PROBABILITIES

Contents

Context

A computer tool, very often, faces uncertain situations. For example, imagine that its purpose is to answer user requests, but that neither the frequency of these requests nor their complexity are known during the tool design. They may be estimated, but the context is uncertain. In such a situation, a probability model is required.

Furthermore, the random variables involved in this model may have an infinite number of possible values. If all the natural numbers are possible values, we have to determine a probability for each $n \in \mathbb{N}$. This defines the probability distribution of the random variable. It can be used in the same way as in the finite case, but the sums involved in the computations are now infinite sums. For example, the expectation and the variance are limits of numerical series.

In practice, working with the whole distribution would require an infinite amount of memory storage: it would be an infinite array of real values, with each cell containing one probability. We hence need to condense the information. The expectation and the variance can be useful for this purpose. The probability-generating function can be another efficient solution.

Many situations require infinite natural random variables. For example:

- number of requests to a server in a given time interval;

- number of times an experiment should be carried out until it succeeds;

- complexity of an algorithm, number of elementary operations required by its execution.

Learning outcomes

The skills that you are expected to acquire are listed below. You should check that:

- you understand what the generating function of a random variable is and how it is connected to the distribution;

- you can handle random variables taking an infinite number of integer values;

- you can connect some typical real-life situations to well-known distributions.

Specifically, for each section:

Generating function of a finite integer-valued random variable

- understand the connection between the distribution and the generating function;

- be able to switch from «distribution reasoning» to «generating function reasoning» in order to solve a probability problem.

Power series

- know and handle the reference power series;

- be able to determine the radius of convergence of a power series;

- be aware of the reasoning mistakes which can result from misusing properties specific to power series.

Infinite integer-valued random variables

- extend to infinite cases the basic reasoning, which you have seen before about finite variables;

- use generating series to answer probability questions;

- connect real-life situations to well-known infinite distributions.

1 Generating function of a finite integer-valued random variable

1.1 Summary

When several random variables are combined, it is difficult to determine the distribution of the resulting variable using pure probability reasoning. The probability-generating function can help to solve this question: when all the variables take natural values, we just need to do basic polynomial operations.

The generating function of a random variable contains all the information displayed by the distribution. This information is just differently encoded. Thus, it should be handled differently.

1.2 Exercises

Exercise 1.1
Let X be the result of rolling a fair dice.

1. What is the generating function G_X of the variable X?

2. Explain why $G_X(1) = 1$.

3. Deduce $E(X)$ and $Var(X)$.

Exercise 1.2
Let X be a natural random variable whose generating function is

$$G_X(t) = a(2t + 1)^2$$

1. What is the value of a?

2. Determine the distribution of X.

3. Determine its expectation and its variance.

Exercise 1.3
Let $(p_x, p_y) \in [0, 1]^2$, let X and Y be two Bernoulli variables of parameters p_x and p_y, that is, they take values in $\{0, 1\}$ with
$$P(X=1) = p_x \quad \text{and} \quad P(Y=1) = p_y$$

We assume that these variables are independent.

1. Determine the generating functions G_X and G_Y.

2. Determine the generating function G_{X+Y} of the variable $X + Y$.

3. Can we find values of p_x and p_y such that $X + Y$ is uniformly distributed on $\{0, 1, 2\}$?

Exercise 1.4
Let $n \in \mathbb{N}^*$, $p \in [0, 1]$, and let $X_1, X_2, \cdots, X_n$ be n independent Bernoulli variables of parameter p. Let $Y = X_1 + \cdots + X_n$. Remember that Y is binomial-distributed with parameters n and p:

$$Y \rightsquigarrow B(n, p)$$

1. Determine the generating function G_{X_i} of the variable X_i, for $i \in [\![1, n]\!]$.

2. Determine the generating function G_Y of the variable Y.

3. Deduce the distribution of Y, then $E(Y)$ and $Var(Y)$.

2   Power series

2.1   Summary

When a random variable can take an infinite number of natural values, its generating function is not a polynomial. It is instead a power series, that is to say, a «polynomial of infinite degree».

Before using power series in a probability framework, it is necessary to understand what they are and to handle them.

2.2   Exercises

Exercise 2.5

1. Consider the power series $\sum x^n$.

   (a) Determine its radius of convergence $r \in \mathbb{R}^+ \cup \{+\infty\}$.

   (b) Let f be the function defined for any $x \in \,]-r, r[$ by $f(x) = \sum_{n=0}^{+\infty} x^n$.

   Express $f(x)$ as a rational fraction.

   (c) Deduce an expression for $g(x) = \sum_{n=0}^{+\infty} (n+1) x^n$ and for $h(x) = \sum_{n=1}^{+\infty} \frac{x^n}{n}$.

2. Consider the power series $\sum \frac{x^n}{n!}$.

   (a) Determine its radius of convergence $r \in \mathbb{R}^+ \cup \{+\infty\}$.

   (b) Let f be the function defined for any $x \in \,]-r, r[$ by $f(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$.

   Find a simple relation between $f'(x)$ and $f(x)$.

   (c) Deduce an expression of $f(x)$ in term of usual function(s).

Exercise 2.6

Let $\sum f_n$ be the series whose general terms is defined by:

$$\forall x \in [0, 2], \quad f_n(x) = \frac{x^{n+1}}{1 + x^{n+1}} - \frac{x^n}{1 + x^n}$$

1. Check that the function f_n is continuous and differentiable.

2. Show that for any $x \in [0, 2]$, the numerical series $\sum f_n(x)$ converges.

   Give an explicit formula for the function $F : x \longmapsto \sum_{n=0}^{+\infty} f_n(x)$.

3. Is the function F continuous on $[0, 2]$? Differentiable?

Exercise 2.7

1. What is the power series expansion $\sum a_n x^n$ of the function $f : x \longmapsto \dfrac{1}{1-x}$? Its radius of convergence r?

2. For any $x \in \,]-r, r[$, let us define the numerical sequence $(R_n(x))$ as the remainder of the numerical series $\sum a_n x^n$:

$$R_n(x) = f(x) - \sum_{k=0}^{n} a_k x^k = \sum_{k=n+1}^{+\infty} a_k x^k$$

Thus, we can write

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + R_n(x)$$

    a. Let $x \in \,]-r, r[$. Can we say that $R_n(x) = o\,(x^n)$ when n tends to $+\infty$?

    b. Let $n \in \mathbb{N}$. Can we say that $R_n(x) = o\,(x^n)$ when x tends to 0?

Exercise 2.8

Determine the radius of convergence of the following power series:

1. $\sum \dfrac{2^n}{n!} x^n$

2. $\sum \dfrac{(n!)^2}{(2n)!} x^n$

3. $\sum (-1)^n n x^n$

Exercise 2.9

For each of the following series, provide an explicit compact expression of the sum and determine the radius of convergence:

1. $\displaystyle\sum_{n=1}^{+\infty} \dfrac{x^n}{n}$

2. $\displaystyle\sum_{n=1}^{+\infty} n x^n$

3. $\displaystyle\sum_{n=0}^{+\infty} n^2 x^n$

4. $\displaystyle\sum_{n=0}^{+\infty} \dfrac{(-1)^n}{n!} x^{n+1}$

Exercise 2.10

Determine the power series expansion (in 0), with the radius of convergence, of the following functions:

1. $\dfrac{1}{(1-x)^2}$

2. $\dfrac{1}{(1-x)^r}$ where $r \in \mathbb{N}$

3. $x\, e^{2x}$

4. **(Bonus)** $\dfrac{\ln(1+x)}{1+x}$ then deduce the power series expansion of $\left(\ln(1+x)\right)^2$

    Express the coefficients a_n and b_n of the latter series as finite sums. Don't try to find explicit values.

3 Infinite integer-valued random variables

3.1 Summary

Random variables from real-life situations can often take an infinite number of values. A typical case is when these possible values are natural numbers.
For example:

- the number of requests to a server, in a given time interval, takes natural values. Those cannot be *a priori* bounded;

- the same situation occurs about the number of times a random experiment must be conducted until it is successful.

Such variables are handled in the same way as finite variables, but the involved sums are infinite sums. They are the limits of some numerical series.

3.2 Exercises

Exercise 3.11 *geometric distribution*
A hacker sends phishing emails to random people. His message urges the receiver to display his Visa card number. When he sends an email, the probability that the receiver is trapped has the value p. We suppose furthermore that the answers of the receivers are mutually independent.
Let X be the number of emails that the hacker needs to send until he gets a first Visa card number.

1. What is the distribution of X? Draw its graph.

2. Compute $P(X \in [\![3, 6]\!])$ and $P(X \geqslant 5)$.

3. Determine the generating function G_X of X.

4. Deduce its expectation and its variance.

Remark: we say that X is «geometric-distributed with parameter p». This is denoted by

$$X \rightsquigarrow \text{Geom}(p)$$

Exercise 3.12 *Pascal distribution*
Let $r \in \mathbb{N}^*$. Consider the hacker from the previous exercise and define Y as the number of emails he must send until he gets r Visa card numbers.

1. Express Y as a sum of geometric variables.

2. Determine the generating function G_Y of Y.

3. Deduce the distribution of Y.

4. Determine its expectation and its variance.

Exercise 3.13 *Poisson distribution*
The road safety service has observed that, on average, there are two accidents every 15 minutes. Let X be the number of accidents in the next 15 minutes.
  We suppose that, at each time t, the date of the next accident depends neither on the previous number of accidents nor on their dates. We accept without proof that, under this hypothesis, there exists $\lambda \in \mathbb{R}_+^*$ such that

$$\forall n \in \mathbb{N}, P(X=n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

1. Draw the graph of the X distribution.

2. Determine its generating function.

3. Deduce its expectation and its variance.

4. Using the observations from the road safety service, determine λ and deduce the probability of the events:

   a. no accident during the next 15 minutes;

   b. at least 3 accidents during the next 15 minutes;

   c. no accident during the next hour.

Remark: we say that X is «Poisson-distributed with parameter λ». This is denoted by

$$X \rightsquigarrow \text{Poisson}(\lambda)$$

Exercise 3.14
Let $(\lambda, \mu) \in \left(\mathbb{R}_+^*\right)^2$, let X and Y be two independent random variables such that

$$X \rightsquigarrow \text{Poisson}(\lambda) \qquad \text{and} \qquad Y \rightsquigarrow \text{Poisson}(\mu)$$

1. What is the generating functions of $X + Y$?

2. Deduce the distribution of $X + Y$.

3. In the situation of the previous exercise, determine the probability of the events:

   a. two accidents during the next 30 minutes;

   b. two accidents during the next hour.

★ **Exercice 3.15**
Let $\lambda \in \mathbb{R}_+^*$ and let Y be a random variable such that $Y \rightsquigarrow \text{Poisson}(\lambda)$.

For any $n \in \mathbb{N}$, consider a binomial variable $X_n \rightsquigarrow \text{B}\left(n, \dfrac{\lambda}{n}\right)$.

1. Display the generating functions G_Y and G_{X_n} of Y and X_n.

2. Let $t \in \mathbb{R}$. Show that $\lim\limits_{n \to +\infty} G_{X_n}(t) = G_Y(t)$.

Remark: we accept without proof that this property implies that «the sequence (X_n) converges in distribution to the variable Y», that is to say,

$$\forall A \subset \mathbb{N}, \quad P(X_n \in A) \xrightarrow[n \to +\infty]{} P(Y \in A)$$