Discrete Random Variable

anis.rezguii@gmail.com

INSAT
Mathematics Department

February 10, 2022

Table of contents

1. Expected learning outcomes
2. Definitions
   - The Mass Distribution Function (MDF)
   - The cumulative distribution function CDF
3. Parameters of a d.r.v
   - The mathematical expectation
   - The variance and the standard deviation
4. Usual probability distributions
   - The uniform distribution
   - The Binomial distribution
   - The Hypergeometric distribution
   - The Geometric distribution
   - The Poisson distribution

Expected learning outcomes

After this lecture the student should:

- Understand the concept of discrete random variable (d.r.v).
- Understand the intrinsic Mass Distribution Function (MDF) associated to a d.r.v
- Know usual examples of MDF.
- Chose the right model for a given random situation.

Definition

Definition

Let $(\Omega, \mathbb{P})$ *be a probability space describing a random experiment and* X *a mapping*

$$\begin{array}{rccl} X: & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & X(\omega). \end{array}$$

X *is said to be a discrete random variable (d.r.v) if its range,* $X(\Omega)$ *is a discrete subset of* $\mathbb{R}$*.*

Discrete subset of $\mathbb{R}$

A discrete subset of $\mathbb{R}$ is any subset which contains only isolated points.

- the subset $\{0, 1, 2, 3\}$ is a discrete subset of $\mathbb{R}$.
- any finite subset is a discrete set.
- the subset of all integers $\mathbb{N}$ is a discrete set.
- the set of rational numbers $\mathbb{Q}$ is not a discrete set.
- Any interval of $\mathbb{R}$ is not a discrete subset of $\mathbb{R}$.

Example of d.r.v

Consider the experiment of tossing two perfect coins. The associated sample space is $\Omega = \{H, T\}^2 = \{HH, TT, HT, TH\}$ where "H" denotes the "head" face and "T" denotes the "tail" face. Let X be the number of heads obtained after each toss.

It is clear that its values set is $X(\Omega) = \{0, 1, 2\}$ so X is a d.r.v.

The d.r.v X can be represented as follows:

Ω	HH	TT	HT	TH
X	2	0	1	1

The Mass Distribution Function (MDF)

Definition

We associate to a given d.r.v, X, its The Mass Distribution Function (MDF) denoted by $\mathbb{P}_X$ and defined by the mapping:

$$\begin{array}{rlcl} \mathbb{P}_X : & X(\Omega) & \longrightarrow & [0,1] \\ & x_i & \longmapsto & \mathbb{P}_X(x_i) = \mathbb{P}\{X = x_i\} \end{array}$$

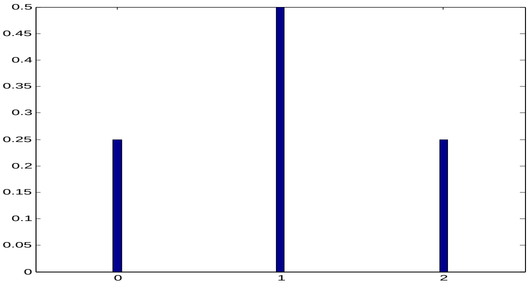
where $\{X = x_i\} = \{\omega \in \Omega : \ X(\omega) = x_i\}$

Table representation of the MDF

In general we represent the MDF of a d.r.v X, $\mathbb{P}_X$, by the following table:

$X(\Omega)$	$\mathbb{P}_X$
x_1	$\mathbb{P}_X(x_i)$
$\vdots$	$\vdots$

Graphical representation of the MDF



Example

We reconsider the same example of tossing two perfect coins and where X was the number of heads. To get its MDF $\mathbb{P}_X$ we need to evaluate:

- $\mathbb{P}_X(0) = \mathbb{P}\{X = 0\} = \mathbb{P}\{TT\} = \mathbb{P}\{T\} \times \mathbb{P}\{T\} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, note that we have used the independence between the two tosses.
- $\mathbb{P}_X(1) = \mathbb{P}\{X = 1\} = \mathbb{P}\{TH, HT\} = \mathbb{P}\{TH\} + \mathbb{P}\{HT\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.
- $\mathbb{P}_X(2) = \mathbb{P}\{X = 2\} = \mathbb{P}\{HH\} = \mathbb{P}\{H\} \times \mathbb{P}\{H\} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

Example

We reconsider the same example of tossing two perfect coins and where X was the number of heads. To get its MDF $\mathbb{P}_X$ we need to evaluate:

- $\mathbb{P}_X(0) = \mathbb{P}\{X = 0\} = \mathbb{P}\{TT\} = \mathbb{P}\{T\} \times \mathbb{P}\{T\} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, note that we have used the independence between the two tosses.
- $\mathbb{P}_X(1) = \mathbb{P}\{X = 1\} = \mathbb{P}\{TH, HT\} = \mathbb{P}\{TH\} + \mathbb{P}\{HT\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.
- $\mathbb{P}_X(2) = \mathbb{P}\{X = 2\} = \mathbb{P}\{HH\} = \mathbb{P}\{H\} \times \mathbb{P}\{H\} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

$\mathbb{P}_X$ is given by $\rightsquigarrow$

$X(\Omega)$	$\mathbb{P}_X$
0	0.25
1	0.5
2	0.25

Notes

1. Note that $\sum_i \mathbb{P}_X(x_i) = \sum_i \mathbb{P}\{X = x_i\} = \mathbb{P}(\Omega) = 1.$

Notes

1. Note that $\sum_i \mathbb{P}_X(x_i) = \sum_i \mathbb{P}\{X = x_i\} = \mathbb{P}(\Omega) = 1.$
2. Note the similarity between a discrete random variable and a discrete statistical variable:

$\mathbf{X(\Omega)}$	$\mathbb{P}_\mathbf{X}$
x_1	$\mathbb{P}_X(x_i)$
$\vdots$	$\vdots$
$\sum$	1

$\mathbf{x_i}$	$\mathbf{f_i}$
x_1	f_1
$\vdots$	$\vdots$
$\sum$	100%

Notes

1. Note that $\sum_i \mathbb{P}_X(x_i) = \sum_i \mathbb{P}\{X = x_i\} = \mathbb{P}(\Omega) = 1.$
2. Note the similarity between a discrete random variable and a discrete statistical variable:

$\mathbf{X(\Omega)}$	$\mathbb{P}_{\mathbf{X}}$
x_1	$\mathbb{P}_X(x_i)$
$\vdots$	$\vdots$
$\sum$	1

$\mathbf{x_i}$	$\mathbf{f_i}$
x_1	f_1
$\vdots$	$\vdots$
$\sum$	100%

3. The latter similarity is used to model a statistical phenomenon by a theoretical d.r.v and then to generate artificial data by computer simulation without any cost. This is very important for prediction and making decision.

The cumulative distribution function CDF

Definition

To any discrete random variable X we could associate its cumulative distribution function defined as follows:

$$\begin{array}{rcl} F_X \quad : \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & F_X(x) = \mathbb{P}\{X \leq x\} \end{array}$$

where $\{X \leq x\} = \{\omega \in \Omega : \; X(\omega) \leq x\}.$

Example

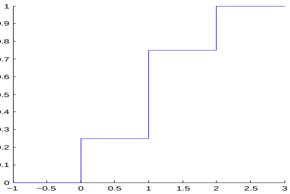
Again, we reconsider the same example of tossing two perfect coins where X was the number of heads, the associated cumulative distribution function is as follows:

1. if $x < 0$ then $\mathbb{P}\{X \leq x\} = \mathbb{P}\{\emptyset\} = 0$
2. if $0 \leq x < 1$ then $\mathbb{P}\{X \leq x\} = \mathbb{P}\{X = 0\} = 0.25$
3. if $1 \leq x < 2$ then $\mathbb{P}\{X \leq x\} = \mathbb{P}\{X = 0 \text{ or } X = 1\} = 0.25 + 0.5 = 0.75$
4. if $x \geq 2$ then $\mathbb{P}\{X \leq x\} = \mathbb{P}\{X = 0 \text{ or } X = 1 \text{ or} X = 2\} = 0.25 + 0.5 + 0.25 = 1$

The cumulative distribution function is summarized as follows:

$$F_X = \begin{cases} 0 & \text{if } x < 0 \\ 0.25 & \text{if } 0 \leq x < 1 \\ 0.75 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

Graphical representation of a CDF



Note 1

Let X a given d.r.v and denote by $\mathbb{P}_X$, F_X its MDF and its CDF respectively. The CDF F_X satisfies:

i. F_X is non-decreasing.

ii. F_X is càdlag i.e continuous to the right and bounded to the left.

iii. $\lim\limits_{x \to -\infty} F_X(x) = 0$ and $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Note 2

It is very important to know that $\mathbb{P}_X$ and F_X are equivalent in the sense that the knowledge of one suffices to determine the other. If we have $\mathbb{P}_X$ it is clear that we can determine F_X, just by definition. Now suppose we have the cumulative distribution function, F_X, we use the following rule to determine $\mathbb{P}_X$:

$$\mathbb{P}\{X = a\} = F_X(a) - F_X(a^-).$$

where $F_X(a^-) = \lim_{x \to a^-} F_X(x)$.

The Mathematical expectation

Definition

Let X be a d.r.v and $X(\Omega) = \{x_i\}$ its set of values.
The mathematical expectation of X, denoted by $\mathbb{E}(X)$, is the quantity:

$$\mathbb{E}(X) = \sum_i x_i \mathbb{P}_X(x_i).$$

Example

Reconsider the last example where X was the number of heads, we get:

$$\begin{aligned}\mathbb{E}(X) &= 0\times\frac{1}{4}+1\times\frac{1}{2}+2\times\frac{1}{4}\\ &= \frac{1}{2}+\frac{1}{2}=1\end{aligned}$$

Example

Reconsider the last example where X was the number of heads, we get:

$$\begin{aligned} \mathbb{E}(X) &= 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

Note: If any random experiment is repeated many times we expect, in the "usual" cases, that most of X's values are "close" to the mathematical expectation $\mathbb{E}(X)$. Thus the mathematical expectation of a d.r.v is considered as a parameter of central tendency.

Exercise

Let X be a d.r.v with three possible values $\{-1, 0, 1\}$ such that its MDF is as follow:

$X(\Omega)$	$\mathbb{P}_X$
-1	q
0	$1/2$
1	p

and its expectation $\mathbb{E}(X) = 0$.

1. Determine the MDF of X, $\mathbb{P}_X$.
2. Suppose now $\mathbb{P}\{X = 0\}$ is also unknown, could you determine $\mathbb{P}_X$?

Solution

1. On one hand $\mathbb{E}(X) = -1 \times p + 0 \times \frac{1}{2} + 1 \times q = 0$ which means $p = q$, on the other hand $p + \frac{1}{2} + q = 1$, then we get $p = q = \frac{1}{4}$.
2. We obtain two equations with three unknowns, so there is infinitely many possibilities.

Solution

1. On one hand $\mathbb{E}(X) = -1 \times p + 0 \times \frac{1}{2} + 1 \times q = 0$ which means $p = q$, on the other hand $p + \frac{1}{2} + q = 1$, then we get $p = q = \frac{1}{4}$.
2. We obtain two equations with three unknowns, so there is infinitely many possibilities.

Note:

In general the only knowledge of the expectation does not determine the MDF of the random variable. However there are many interesting particular cases, of random variables, where it is the case, see section usual MDF.

The Markov Inequality

The next theorem gives a rough estimation of some probabilities by the only use of the expectation.

Theorem

Let $(\Omega, \Sigma, \mathbb{P})$ *be a probability space,* X *a positive d.r.v and* λ *a positive real number. Then*

$$\mathbb{P}\{X \geq \lambda\} \leq \frac{\mathbb{E}(X)}{\lambda}.$$

The variance

Definition

Let X be a d.r.v and $X(\Omega) = \{x_i\}$ its range. Its variance, denoted by $\mathbb{V}(X)$ or σ_X^2, is the non negative quantity:

$$\mathbb{V}(X) = \sigma_X^2 = \sum_i (x_i - \mathbb{E}(X))^2 \mathbb{P}_X(x_i).$$

Its standard deviation is

$$\sigma_X = \sqrt{\mathbb{V}(X)}.$$

Notes

1. Suppose that a given random experiment is repeated many times, we expect, in the usual cases, that most of X's values lies in the interval $[\mathbb{E}(X) - \sigma_X, \mathbb{E}(X) + \sigma_X]$.

2. The variance as well as the standard deviation are parameters of dispersion. In fact they give idea about how expected data are dispersed around the mathematical expectation.

The Tchebychev Inequality

Theorem

Let $(\Omega, \Sigma, \mathbb{P})$ *be a probability space, and* X *a d.r.v, then for any* $\lambda > 0$ *we have*

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq \lambda\} \leq \frac{\mathbb{V}(X)}{\lambda^2}.$$

The Tchebychev Inequality

Theorem

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, and X a d.r.v, then for any $\lambda > 0$ we have

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq \lambda\} \leq \frac{\mathbb{V}(X)}{\lambda^2}.$$

Proof:

Let $\lambda > 0$

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq \lambda\} = \mathbb{P}\{|X - \mathbb{E}(X)|^2 \geq \lambda^2\}$$

and so by using Markov inequality, we get

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq \lambda\} \leq \frac{\mathbb{E}\{|X - \mathbb{E}(X)|^2\}}{\lambda^2} = \frac{\mathbb{V}(X)}{\lambda^2},$$

which finishes the proof.

Note

Tchebychev inequality is often stated as follow:

$$\mathbb{P}\{|X - \mathbb{E}(X)| \geq n\sigma\} \leq \frac{1}{n^2}$$

which can be read

$$\mathbb{P}\left\{X \in [\mathbb{E}(X) - n\sigma, \mathbb{E}(X) + n\sigma]\right\} > 1 - \frac{1}{n^2}$$

If for example we take $n = 6$ we obtain that for any random variable X, which means for any phenomenon we have the following estimation:

$$\mathbb{P}\left\{X \in [\mathbb{E}(X) - 6\sigma, \mathbb{E}(X) + 6\sigma]\right\} > 1 - \frac{1}{6^2} = 97.22\%$$

.

Note

We have a different, but, easy to prove and useful different formulation of the variance: For any discrete random variable X:

$$\begin{aligned}\mathbb{V}(X) &= \sum_i \Big(x_i - \mathbb{E}(X)\Big)^2 \mathbb{P}\{X = x_i\} \\ &= \sum_i \Big(x_i^2 - 2\mathbb{E}(X)x_i + \mathbb{E}(X)^2\Big)\mathbb{P}\{X = x_i\} \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2.\end{aligned}$$

The uniform distribution

Definition

Let X *be a d.r.v with a finite range,* $X(\Omega) = \{x_1, \cdots, x_n\}$*, for some* $n \in \mathbb{N}$*. We say that* X *follows the uniform distribution on* $\{x_1, \cdots, x_n\}$ *and we denote*

$$X \hookrightarrow \mathcal{U}\{x_1, \cdots, x_n\}$$

if

$$\mathbb{P}\{X = x_1\} = \cdots = \mathbb{P}\{X = x_n\} = \frac{1}{n}.$$

Examples

1. If we toss a perfect coin and set $X = 0$ for "Tail" and $X = 1$ for "Head". It is easy to check that $X \hookrightarrow \mathcal{U}\{0, 1\}$.
2. If we throw a perfect die and set X as the obtained digit. It is easy to check that $X \hookrightarrow \mathcal{U}\{1, \cdots, 6\}$.

Examples

1. If we toss a perfect coin and set $X = 0$ for "Tail" and $X = 1$ for "Head". It is easy to check that
$X \hookrightarrow \mathcal{U}\{0, 1\}$.
2. If we throw a perfect die and set X as the obtained digit. It is easy to check that
$X \hookrightarrow \mathcal{U}\{1, \cdots, 6\}$.

Note:

In case of uniform distribution we have the following formula:

$$\mathbb{P}\{X \in A\} = \frac{\#A}{n},$$

for any $A \subset X(\Omega)$.

Expectation and variance

Let $X \hookrightarrow \mathcal{U}\{x_1, \cdots, x_n\}$, we can check easily that:

i.
$$\mathbb{E}(X) = \frac{x_1 + \cdots + x_n}{n}.$$

ii.
$$\mathbb{V}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2.$$

Expectation and variance

Let $X \hookrightarrow \mathcal{U}\{x_1, \cdots, x_n\}$, we can check easily that:

i.
$$\mathbb{E}(X) = \frac{x_1 + \cdots + x_n}{n}.$$

ii.
$$\mathbb{V}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\Big(\sum_{i=1}^{n} x_i\Big)^2.$$

Note: Note the similarity with the sample mean and the sample variance of any set of data.

The Binomial distribution

Definition

Let X be a d.r.v, $n \in \mathbb{N}$ and $p \in]0,1[$. X is said to follow a binomial distribution with n repetitions and a "success" probability p, or just with parameters (n,p), if:

i) $X(\Omega) = \{0, 1, \cdots, n\}$,

ii) *for any* $k = 0, 1, \cdots, n$, $\mathbb{P}\{X = k\} = C_n^k \, p^k (1-p)^{n-k}$.

We denote

$$X \hookrightarrow \beta(n,p).$$

Typical Example

Consider the experiment of throwing a perfect die 10 times and X the number of occurrences of the digit "6". We can check easily that

$$X \hookrightarrow \beta(10, \frac{1}{6}).$$

Checking

– It is clear that $X(\Omega) = \{0, 1, \cdots, 10\}$.

– The event
$\{X = 0\} = \Big\{\text{"6" doesn't appear}\Big\} = \Big\{10 \text{ times } \{1, \cdots, 5\}\Big\}$,
this implies that:

$$\mathbb{P}\{X = 0\} = \mathbb{P}\{1, \cdots, 5\} \times \cdots \times \mathbb{P}\{1, \cdots, 5\} = \frac{5}{6} \times \cdots \times \frac{5}{6} = (\frac{5}{6})^{10}.$$

– The event $\{X = 1\} = \Big\{\text{"6" appears only once}\Big\}$, then

$$\mathbb{P}\{X = 1\} = {\color{red}10} \times \mathbb{P}\{X = 6\} \times (\mathbb{P}\{1, \cdots, 5\})^9 = 10 \times \frac{1}{6} \times (\frac{5}{6})^9,$$

where 10 in the formula denotes the 10 possible positions of occurrence of the digit "6".

When we use the binomial model?

In general we use the binomial distribution when we repeat a given random experiment n times, independently, and when X denotes the number of occurrences of a particular event called "success or failure" with a given probability $p \in]0, 1[$. In that case we have

$$X \hookrightarrow \beta(n, p).$$

Expectation and variance

Let X be a d.r.v following a binomial distribution $\beta(n, p)$ for $n \in \mathbb{N}^*$ and $p \in]0, 1[$:

i. its mathematical expectation is:

$$\mathbb{E}(X) = n \times p,$$

ii. its variance is:

$$\mathbb{V}(X) = n \times p \times (1 - p).$$

The Hypergeometric distribution

Definition

Let N_1, N_2 and n three integers, such that $N_1 + N_2 > n$, and X a given discrete random variable. It is said to follow a hypergeometric distribution of parameters (N_1, N_2, n), and denoted $X \hookrightarrow \mathcal{HG}(N_1, N_2, n)$ if

i) $X(\Omega) = \{0, 1, \cdots, n\}$.

ii) *for any* $k = 0, \cdots, n$

$$\mathbb{P}\{X = k\} = \frac{C_{N_1}^{k} C_{N_2}^{n-k}}{C_{N_1+N_2}^{n}}.$$

Typical example

X is the number of red balls drawn after n successive draws of one ball, without replacement, from a box that contains N_1 red balls and N_2 white balls. Check that X follows a hypergeometric distribution, $X \hookrightarrow \mathcal{HG}(N_1, N_2, n)$.

Expectation and Variance

On can check that

$$\mathbb{E}(X) = np$$

and

$$\mathbb{V}(X) = np(1-p)\frac{N-n}{N-1}$$

where $p = \dfrac{N_1}{N}$ and $N = N_1 + N_2$.

Definition

Let $p \in]0,1[$ *and* X *a discrete random variable.* X *is said to follow a geometric distribution of parameter* p, *and is denoted* $X \hookrightarrow \mathcal{G}(p)$, *if*

i) $X(\Omega) = \mathbb{N}^*$.

ii) *For any* $k \geq 1$,

$$\mathbb{P}\{X = k\} = p(1-p)^{k-1}.$$

Typical example

Let X be the position of the first occurrence of the face "head" when we toss a balanced coin infinitely many times. Check that X follows a geometric distribution of parameter $\frac{1}{2}$.

The expectation and the variance

It is easy to show that

$$\mathbb{E}(X) = \frac{1}{p}$$

and

$$\mathbb{V}(X) = \frac{1-p}{p^2}.$$

The Poisson distribution

Definition

Let $\lambda \in \mathbb{R}_+^*$ *and* X *a d.r.v, we say that* X *follows a Poisson distribution with parameter* λ, *and we denote* $X \hookrightarrow \mathcal{P}(\lambda)$, *if*

i) $X(\Omega) = \mathbb{N}$.

ii) *for any* $k \geq 0$,

$$\mathbb{P}\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Typical example

- In general, the Poisson distribution is used to model the number of clients (customers) which arrive during a time period T, with a frequency f, to ask for a given service.
- In that case their random number can be modeled by a Poisson distribution of parameter the average number of clients $\lambda = T \times f$.

Expectation and variance

Let $\lambda \in \mathbb{R}_+^*$ and $X \hookrightarrow \mathcal{P}(\lambda)$. We prove that:

i.
$$\mathbb{E}(X) = \lambda.$$

ii.
$$\mathbb{V}(X) = \lambda.$$

Exercise

Consider the records of highway accidents between Makkah and AL Madinah during 50 days.

accidents number	number of days
0	21
1	18
2	7
3	3
4	1

1. Compute the average number of accident, m.

Still the exercise

2. Suppose that the theoretical probability distribution of the accidents number is of Poisson type with parameter λ. Determine an approximation of λ and compute the probability that no accident happen, only one accident happen, 2 accident ...
3. Compute the theoretical number of days t_i such that we observe i accidents for $i = 0, \cdots, 4$. Compare theoretical and empirical results.
4. How many days without accidents we will observe during a year.

Solution

1. The average m is just the total number of accidents divided by the number of days.

$$m = \frac{1 \times 18 + 2 \times 7 + 3 \times 3 + 4 \times 1}{50} = 0,9.$$

2. Let $X \hookrightarrow \mathcal{P}(\lambda)$, if the latter is a model for the accidents number we must have $\mathbb{E}(X) = m$. Thus we have $\lambda = m = 0,9$. We obtain the probabilities that happen i accidents during 50 days for $i = 0, \cdots, 4$ by using the formula $\mathbb{P}\{X = i\} = e^{-0,9}\frac{(0,9)^i}{i!}$. And so

X	0	1	2	3	4
$\mathbb{P}_X$	$40,6\%$	$36,5\%$	$16,4\%$	$4,9\%$	$1,1\%$

Still solution

3. The theoretical number of days where there was exactly i accidents is $t_i = \mathbb{P}_X(i) \times 50.$

i	0	1	2	3	4
t_i	20,328	18,295	8,233	2,469	0,555

It is clear that the theoretical frequency distribution is very close to the experimental (statistical) one and so, we have a good reason to trust the Poisson model.

4. Without waiting for a year, we can estimate the number of days without any accidents

$$\text{accidents number per year} = 365 \times \mathbb{P}_X(0) = 148,397.$$