0.2 \end{pmatrix}$$

(c)

$$\sum_{i=1}^{4} u_i = 0, \qquad \sum_{i=1}^{4} u_i x_i = 0$$