These are lecture notes for an introductory course on probability taught as part of the first year's program at École des Ponts ParisTech. This course is currently under the direction of Aurélien Alfonsi (CERMICS) and based on the following textbook:

B. Jourdain, Probabilités et statistique, Ellipses 2009, 2nd edition 2016.

Chapter 2

Discrete random variables

2.1 Probability measure

From now on, we consider a random experiment which may have infinitely many possible outcomes. The set of all possible outcomes of this experiment is still called *sample space* and denoted Ω. In the present lesson we will call *event* any subset of Ω. For the sake of accuracy, let us mention that this definition actually raises an issue when Ω is uncountably infinite. The power set $\mathcal{P}(\Omega)$ is indeed proven to be "too big" to be considered as the set of events in many natural settings. One of the main troubles is the possible non-existence of a satisfying probability measure on $\mathcal{P}(\Omega)$. Therefore, we must restrict the set of events to a subset of $\mathcal{P}(\Omega)$. Such a set must be compatible with the usual operations on sets, namely the complement, union and intersection, which leads to the definition of a σ-algebra.

Definition 2.1.1 (σ-algebra)**.** *A* σ-algebra *on a set Ω is a class $\mathcal{A}$ of subsets of Ω ($\mathcal{A} \subset \mathcal{P}(\Omega)$) such that*

(i) $\Omega \in \mathcal{A}$ *;*

(ii) If $A \in \mathcal{A}$, then $A^{\complement} \in \mathcal{A}$;

(iii) If for all $n \in \mathbb{N}$, $A_n \in \mathcal{A}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

We then say that $(\Omega, \mathcal{A})$ is a measurable space.

For the sake of accuracy again, let us mention that a map between two measurable spaces is in general not compatible with their respective underlying σ-algebras, hence the following definition.

Definition 2.1.2 (Measurable map)**.** *Let* $(E, \mathcal{E})$ *and* $(F, \mathcal{F})$ *be two measurable spaces. We call* measurable *map between* $(E, \mathcal{E})$ *and* $(F, \mathcal{F})$ *any map* $f : E \to F$ *such that*

$$\forall B \in \mathcal{E}, \quad f^{-1}(B) = \{x \in A \mid f(x) \in B\} \in \mathcal{E}.$$

From now on, we will consider that the sample space Ω is endowed with a σ-algebra $\mathcal{A}$ which is seen as the set of events. One should then consider in order to be accurate that $A \subset \Omega$ is an event iff $A \in \mathcal{A}$. However, all the σ-algebras we will encounter in the present lesson will always be big enough to contain all the subsets we will consider. Therefore, we will still call event any subset of Ω. For similar reasons, we will purposely never worry about the measurability of a function, as all the functions encountered in the present lessons will be measurable. Definitions 2.1.1 and 2.1.2 can therefore be ignored at first reading.

Definition 2.1.3 (Probability measure)**.** *A* probability measure *on a sample space* Ω *is a map* $\mathbb{P}$ *from the set of events to* $[0, 1]$ *such that*

(i) $\mathbb{P}(\Omega) = 1$ *;*

(ii) If $(A_i)_{i \in I}$ *is an at most countable disjoint family of events, then*

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i).$$

If $\mathcal{A}$ *denotes the class of events, then we say that* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space.*

Remark 2.1.4. A map $\mathbb{P}$ which satisfies (ii) is called σ-additive.

Throughout the rest of the present chapter and the next ones and unless explicitly mentioned otherwise, $(\Omega, \mathcal{A}, \mathbb{P})$ will always refer to a probability space such that

- Ω is the sample space we work on;
- $\mathcal{A}$ is the class of events (which can be harmlessly considered to be the power set $\mathcal{P}(\Omega)$);
- $\mathbb{P}$ is the probability measure which measures the likelihood of each event.

Except for the uniform distribution, the definitions and propositions given in the previous chapter for a finite sample space hold for a general sample space, including the notions of conditional probability, independence and Bayes' theorem.

2.2 Discrete random variables

2.2.1 Definition

We introduce here the fundamental notion of *random variable*, which is an expression whose value depends on the outcome of a random experiment. In this chapter we only consider random variables which have finitely or countably infinitely many different possible values.

Definition 2.2.1 (Discrete random variable)**.** *A* discrete random variable *is a (measurable) map $X : \Omega \to E$ where E is an at most countable set.*

For any event A, we denote

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}.$$

The family $(\mathbb{P}(\{X = x\}))_{x \in E}$ is called the probability distribution *of X.*

Example 2.2.2. Let A be an event. We recall that the indicator function of A is defined by

$$\forall \omega \in \Omega, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{else} \end{cases}.$$

Therefore, $\mathbb{1}_A : \Omega \to \{0, 1\}$ is a discrete random variable.

Definition 2.2.3 (Equality in distribution)**.** *Let E be an at most countable set and $X : \Omega \to E$ and $Y : \Omega \to E$ be two discrete random variables. We say that X and Y are equal in distribution if the probability distribution of X is equal to the probability distribution of Y, that is*

$$\forall z \in E, \quad \mathbb{P}(\{X = z\}) = \mathbb{P}(\{Y = z\}).$$

In that case, we denote $X \stackrel{d}{=} Y$.

2.2.2 Independence

Definition 2.2.4 (Independence)**.** *Let E and F be two at most countable sets. Two random variables $X : \Omega \to E$ and $Y : \Omega \to F$ are said to be* independent *if*

$$\forall (x, y) \in E \times F, \quad \mathbb{P}(\{X = x, Y = y\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\}).$$

In that case, we denote $X \perp\!\!\!\perp Y$.

We already saw in the previous chapter a notion of independence which concerns events. Those two notions coincide in the following sense: the event A is independent of the event B iff the random variable $\mathbb{1}_A$ is independent of the random variable $\mathbb{1}_A$.

Definition 2.2.5 (Independence)**.** *Let $n \in \mathbb{N}^*$. For $1 \leq k \leq n$, let E_k be an at most countable set and $X_k : \Omega \to E_k$ be a discrete random variable. The family of discrete random variables $(X_k)_{1 \leq k \leq n}$ is said to be* mutually independent *if*

$$\forall (x_1, \cdots, x_n) \in E_1 \times \cdots \times E_n, \quad \mathbb{P}(\{X_1 = x_1, \cdots, X_n = x_n\}) = \mathbb{P}(\{X_1 = x_1\}) \times \cdots \times \mathbb{P}(\{X_n = x_n\}).$$

We say that a family $(X_i)_{i \in I}$ of discrete random variables is mutually independent *if any finite subfamily of $(X_i)_{i \in I}$ is mutually independent.*

We say that a family $(X_i)_{i \in I}$ of discrete random variables is pairwise independent *if for all $i, j \in I$ such that $i \neq j$, X_i is independent of X_j.*

Remark 2.2.6. If $(X_i)_{i\in I}$ is mutually independent, then it is pairwise independent, but the converse in not true.

Definition 2.2.7 (i.i.d.)**.** *A family of discrete random variables is called* independent and identically distributed*, usually abbreviated* i.i.d.*, if this family is mutually independent and all its elements are equal in distribution.*

2.2.3 Common discrete probability distributions

We present here the most common discrete probability distributions.

2.2.3.1 The degenerate univariate distribution

Definition 2.2.8. *Let $\alpha \in \mathbb{R}$. We say that a discrete random variable X follows the degenerate univariate distribution with parameter α if*

$$\mathbb{P}(\{X = \alpha\}) = 1.$$

In that case, we denote $X \sim \delta_\alpha$.

The degenerate univariate distribution is the distribution of an almost constant discrete random variable. It is a way to see a deterministic variable as a particular case of random variable.

2.2.3.2 The Bernoulli distribution

Definition 2.2.9. *Let $p \in [0, 1]$. We say that a discrete random variable X follows the Bernoulli distribution with parameter p if*

$$\mathbb{P}(\{X = 1\}) = p \quad \textit{and} \quad \mathbb{P}(\{X = 0\}) = 1 - p.$$

In that case, we denote $X \sim \mathcal{B}(p)$.

Remark 2.2.10.

- For any event A, $\mathbb{1}_A \sim \mathcal{B}(\mathbb{P}(A))$.
- X follows the Bernoulli distribution with parameter p iff

$$\forall x \in \{0, 1\}, \quad \mathbb{P}(\{X = x\}) = p^x(1 - p)^{1-x}.$$

- $\mathcal{B}(0) = \delta_0$ and $\mathcal{B}(1) = \delta_1$.

The Bernoulli distribution models a random experiment which has two possible outcomes: head or tail, true or false, yes or no, etc. This kind of experiment is called a Bernoulli trial. When it makes sense, we usually interpret the event $\{X = 1\}$ as the success of the experiment and $\{X = 0\}$ as the failure. Therefore, p is often interpreted as the probability of success of a Bernoulli trial.

2.2.3.3 The binomial distribution

Definition 2.2.11. *Let $n \in \mathbb{N}^*$ and $p \in [0,1]$. We say that a discrete random variable X follows the binomial distribution with parameters n and p if*

$$\forall 0 \leq k \leq n, \quad \mathbb{P}(\{X = k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

In that case, we denote $X \sim \mathcal{B}(n,p)$.

Remark 2.2.12. $\mathcal{B}(1,p) = \mathcal{B}(p)$.

Proposition 2.2.13. *Let $n \in \mathbb{N}^*$ and $p \in [0,1]$. Let $X : \Omega \to \mathbb{N}$ be a discrete random variable and $X_1, \cdots, X_n$ be i.i.d. random variables, each having a Bernoulli distribution with parameter p. Then*

$$X \sim \mathcal{B}(n,p) \iff X \stackrel{d}{=} X_1 + \cdots + X_n.$$

Proof. Let us show that $X_1 + \cdots + X_n \sim \mathcal{B}(n,p)$. Let $k \in \{0, \cdots, n\}$. By σ-additivity, mutual independence of $(X_i)_{1 \leq i \leq n}$ and Remark 2.2.10, we have

$$\begin{aligned}
\mathbb{P}(\{X_1 + \cdots + X_n = k\}) &= \mathbb{P}\left(\bigcup_{\substack{x_1, \cdots, x_n \in \{0,1\} \\ x_1 + \cdots + x_n = k}} \{X_1 = x_1, \cdots, X_n = x_n\} \right) \\
&= \sum_{\substack{x_1, \cdots, x_n \in \{0,1\} \\ x_1 + \cdots + x_n = k}} \mathbb{P}(\{X_1 = x_1, \cdots, X_n = x_n\}) \\
&= \sum_{\substack{x_1, \cdots, x_n \in \{0,1\} \\ x_1 + \cdots + x_n = k}} \mathbb{P}(\{X_1 = x_1\}) \times \cdots \times \mathbb{P}(\{X_n = x_n\}) \\
&= \sum_{\substack{x_1, \cdots, x_n \in \{0,1\} \\ x_1 + \cdots + x_n = k}} p^{x_1}(1-p)^{1-x_1} \times \cdots \times p^{x_n}(1-p)^{1-x_n} \\
&= \sum_{\substack{x_1, \cdots, x_n \in \{0,1\} \\ x_1 + \cdots + x_n = k}} p^k (1-p)^{n-k} \\
&= C \times p^k (1-p)^{n-k},
\end{aligned}$$

where C is the cardinality of the set $\{(x_1, \cdots, x_n) \in \{0,1\}^n \mid x_1 + \cdots + x_n = k\}$. Choosing $x_1, \cdots, x_n \in \{0,1\}$ such that $x_1 + \cdots + x_n = k$ is equivalent to assigning 1 to k components of a vector of $\{0,1\}^n$ and 0 to the $n-k$ other components. The latter is itselft equivalent to choosing k elements among n. We deduce that $C = \binom{n}{k}$. This proves that $X_1 + \cdots + X_n$ follows a binomial distribution with parameters n and p. Therefore, $X \sim \mathcal{B}(n,p)$ iff X is equal in distribution to $X_1 + \cdots + X_n$. □

The best understanding of the binomial distribution is given by Proposition 2.2.13. For $X \sim \mathcal{B}(n,p)$, $\mathbb{P}(X = k)$ is the probability that exactly k successes occur in n independent Bernoulli trials.

2.2.3.4 The Poisson distribution

Definition 2.2.14. *Let $\lambda > 0$. We say that a discrete random variable X follows the Poisson distribution with parameter λ if*

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

In that case, we denote $X \sim \mathcal{P}(\lambda)$.

A random variable X which describes the number of events which happen during a certain time interval is typically modeled by a Poisson distribution.

2.2.3.5 The geometric distribution

Definition 2.2.15. *Let $p \in (0, 1]$. We say that a discrete random variable X follows the geometric distribution with parameter p if*

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P}(\{X = n\}) = p(1-p)^{n-1}.$$

In that case, we denote $X \sim \mathcal{G}eo(p)$.

Remark 2.2.16. $\mathcal{G}eo(1) = \delta_1$.

Proposition 2.2.17. *Let $p \in (0, 1]$. Let $(X_n)_{n \in \mathbb{N}^*}$ be a family of i.i.d. random variables, each having a Bernoulli distribution with parameter p. Then*

$$X \sim \mathcal{G}eo(p) \iff X \overset{d}{=} \inf\{n \in \mathbb{N}^* \mid X_n = 1\}.$$

Proof. Let $G = \inf\{n \in \mathbb{N}^* \mid X_n = 1\}$. Let us show that $G \sim \mathcal{G}eo(p)$. Let $n \in \mathbb{N}^*$. By definition of the infimum and mutual independence of $(X_k)_{k \in \mathbb{N}^*}$, we have

$$\begin{aligned}
\mathbb{P}(\{G = n\}) &= \mathbb{P}(\{X_1 = 0, \cdots, X_{n-1} = 0, X_n = 1\}) \\
&= \mathbb{P}(\{X_1 = 0\}) \times \cdots \times \mathbb{P}(\{X_{n-1} = 0\}) \times \mathbb{P}(\{X_n = 1\}) \\
&= (1-p)^{n-1} \times p,
\end{aligned}$$

so $G \sim \mathcal{G}eo(p)$. Therefore, $X \sim \mathcal{G}eo(p)$ iff $X \overset{d}{=} G$. □

Proposition 2.2.17 gives us a better understanding of the Geometric distribution. For $X \sim \mathcal{G}eo(p)$, $\mathbb{P}(X = n)$ is the probability that exactly n attempts are needed to witness the first success in a series of independent Bernoulli trials.

2.2.4 Marginal distribution

Let E and F be two at most countable sets. Let $X : \Omega \to E$ and $Y : \Omega \to F$ be two discrete random variables. Since $E \times F$ is at most countable, the map

$$(X,Y) \; : \; \begin{array}{ccc} \Omega & \to & E \times F \\ \omega & \mapsto & (X(\omega), Y(\omega)) \end{array}$$

is a discrete random variable. The probability distribution of X (resp. Y) is called *first* (resp *second*) *marginal distribution* of (X,Y). The probability distribution of (X,Y) is called the *joint probability distribution* for X and Y.

For all $x \in E$, by σ-additivity, we have

$$\mathbb{P}(\{X = x\}) = \mathbb{P}\left(\bigcup_{y \in F} \{X = x, Y = y\}\right) = \sum_{y \in F} \mathbb{P}(\{X = x, Y = y\}).$$

Similarly, for all $y \in F$, $\mathbb{P}(\{Y = y\}) = \sum_{x \in E} \mathbb{P}(\{X = x, Y = y\})$. We deduce that the marginal distributions can be deduced from the joint probability distribution. However, the converse is in general not true. Nevertheless let us mention that in the particular case of independence between X and Y, the joint probability distribution can be deduced from the marginal distributions. Indeed, if X is independent of Y, then for all $(x,y) \in E \times F$, $\mathbb{P}(\{X = x, Y = y\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\})$.

Exercise 2.2.18. Let E and F be two at most countable sets. Let $X : \Omega \to E$ and $Y : \Omega \to F$ be two discrete random variables such that there exist $c \in \mathbb{R}$, $\mu : E \to \mathbb{R}$ and $\nu : F \to \mathbb{R}$ which satisfy

$$\forall (x,y) \in E \times F, \quad \mathbb{P}(\{X = x, Y = y\}) = c\mu(x)\nu(y).$$

1. Compute c.

2. What can we say about X and Y?

2.3 Expected value and variance

2.3.1 Expected value

Definition 2.3.1 (Expected value)**.** *Let E be an at most countable subset of $\mathbb{R}$ and $X : \Omega \to E$ be a real-valued discrete random variable. We say that X is* integrable *and denote $X \in L^1$ if*

$$\sum_{x \in E} |x| \mathbb{P}(\{X = x\}) < +\infty.$$

In that case, the expected value *of X is denoted $\mathbb{E}[X]$ and defined by*

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

Remark 2.3.2. – The integrability of X and its expected value depend only on the probability distribution of X ;

– If E is finite, then X is integrable ;

– If $\lambda \in \mathbb{R}$, then $\mathbb{E}[\lambda] = \lambda$;

– If A is an event, then $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.

Proposition 2.3.3. *Let X and Y be two integrable discrete random variables.*

(i) Linearity: For all $\lambda \in \mathbb{R}$, $\lambda X + Y$ is an integrable discrete random variable and

$$\mathbb{E}[\lambda X + Y] = \lambda \mathbb{E}[X] + \mathbb{E}[Y];$$

(ii) Positivity and non-degeneracy: If $\mathbb{P}(X \geq 0) = 1$, then $\mathbb{E}[X] \geq 0$. If in addition $\mathbb{E}[X] = 0$, then $\mathbb{P}(X = 0) = 1$.

(iii) Monotonicity: If $\mathbb{P}(X \leq Y) = 1$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Proof. Let E and F be the two at most countable sets such that $X : \Omega \to E$ and $Y : \Omega \to F$.

(i) Let $\lambda \in \mathbb{R}$ and $G = \{\lambda x + y \mid (x, y) \in E \times F\}$. The set G is at most countable, so $\lambda X + Y : \Omega \to G$ is a discrete random variable. Using σ-additivity for the second equality, Fubini's theorem (for nonnegative series) for the fourth equality and the triangle inequality for the first inequality, we have

$$\begin{aligned}
\sum_{z \in G} |z| \mathbb{P}(\{\lambda X + Y = z\}) &= \sum_{z \in G} |z| \mathbb{P}\left(\bigcup_{\substack{(x,y) \in E \times F \\ \lambda x + y = z}} \{X = x, Y = y\} \right) \\
&= \sum_{z \in G} |z| \sum_{\substack{(x,y) \in E \times F \\ \lambda x + y = z}} \mathbb{P}(\{X = x, Y = y\}) \\
&= \sum_{z \in G} \sum_{(x,y) \in E \times F} \mathbb{1}_{\{\lambda x + y = z\}} |z| \mathbb{P}(\{X = x, Y = y\}) \\
&= \sum_{(x,y) \in E \times F} \sum_{z \in G} \mathbb{1}_{\{\lambda x + y = z\}} |z| \mathbb{P}(\{X = x, Y = y\}) \\
&= \sum_{(x,y) \in E \times F} |\lambda x + y| \mathbb{P}(\{X = x, Y = y\}) \\
&\leq |\lambda| \sum_{(x,y) \in E \times F} |x| \mathbb{P}(\{X = x, Y = y\}) + \sum_{(x,y) \in E \times F} |y| \mathbb{P}(\{X = x, Y = y\}) \\
&= |\lambda| \sum_{x \in E} |x| \mathbb{P}(\{X = x\}) + \sum_{y \in F} |y| \mathbb{P}(\{Y = y\}) \\
&< +\infty,
\end{aligned}$$

so $\lambda X+Y$ is integrable. We now reproduce the same calculation as above but we remove the absolute values. This time we use Fubini's theorem for absolutely convergent series and the triangle inequality becomes an equality, so that

$$\begin{aligned}
\sum_{z\in G} z\mathbb{P}(\{\lambda X+Y=z\}) &= \sum_{z\in G}\sum_{(x,y)\in E\times F} \mathbb{1}_{\{\lambda x+y=z\}} z\mathbb{P}(\{X=x,Y=y\}) \\
&= \sum_{(x,y)\in E\times F}\sum_{z\in G} \mathbb{1}_{\{\lambda x+y=z\}} z\mathbb{P}(\{X=x,Y=y\}) \\
&= \lambda \sum_{(x,y)\in E\times F} x\mathbb{P}(\{X=x,Y=y\}) + \sum_{(x,y)\in E\times F} y\mathbb{P}(\{X=x,Y=y\}) \\
&= \lambda \sum_{x\in E} x\mathbb{P}(\{X=x\}) + \sum_{y\in F} y\mathbb{P}(\{Y=y\}) \\
&= \lambda\mathbb{E}[X]+\mathbb{E}[Y].
\end{aligned}$$

(ii) If $\mathbb{P}(\{X\geq 0\})=1$, then for all $x\in E\cap\mathbb{R}_-^*$, $\mathbb{P}(\{X=x\})=0$, so

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x\in E} x\mathbb{P}(\{X=x\}) = \sum_{x\in E\cap\mathbb{R}_+} x\mathbb{P}(\{X=x\}) + \sum_{x\in E\cap\mathbb{R}_-^*} x\mathbb{P}(\{X=x\}) \\
&= \sum_{x\in E\cap\mathbb{R}_+} x\mathbb{P}(\{X=x\}) \geq 0.
\end{aligned}$$

If in addition $\mathbb{E}[X]=0$, then $\sum_{x\in E\cap\mathbb{R}_+} x\mathbb{P}(\{X=x\}=0$, so for all $x\in E\cap\mathbb{R}_+$, $x\mathbb{P}(\{X=x\})=0$. We deduce that for all $x\in E$ such that $x\neq 0$, $\mathbb{P}(\{X=x\})=0$, so $\mathbb{P}(\{X=0\})=1$.

(iii) If $\mathbb{P}(\{X\leq Y\})=1$ then $\mathbb{P}(\{Y-X\geq 0\})$, so by linearity and positivity, $\mathbb{E}[Y]-\mathbb{E}[X]=\mathbb{E}[Y-X]\geq 0$, hence $\mathbb{E}[Y]\geq\mathbb{E}[X]$.

□

The next proposition, known as the *law of the unconscious statistician*, usually abbreviated *LOTUS*, is very useful in practice.

Proposition 2.3.4 (LOTUS). *Let E be a an at most countable subset of $\mathbb{R}$, $X:\Omega\to E$ be a discrete random variable and $f:E\to\mathbb{R}$ be a (measurable) map. Then $f(X):\omega\in\Omega\mapsto f(X(\omega))$ is a real-valued discrete random variable. Moreover,*

$$f(X)\in L^1 \iff \sum_{x\in E}|f(x)|\mathbb{P}(\{X=x\})<+\infty.$$

In that case,

$$\mathbb{E}[f(X)] = \sum_{x\in E} f(x)\mathbb{P}(\{X=x\}).$$

Proof. Since E is at most countable, $f(E)$ is at most countable as well, so $f(X) : \Omega \to f(E)$ is a real-valued discrete random variable. Using σ-additivity for the second equality and Fubini's theorem (for nonnegative series) for the fourth equality, we have

$$\begin{aligned}
\sum_{y \in f(E)} |y| \mathbb{P}(\{f(X) = y\}) &= \sum_{y \in f(E)} |y| \mathbb{P}\left(\bigcup_{\substack{x \in E \\ f(x) = y}} \{X = x\} \right) \\
&= \sum_{y \in f(E)} |y| \sum_{\substack{x \in E \\ f(x) = y}} \mathbb{P}(\{X = x\}) \\
&= \sum_{y \in f(E)} \sum_{x \in E} \mathbb{1}_{\{f(x) = y\}} |y| \mathbb{P}(\{X = x\}) \\
&= \sum_{x \in E} \sum_{y \in f(E)} \mathbb{1}_{\{f(x) = y\}} |y| \mathbb{P}(\{X = x\}) \\
&= \sum_{x \in E} |f(x)| \mathbb{P}(\{X = x\}).
\end{aligned}$$

Therefore, $f(X) \in L^1 \iff \sum_{y \in f(E)} |y| \mathbb{P}(\{f(X) = y\}) < +\infty \iff \sum_{x \in E} |f(x)| \mathbb{P}(\{X = x\}) < +\infty$.

Suppose now that $f(X) \in L^1$. We reproduce the same calculation as above but we remove the absolute values. This time we use Fubini's theorem for absolutely convergent series, so that

$$\begin{aligned}
\mathbb{E}[f(X)] = \sum_{y \in f(E)} y \mathbb{P}(\{f(X) = y\}) &= \sum_{y \in f(E)} \sum_{x \in E} \mathbb{1}_{\{f(x) = y\}} y \mathbb{P}(\{X = x\}) \\
= \sum_{x \in E} \sum_{y \in f(E)} \mathbb{1}_{\{f(x) = y\}} y \mathbb{P}(\{X = x\}) &= \sum_{x \in E} f(x) \mathbb{P}(\{X = x\}).
\end{aligned}$$

□

Proposition 2.3.5. *Let E and F be two at most countable sets and $X : \Omega \to E$ and $Y : \Omega \to F$ be two discrete random variables.*

(i) If X is independent of Y, then for all (measurable) maps $f : E \to \mathbb{R}$ and $g : F \to \mathbb{R}$ such that $f(X), g(Y) \in L^1$, we have $f(X)g(Y) \in L^1$ and

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]. \tag{2.1}$$

(ii) Conversely, if (2.1) *holds for all (measurable) bounded maps $f : E \to \mathbb{R}$ and $g : F \to \mathbb{R}$, then X is independent of Y.*

Proof. (i) Suppose that X is independent of Y. Let $f : E \to \mathbb{R}$ and $g : F \to \mathbb{R}$ be two (measurable) maps. Let $h : E \times F \to \mathbb{R}$ be defined for all $(x, y) \in E \times F$ by

$h(x,y) = f(x)g(y)$. By Fubini's theorem for nonnegative series and independence of X and Y, we get

$$\begin{aligned}\sum_{(x,y)\in E\times F} |h(x,y)|\mathbb{P}(\{X=x, Y=y\}) &= \sum_{x\in E}\sum_{y\in F} |f(x)g(y)|\mathbb{P}(\{X=x\})\mathbb{P}(\{Y=y\}) \\ &= \sum_{x\in E} |f(x)|\mathbb{P}(\{X=x\}) \sum_{y\in F} |g(y)|\mathbb{P}(\{Y=y\}).\end{aligned}$$

According to Proposition 2.3.4, $h(X,Y) = f(X)g(Y) \in L^1$. Using LOTUS, Fubini's theorem for absolutely convergent series and independence of X and Y, we have

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] = \mathbb{E}[h(X,Y)] &= \sum_{(x,y)\in E\times F} h(x,y)\mathbb{P}(\{X=x, Y=y\}) \\ &= \sum_{x\in E}\sum_{y\in F} f(x)g(y)\mathbb{P}(\{X=x\})\mathbb{P}(\{Y=y\}) \\ &= \sum_{x\in E} f(x)\mathbb{P}(\{X=x\}) \sum_{y\in F} g(y)\mathbb{P}(\{Y=y\}) = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].\end{aligned}$$

(ii) Suppose now that (2.1) holds for all (measurable) bounded maps $f : E \to \mathbb{R}$ and $g : F \to \mathbb{R}$. Let $(x,y) \in (E,F)$. Then (2.1) for $f = \mathbb{1}_{\{x\}}$ and $g = \mathbb{1}_{\{y\}}$ writes $\mathbb{E}[\mathbb{1}_{\{X=x,Y=y\}}] = \mathbb{E}[\mathbb{1}_{\{X=x\}}]\mathbb{E}[\mathbb{1}_{\{Y=y\}}]$, that is $\mathbb{P}(\{X=x, Y=y\}) = \mathbb{P}(\{X=x\})\mathbb{P}(\{Y=y\})$. So X is independent of Y.

□

2.3.2 Variance

Definition 2.3.6 (Variance)**.** *Let E be an at most countable subset of $\mathbb{R}$ and $X : \Omega \to E$ be a real-valued discrete random variable. We say that X is* square-integrable *and denote $X \in L^2$ if*

$$\sum_{x\in E} x^2\mathbb{P}(\{X=x\}) < +\infty.$$

In that case, the variance *of X is denoted* $\operatorname{Var} X$ *and defined by*

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

The square root of the variance of X is called the standard deviation *of X.*

Remark 2.3.7. According to LOTUS, $X \in L^2$ iff $X^2 \in L^1$.

Proposition 2.3.8. *Let X be a square-integrable discrete random variable. Then*

(i) *X is integrable.*

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

(iii) For all $a, b \in \mathbb{R}$, $aX + b \in L^2$ and

$$\operatorname{Var}(aX+b) = a^2 \operatorname{Var} X.$$

Proof. (i) $|X| \leq \frac{1}{2}(1+X^2)$. By hypothesis, $X^2 \in L^1$, so $|X|$ is bounded from above by an integrable random variable. Therefore, $X \in L^1$.

(ii) Let $\mu = \mathbb{E}[X]$, which is well defined according to (i). Using LOTUS, we get

$$\begin{aligned}
\operatorname{Var} X = \mathbb{E}[(X-\mathbb{E}[X])^2] &= \sum_{x\in E}(x-\mu)^2\mathbb{P}(\{X=x\}) \\
&= \sum_{x\in E} x^2\mathbb{P}(\{X=x\}) - 2\mu\sum_{x\in E} x\mathbb{P}(\{X=x\}) + \mu^2\sum_{x\in E}\mathbb{P}(\{X=x\}) \\
&= \sum_{x\in E} x^2\mathbb{P}(\{X=x\}) - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.
\end{aligned}$$

(iii) Let $a, b \in \mathbb{R}$. Then $(aX+b)^2 = aX^2 + 2abX + b^2$. By hypothesis and (i), $aX^2 \in L^1$ and $2abX \in L^1$ so the sum $(aX+b)^2$ is integrable, hence $aX + b \in L^2$. According to (ii), we have

$$\begin{aligned}
\operatorname{Var}(aX+b) &= \mathbb{E}[(aX+b)^2] - \mathbb{E}[aX+b]^2 = \mathbb{E}[a^2X^2 + 2abX + b^2] - (a\mathbb{E}[X]+b)^2 \\
&= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - b^2 \\
&= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2 \operatorname{Var} X.
\end{aligned}$$

□

Proposition 2.3.9. *Let $n \in \mathbb{N}^*$ and $X_1, \cdots, X_n$ be square-integrable discrete random variables. Then their sum $X_1 + \cdots + X_n$ is square-integrable.*

If in addition $(X_k)_{1\leq k\leq n}$ is pairwise independent, then

$$\operatorname{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \operatorname{Var}(X_k).$$

Proof. Expanding the square, we get

$$\begin{aligned}
\operatorname{Var}\left(\sum_{k=1}^n X_k\right) &= \mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] - \mathbb{E}\left[\sum_{k=1}^n X_k\right]^2 = \mathbb{E}\left[\sum_{k,l=1}^n X_kX_l\right] - \mathbb{E}\left[\sum_{k=1}^n X_k\right]\mathbb{E}\left[\sum_{l=1}^n X_l\right] \\
&= \sum_{k,l=1}^n \mathbb{E}[X_kX_l] - \sum_{k=1}^n\sum_{l=1}^n \mathbb{E}[X_k]\mathbb{E}[X_l] = \sum_{k,l=1}^n (\mathbb{E}[X_kX_l] - \mathbb{E}[X_k]\mathbb{E}[X_l]).
\end{aligned}$$

Let $k, l \in \{1, \cdots, n\}$ be such that $k \neq l$. By independence of X_k and X_l and Proposition 2.3.5, $\mathbb{E}[X_kX_l] - \mathbb{E}[X_k]\mathbb{E}[X_l] = 0$. We deduce that

$$\operatorname{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n(\mathbb{E}[X_k^2] - \mathbb{E}[X_k]^2) = \sum_{k=1}^n \operatorname{Var} X_k.$$

□

2.3.3 Moment-generating function of an integer-valued random variable

Definition 2.3.10 (Moment-generating function)**.** *Let $X : \Omega \to \mathbb{N}$ be a nonnegative integer-valued random variable. We call* moment-generating function *of X and denote g_X the map*

$$g_X \ : \ \begin{array}{ccc} [-1,1] & \to & \mathbb{R} \\ u & \mapsto & \mathbb{E}[u^X] = \sum_{n\in\mathbb{N}} u^n \mathbb{P}(\{X=u\}) \end{array} \cdot$$

Proposition 2.3.11. *The moment-generating function characterises the probability distribution of an integer-valued random variable: if $X : \Omega \to \mathbb{N}$ and $Y : \Omega \to \mathbb{N}$ are two nonnegative integer-valued random variables, then*

$$X \stackrel{d}{=} Y \iff \forall u \in [-1,1], \quad g_X(u) = g_Y(u).$$

Proof. Let $u \in [-1,1]$. According to LOTUS, the value of $g_X(u)$ depends only on the probability distribution of X. Therefore, if $X \stackrel{d}{=} Y$, then $g_X(u) = g_Y(u)$.

Conversely, suppose that $g_X(u) = g_Y(u)$ for $u \in [-1,1]$. Since the power series $\sum_{n\geq 0} u^n \mathbb{P}(\{X = n\})$ has a radius of convergence of at least 1, g_X is infinitely differentiable on $(-1,1)$. Moreover, for all $k \in \mathbb{N}$, the k-th derivative of g_X, denoted $g_X^{(k)}$, satisfies

$$\forall u \in (-1,1),\ g_X^{(k)}(u) = \sum_{n=k}^{+\infty} \frac{n!}{(n-k)!} u^{n-k} \mathbb{P}(\{X=n\}) \quad \text{and} \quad \mathbb{P}(\{X=n\}) = \frac{g_X^{(n)}(0)}{n!}\cdot \tag{2.2}$$

This implies that for all $n \in \mathbb{N}$, $\mathbb{P}(\{X = n\}) = \frac{g_X^{(n)}(0)}{n!} = \frac{g_Y^{(n)}(0)}{n!} = \mathbb{P}(\{Y = n\})$, hence $X \stackrel{d}{=} Y$. □

Proposition 2.3.12. *Let $X : \Omega \to \mathbb{N}$ be a nonnegative integer-valued random variable.*

(i) X is integrable iff its moment-generating function g_X is such that $\lim_{u\to 1^-} g'_X(u)$ is finite. In that case,

$$\mathbb{E}[X] = \lim_{u\to 1^-} g'_X(u).$$

(ii) X is square-integrable iff $\lim_{u\to 1^-} g''_X(u)$ is finite. In that case,

$$\operatorname{Var} X = \left(\lim_{u\to 1^-} g''_X(u)\right) - \left(\lim_{u\to 1^-} g'_X(u)\right)^2 + \left(\lim_{u\to 1^-} g'_X(u)\right).$$

Proof. (i) According to (2.2), g_X is differentiable on $(-1,1)$ and for all $u \in (-1,1)$, $g'_X(u) = \sum_{n=1}^{+\infty} n u^{n-1} \mathbb{P}(\{X = n\})$. The map g'_X is nonnegative and nondecreasing on $[0,1)$, so it has a limit $l \in \mathbb{R}_+ \cup \{+\infty\}$ at 1. On the one hand, for all $u \in [0,1)$, $0 \leq u \leq 1$ so $g'_X(u) \leq \sum_{n=1}^{+\infty} n\mathbb{P}(\{X = n\})$. For $u \to 1^-$ the latter inequality yields $l \leq \sum_{n=1}^{+\infty} n\mathbb{P}(\{X = n\})$. On the other hand, for all $N \in \mathbb{N}^*$ and $u \in [0,1)$, $\sum_{n=1}^{N} n u^{n-1} \mathbb{P}(\{X = n\}) \leq g'_X(u)$. For

$u \to 1^-$, this yields $\sum_{n=1}^{N} n\mathbb{P}(\{X = n\}) \le l$, which itself yields $\sum_{n=1}^{+\infty} n\mathbb{P}(\{X = n\}) \le l$ for $N \to +\infty$. We deduce that

$$\lim_{u \to 1^-} g_X'(u) = \sum_{n=1}^{+\infty} n\mathbb{P}(\{X = n\}) \in \mathbb{R}_+ \cup \{+\infty\}.$$

Therefore, $\lim_{u \to 1^-} g_X'(u)$ is finite iff $\sum_{n=1}^{+\infty} n\mathbb{P}(\{X = n\}$ is finite, in which case they are equal.

(ii) According to (2.2), g_X is twice differentiable on $(-1, 1)$ and for all $u \in (-1, 1)$, $g_X''(u) = \sum_{n=2}^{+\infty} n(n-1)u^{n-2}\mathbb{P}(\{X = n\})$. With a similar reasoning as for (i), we prove that

$$\lim_{u \to 1^-} g_X''(u) = \sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(\{X = n\}) \in \mathbb{R}_+ \cup \{+\infty\}. \tag{2.3}$$

If X is square-integrable, then $\sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(\{X = n\} \le \sum_{n=2}^{+\infty} n^2\mathbb{P}(\{X = n\}) < +\infty$ so $\lim_{u \to 1^-} g_X''(u)$ is finite.

Conversely, if $\lim_{u \to 1^-} g_X''(u)$ is finite, then $\sum_{n=2}^{+\infty} n^2\mathbb{P}(\{X = n\}) \le 2\sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(\{X = n\}) < +\infty$ so X is square-integrable.

If $X \in L^2$, then (i) and (2.3) yield

$$\begin{aligned}\lim_{u \to 1^-} g_X''(u) &= \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X] = \operatorname{Var} X + \mathbb{E}[X]^2 - \mathbb{E}[X] \\ &= \operatorname{Var} X + \left(\lim_{u \to 1^-} g_X'(u)\right)^2 - \left(\lim_{u \to 1^-} g_X'(u)\right).\end{aligned}$$

□

Proposition 2.3.13. *Let $X : \Omega \to \mathbb{N}$ and $Y : \Omega \to \mathbb{N}$ be two nonnegative integer-valued random variables. If X is independent of Y, then*

$$\forall u \in [-1, 1], \quad g_{X+Y}(u) = g_X(u)g_Y(u).$$

Proof. Let $u \in [-1, 1]$. If X if independent of Y, then by Proposition 2.3.5, we have

$$g_{X+Y}(u) = \mathbb{E}[u^{X+Y}] = \mathbb{E}[u^X u^Y] = \mathbb{E}[u^X]\mathbb{E}[u^Y] = g_X(u)g_Y(u).$$

□

2.3.4 Review of common discrete probability distributions

We end the present chapter by enumerating the main properties of the probability distributions given in Section 2.2.3.

Proposition 2.3.14. *Let $p \in [0, 1]$, $n \in \mathbb{N}^*$, $\lambda > 0$, $u \in [-1, 1]$ and X be a discrete random variable.*

(i) If $X \sim \mathcal{B}(p)$, then

$$\mathbb{E}[X] = p, \quad \operatorname{Var} X = p(1-p) \quad \textit{and} \quad g_X(u) = 1 - p + pu.$$

(ii) If $X \sim \mathcal{B}(n, p)$, then

$$\mathbb{E}[X] = np, \quad \operatorname{Var} X = np(1-p) \quad \textit{and} \quad g_X(u) = (1 - p + pu)^n.$$

(iii) If $X \sim \mathcal{P}(\lambda)$, then

$$\mathbb{E}[X] = \lambda, \quad \operatorname{Var} X = \lambda \quad \textit{and} \quad g_X(u) = e^{\lambda(u-1)}.$$

(iv) If $p \neq 0$ and $X \sim \mathcal{G}eo(p)$, then

$$\mathbb{E}[X] = \frac{1}{p}, \quad \operatorname{Var} X = \frac{1-p}{p^2} \quad \textit{and} \quad g_X(u) = \frac{pu}{1-(1-p)u}.$$

Proof. (i) Suppose $X \sim \mathcal{B}(p)$. Then

$$\begin{aligned}
\mathbb{E}[X] &= 0 \times (1-p) + 1 \times p = p; \\
\operatorname{Var} X &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 0^2 \times (1-p) + 1^2 \times p^2 - p^2 = p(1-p); \\
g_X(u) &= u^0 \times (1-p) + u^1 \times p = 1 - p + pu.
\end{aligned}$$

(ii) Suppose $X \sim \mathcal{B}(n, p)$. Let $X_1, \cdots, X_n$ be i.i.d. discrete random variables, each heaving a Bernoulli distribution with parameter p. According to Proposition 2.2.13, $X \overset{d}{=} X_1 + \cdots + X_n$, so using (i), Proposition 2.3.9 and Proposition 2.3.13, we have

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np; \\
\operatorname{Var} X &= \operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var} X_1 + \cdots + \operatorname{Var} X_n = np(1-p); \\
g_X(u) &= g_{X_1 + \cdots + X_n}(u) = g_{X_1}(u) \times \cdots \times g_{X_n}(u) = (1 - p + pu)^n.
\end{aligned}$$

(iii) Suppose $X \sim \mathcal{P}(\lambda)$. Then

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n=0}^{+\infty} n \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} = \mathrm{e}^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^n}{(n-1)!} = \lambda \mathrm{e}^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} = \lambda \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda} = \lambda; \\
\mathbb{E}[X(X-1)] &= \sum_{n=0}^{+\infty} n(n-1) \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} = \mathrm{e}^{-\lambda} \sum_{n=2}^{+\infty} \frac{\lambda^n}{(n-2)!} = \lambda^2 \mathrm{e}^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} = \lambda^2; \\
\operatorname{Var} X &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda; \\
g_X(u) &= \sum_{n=0}^{+\infty} u^n \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} = \mathrm{e}^{-\lambda} \sum_{n=0}^{+\infty} \frac{(\lambda u)^n}{n!} = \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda u} = \mathrm{e}^{\lambda(u-1)}.
\end{aligned}$$

(iv) Suppose $p \neq 0$ and $X \sim \mathcal{G}eo(p)$. Then

$$g_X(u) = \sum_{n=1}^{+\infty} u^n p(1-p)^{n-1} = pu \sum_{n=0}^{+\infty} ((1-p)u)^n = \frac{pu}{1-(1-p)u}.$$

Moreover $g_X'(u) = \frac{p}{(1-(1-p)u)^2} \underset{u \to 1^-}{\longrightarrow} \frac{1}{p}$ and $g_X''(u) = \frac{2p(1-p)}{(1-(1-p)u)^3} \underset{u \to 1^-}{\longrightarrow} \frac{2(1-p)}{p^2}$. We deduce from Proposition 2.3.12 that

$$\mathbb{E}[X] = \frac{1}{p};$$

$$\operatorname{Var} X = \frac{2(1-p)}{p^2} - \frac{1}{p^2} + \frac{1}{p} = \frac{2(1-p)-1+p}{p^2} = \frac{1-p}{p^2}.$$

$\square$