1 Discrete Random Variables

1.1 Definition

Let Ω be a sample space. A **Discrete random variable** is a *function* $X : \omega \in \Omega \mapsto X(\omega) \in \mathbb{R}$ that takes on a **finite number** of values $a_1, a_2, \ldots, a_N$, or a **countably infinite** number of values $a_1, a_2, \ldots, a_N, \ldots$.

1.2 Examples

- Consider a single toss of a coin, with sample space $\Omega = \{H, T\}$.
  The function X on Ω defined by
  $$X(H) = 1$$
  $$X(T) = -1$$
  is a discrete random variable.
- Consider N successive tosses of a coin, with sample space
  $$\Omega = \underbrace{\{H,T\}, \times\{H,T\} \times \ldots \times \{H,T\}}_{N \text{ terms}}$$
  The function X on Ω defined by
  $$X(\omega) = \text{number of } H\text{'s in } \omega \text{ for } \omega \in \Omega$$
  is a discrete random variable.
- Consider an experiment in which one rolls two dice, with sample space $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$.
  The function X on Ω defined by
  $$X(\omega_1, \omega_2) = \max(\omega_1, \omega_2) \qquad \text{for } (\omega_1, \omega_2) \in \Omega$$
  is a discrete random variable.

2 Characterizing random variables

As we have just seen, a discrete random variable is a mathematical procedure for assigning a numerical quantity to each physical outcome of an experiment.

Instead of associating probabilities to events in a sample space, we can think of characterizing a discrete random variable X through its statistical properties. This is done as follows.

2.1 Probability mass function

2.1.1 Definition

The **probability mass function** p of a discrete random variable X is the function
$$p : a \in \mathbb{R} \mapsto p(a) \in [0, 1]$$
defined by
$$p(a) = P(X = a) \quad \text{for } a \in (-\infty, \infty) \tag{1}$$

By definition, this means that if a discrete random variable X takes values $a_1, a_2, \ldots, a_N, \ldots$, we can say that

- $p(a_i) > 0$
- $\sum_i p(a_i) = 1$
- $p(a) = 0$ for any a such that $a \neq a_i$ for all i.

2.1.2 Example

Let us return to our first example of a discrete random variable:

$$X(H) = 1$$
$$X(T) = -1$$

The probability mass function p of X is such that

$$p(-1) = \frac{1}{2} \quad , \quad p(1) = \frac{1}{2} \quad , \quad \text{and } \ p(a = 0) \text{ for all } a \ \text{ such that } a \neq -1 \ , \ a = 1$$

We can verify that

$$p(-1) + p(1) = \frac{1}{2} + \frac{1}{2}$$

2.2 Cumulative distribution function

2.2.1 Definition

The **distribution function** (or **cumulative distribution function**, abbreviated c.d.f.) of a random variable X is the function $F : a \in \mathbb{R} \mapsto F(a) \in [0, 1]$ defined by

$$F(a) = P(X \leq a) \qquad \text{for } a \in (-\infty, \infty) \tag{2}$$

2.2.2 Connection between probability mass function and distribution function

For a discrete random variable, it is straightforward to see how the probability mass function and distribution function are connected.

Let X be a discrete random variable which takes on the values $a_1, a_2, \ldots$. By definition, we have

$$F(a) = P(X \leq a) \qquad \forall \ a \in (\infty, +\infty)$$

Now, let us fix $a \in \mathbb{R}$, and let $a_1, a_2, \ldots, a_N$ be the N values such that

$$a_i \leq a \ , \ i = 1, \ldots, N$$

We can write

$$P(X \leq a) = P(X = a_1) + P(X = a_2) + \ldots + P(X = a_N) = p(a_1) + p(a_2) + \ldots + p(a_N)$$

Hence, we have the general formula linking probability mass function and cumulative distribution function:

$$F(a) = \sum_{a_i \leq a} p(a_i)$$

It is clear that both F and p contain all the probabilistic information of the discrete random variable X. They can be used interchangeably.

We will soon see that the situation is different for continuous random variables, for which only the distribution function can be defined.

2.2.3 Immediate properties of distribution functions

- Let $x \in \mathbb{R}$, $y \in \mathbb{R}$ such that $x \leq y$.

  $$\{X \leq x\} \subset \{X \leq y\}$$

  so

  $$P(X \leq x) \leq P(X \leq y)$$

  We conclude that

  $$\underline{x \leq y \qquad \Rightarrow \qquad F(x) \leq F(y)}$$

  F **is a nondecreasing function**.

- $$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \leq x) = 0$$
  $$\lim_{x \to +\infty} F(x) = \lim_{x \to +\infty} P(X \leq x) = 1$$

- To prove the last property, let us first state a theorem we will not prove:

  (i) Let $(A_1, A_2, A_3, \ldots)$ be an increasing sequence of events, i.e. a sequence such that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$.
  Then
  $$\lim_{n \to +\infty} P(A_n) = P(\bigcup_{i=1}^{+\infty} A_i)$$

  (ii) Let $(A_1, A_2, A_3, \ldots)$ be a decreasing sequence of events, i.e. a sequence such that $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$.
  Then
  $$\lim_{n \to +\infty} P(A_n) = P(\bigcap_{i=1}^{+\infty} A_i)$$

  We will now use part (ii) of this theorem to prove a third property of cumulative distribution functions.

  Let x_n be a decreasing sequence such that $x_n \xrightarrow[n \to +\infty]{} x$. Then $\{X \leq x\} \subseteq \{X \leq x_n\}$ for all n, and $\{X \leq x\}$ is the largest set for which this holds. Hence,

  $$\bigcap_{n=1}^{+\infty} \{X \leq x_n\} = \{X \leq x\}$$

  We can therefore write

  $$\begin{aligned} \lim_{n \to +\infty} F(x_n) &= \lim_{n \to +\infty} P(X \leq x_n) \\ &= P(\bigcap_{i=1}^{+\infty} \{X \leq x_i\} \qquad \text{(Using part (ii) of theorem)} \\ &= P(X \leq x) = F(x) \end{aligned}$$

  Since this is true for any decreasing sequence x_n such that $x_n \xrightarrow[n \to +\infty]{} x$,

  $$\underline{\lim_{\epsilon \to 0^+} F(x + \epsilon) = F(x)}$$

  In other words, F **is right-continuous**.

Illustration: Let us plot the functions p and F for the random variable X defined by

$$\begin{cases} X(H) = 1, & \text{with probability } \frac{1}{2} \\ X(T) = -1, & \text{with probability } \frac{1}{2} \end{cases} \tag{3}$$

which may be associated with the single toss of a fair coin.

The plots are shown in Figure 1.

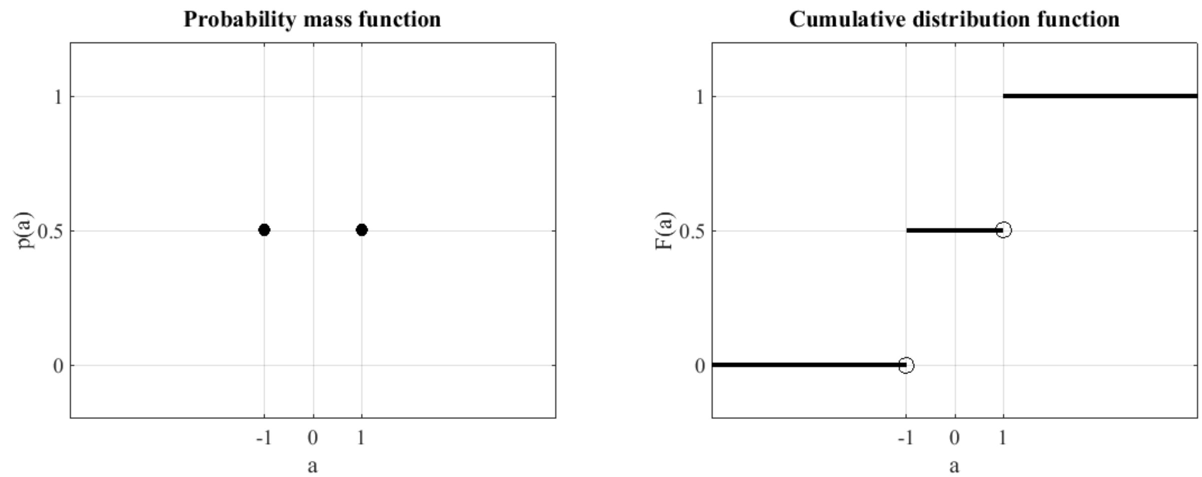


Figure 1: Probability mass function and cumulative distribution function for the random variable defined in Eq.(3)

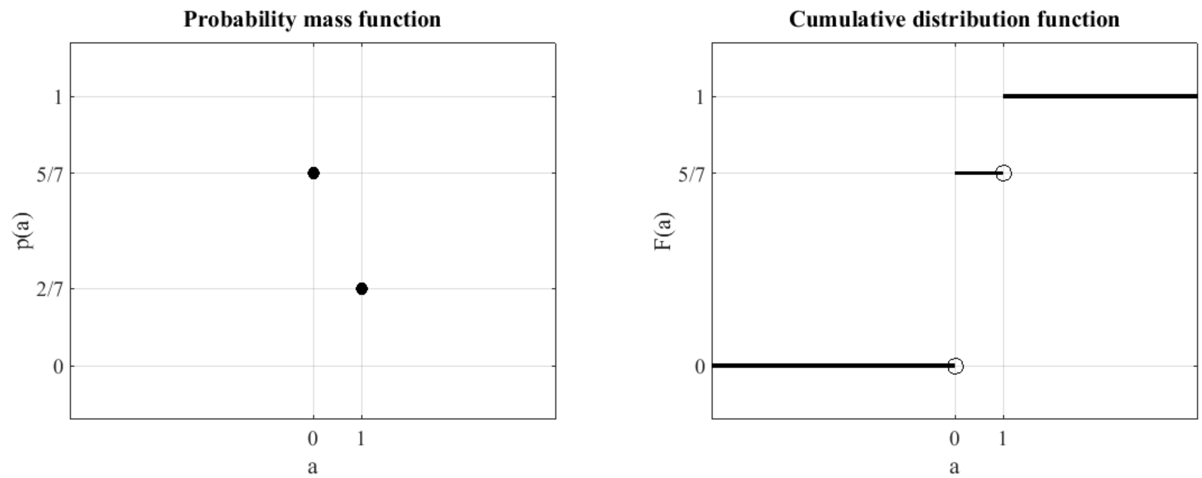


Figure 2: Probability mass function and cumulative distribution function for a Bernoulli random variable with parameter $\frac{2}{7}$.

3 Some important random variables/distributions

3.1 Bernoulli distribution

A discrete random variable X has a **Bernoulli distribution** with **parameter** p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(0) = P(X = 0) = 1 - p \quad \text{and} \quad p_X(1) = P(X = 1) = p \qquad \text{Bernoulli distribution} \tag{4}$$

The probability mass function and cumulative distribution function for a Bernoulli distributed random variable X are shown in Figure 2 for $p = \frac{2}{7}$. We call such a random variable/distribution Ber(p).

3.2 The binomial distribution

This corresponds to the situation in which N independent Bernoulli trials are performed (e.g. N independent experiments with probability p of success and probability $1 - p$ of failure), and X is the number of successes.

X takes on values in $\{0, 1, 2, \ldots, N\}$. Let us derive the formula for $P(X = k)$ for any $k = 0, \ldots, N$.

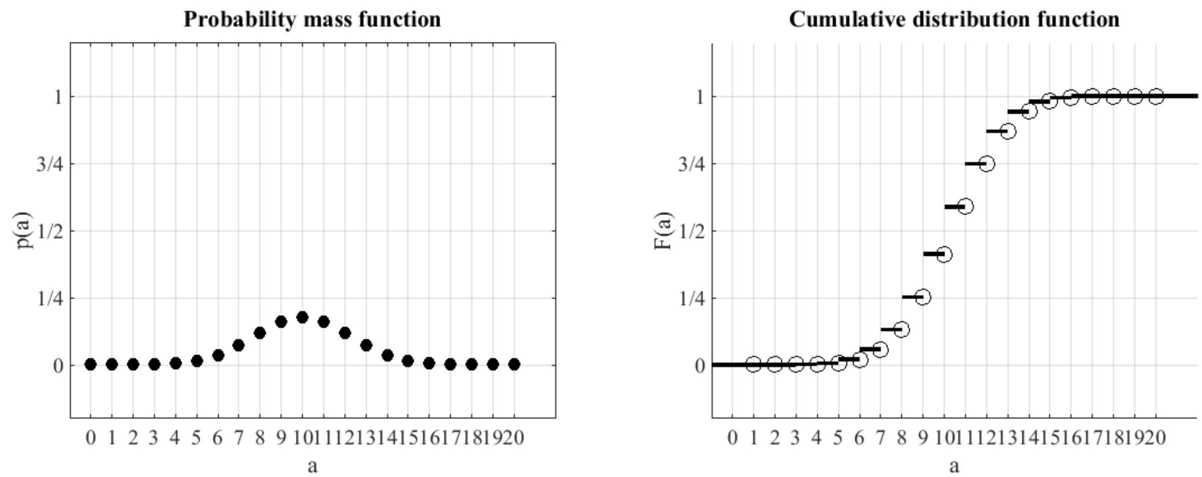


Figure 3: Probability mass function and cumulative distribution function for a binomial random variable with parameters 20 and $\frac{1}{2}$.

The probability of any sequence with k successes and $N-k$ failures is

$$p^k(1-p)^{N-k}$$

Following the same reasoning as in problem 3 in Homework 2, we know that there are

$$\binom{N}{k}$$

different sequences with k successes. Hence

$$P(X=k)=\binom{N}{k}p^k(1-p)^{N-k}$$

This result leads to the following definition.

Definition: A discrete random variable X has **a binomial distribution with parameters n and p**, where n is a positive integer and $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k)=P(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \quad , \quad k=0,1,\ldots,n \qquad \text{Binomial distribution} \tag{5}$$

Such a random variable is often written $\mathrm{Bin}(n,p)$ or $\mathrm{B}(n,p)$. The probability mass function and cumulative distribution function for $\mathrm{B}(20,\frac{1}{2})$ are shown in Figure 3.

3.3 The geometric distribution

This corresponds to the situation in which independent Bernoulli trials are performed, and X is the number of the trial on which the first success occurred.

X takes on values in $\{1,2,\ldots\}$. The event $\{X=k\}$ occurs as follows:

$$\underset{\text{1st trial}}{\underline{F}}\;\underset{\text{2nd trial}}{\underline{F}}\cdots\underset{(k-1)^{\text{th}}\text{ trial}}{\underline{F}}\;\underset{k^{\text{th}}\text{ trial}}{\underline{S}}$$

Hence,

$$P(X=k)=(1-p)^{k-1}p \qquad \text{for } k=1,2,\ldots$$

This short derivation motivates the following definition:

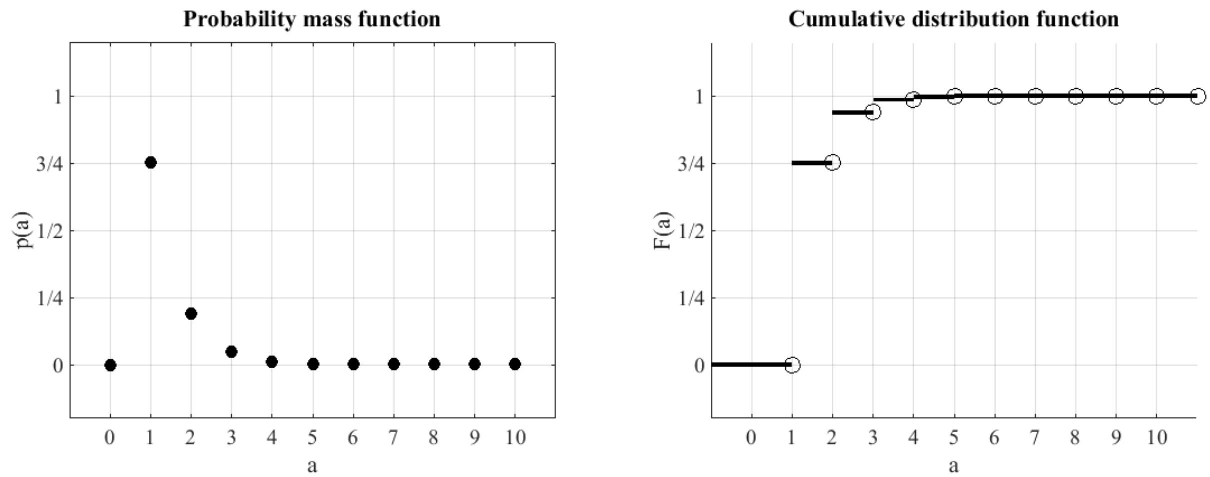


Figure 4: Probability mass function and cumulative distribution function for a geometric random variable with parameter $\frac{3}{4}$. The plots are truncated after $a = 10$ because the $p_X(a) \leq 10^{-6}$ is very small past that point.

Definition: A discrete random variable X has a **geometric distribution with parameter** p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = (1-p)^{k-1}p \quad , \quad k = 1, 2, \ldots \qquad \text{Geometric distribution} \tag{6}$$

A random variable with such a distribution is often written Geo(p). The probability mass function and cumulative distribution function for Geo($\frac{3}{4}$) are shown in Figure 4. Note that a complete plot would require us to plot the probability mass function and cumulative distribution function for all $a \in \mathbb{N}^*$. Indeed, $\forall a \in \mathbb{N}^*$, $p_X(a)$ is finite. However, for this particular example we stop at $a = 10$ because the decay of the probability mass function of Geo($\frac{3}{4}$) with a is very fast: $p_X(10) = 2.861022949218750 \cdot 10^{-6}$, and $p_X(11)$, which we chose not to plot, is such that $p_X(11) = 7.152557373046875 \cdot 10^{-7}$.
As an exercise, let us verify that for such a random variable, $\lim_{x \to +\infty} F(x) = 1$.

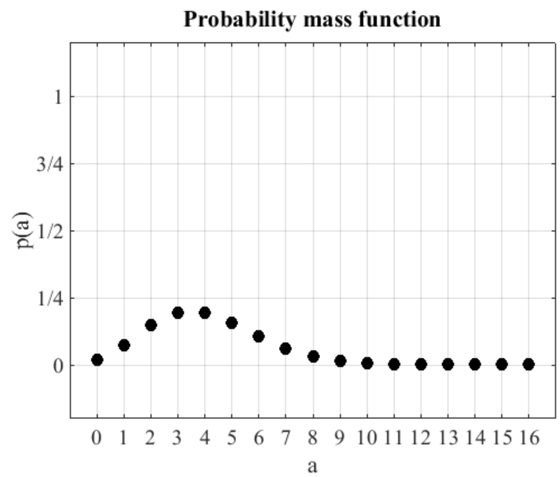
$$\begin{aligned}
\lim_{x \to +\infty} F(x) &= \lim_{x \to +\infty} P(X \leq x) = \sum_{k=1}^{+\infty} p_X(k) \\
&= \sum_{k=1}^{+\infty} (1-p)^{k-1}p = p \sum_{k=1}^{+\infty} (1-p)^{k-1} \\
&= p \frac{1}{1-(1-p)} = 1
\end{aligned}$$

3.4 The Poisson distribution

Definition: A discrete random variable X has a Poisson distribution with parameter λ, where λ is a positive real number, if its probability mass function is given by

$$p_X(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad , \quad k = 0, 1, 2, \ldots \qquad \text{Poisson distribution} \tag{7}$$

This distribution is often written $P(\lambda)$ or Poisson(λ). The probability mass function and cumulative distribution function for Poisson(4) are shown in Figure 5. As for the geometric distribution, a complete plot would require us to plot the probability mass function and cumulative distribution function for all $a \in \mathbb{N}^*$. Indeed, $\forall a \in \mathbb{N}^*$, $p_X(a)$ is finite. However, for this particular example we stop at $a = 16$ because for our



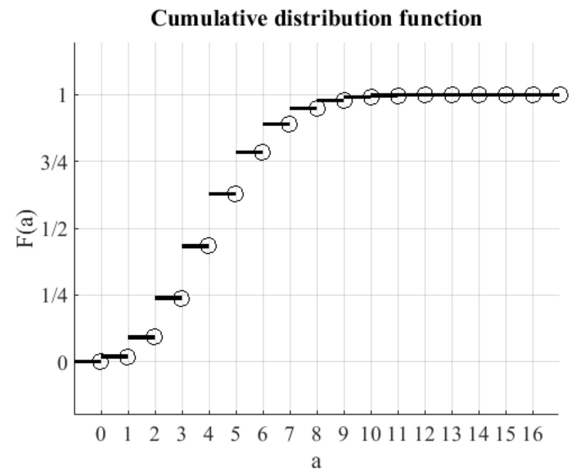


Figure 5: Probability mass function and cumulative distribution function for a Poisson random variable with parameters 4. The plots are truncated after $a = 16$ because the $p_X(a) \leq 10^{-6}$ is very small past that point.

choice $\lambda = 4$, $p_X(17)$, which we chose not to plot, is such that $p_X(17) = 8.846539272254220 \cdot 10^{-7}$, and $p_X(a) < p_X(17) \quad, \forall a > 17$.

A Poisson random variable is often used to represent the number of times an event occurs in a given interval of time, if this event occurs with a known constant mean rate, and independently of the time since the last event. λ then corresponds to the average number of events per interval. We will return to the Poisson distribution later in this course.

A Poisson random variable can be seen as the limit of the binomial distribution when the number of experiments n is very large, and the probability p is very small.

Once again, one can verify as an exercise that, for a Poisson random variable,

$$\lim_{x \to +\infty} F(x) = \lim_{x \to +\infty} P(X \leq x) = \sum_{k=0}^{+\infty} p_X(k)$$
$$= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$