2 Cumulative Distribution Function

Definition 2.1. *Let X be a random variable (of <u>any</u> type, discrete or continuous). The function $F = F_X : \mathbb{R} \to \mathbb{R}$, defined by*

$$F_X(x) \;=\; P(X \le x), \tag{2.1}$$

is called the ***(cumulative) distribution function (cdf)*** *of X.*

Example 2.2. Let us go back to Example (1.4) (or (1.10)) in Lecture 3 (the indicator of an event). Its pdf is

$$X_A \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix},\; p = P(A).$$

From the analysis we did, let us recall:
For $x < 0$,

$$P(X_A \le x) \;=\; P(\emptyset) \;=\; 0.$$

If $0 \le x < 1$,

$$P(X_A \le x) \;=\; P(X_A = 0) \;=\; 1 - p.$$

Finally for $x \ge 1$,

$$P(X_A \le x) \;=\; P(\{X_A = 0\} \cup \{X_A = 1\}) \;=\; 1 - p + p \;=\; 1.$$

So, we find now the cdf of X_A to be

$$F_A(x) \;=\; \begin{cases} 0, & \text{if} \quad x < 0 \\ 1-p, & \text{if} \quad 0 \le x < 1 \\ 1, & \text{if} \quad x \ge 1. \end{cases}$$

The graphic representation of F_A is given in Figure 1.

Remark 2.3. It easily follows from the previous example that for a discrete random variable with pdf

$$X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I},$$

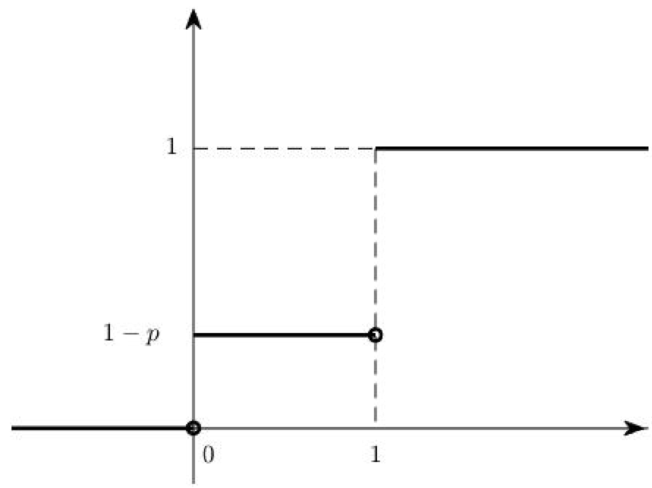

Fig. 1: Cumulative distribution function for the indicator random variable

the cdf is computed by

$$F(x) \quad = \quad \sum_{x_i \le x} p_i \tag{2.2}$$

and for every $A \subseteq \mathbb{R}$,

$$P(X \in A) \quad = \quad \sum_{x_i \in A} p_i. \tag{2.3}$$

Theorem 2.4. *Let X be a random variable with cdf $F : \mathbb{R} \to \mathbb{R}$. Then F has the following properties:*

a) *If $a < b$ are real numbers, then $P(a < X \le b) = F(b) - F(a)$.*

b) *F is monotonely increasing, i.e. if $a < b$, then $F(a) \le F(b)$.*

c) *F is right continuous, i.e. $F(x+0) = F(x)$, for every $x \in \mathbb{R}$, where $F(x+0) = \lim\limits_{y \searrow x} F(y)$ is the limit from the right at x.*

d) $\lim\limits_{x \to -\infty} F(x) = 0$ *and* $\lim\limits_{x \to \infty} F(x) = 1$.

e) *For every $x \in \mathbb{R}$, $P(X < x) = F(x-0) = \lim\limits_{y \nearrow x} F(y)$ and $P(X = x) = F(x) - F(x-0)$.*

Proof. (Selected)
a) If $a < b$, then $X \le a$ implies $X \le b$, so, as events,

$$(X \le a) \subseteq (X \le b) \ \text{ and } \ (X \le a) \cap (X \le b) = (X \le a).$$

Then, (by Theorem 2.5 c) in Chapter 1 (Lecture 1) and the fact that $A \cap \overline{B} = A \setminus B$),

$$\begin{aligned} P(a < X \le b) &= P\Big((X \le b) \cap (\overline{X \le a})\Big) \ = \ P\Big((X \le b) \setminus (X \le a)\Big) \\ &= P(X \le b) - P(X \le a) \ = \ F(b) - F(a). \end{aligned}$$

b) If $a < b$, then $F(b) - F(a) = P(a < X \le b) \ge 0$, since it is a probability.
d) We have

$$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \le x) = P(\emptyset) = 0.$$

and

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} P(X \le x) = P(S) = 1.$$

e) Just the second part:

$$P(X = x) = P\Big((X \le x) \setminus (X < x)\Big) = P(X \le x) - P(X < x) = F(x) - F(x - 0).$$

$\square$

3 Common Discrete Distributions

Bernoulli Distribution $Bern(p)$

A random variable X has a Bernoulli distribution with parameter $p \in (0, 1)$, if its pdf is

$$X \begin{pmatrix} 0 & 1 \\ 1 - p & p \end{pmatrix}. \tag{3.1}$$

Notice that this is the pdf of the indicator random variable from Example 1.10 in Chapter 1 (Lecture 3). A Bernoulli r.v. models the occurrence or nonoccurrence of an event.

Discrete Uniform Distribution $U(m)$

A random variable X has a Discrete Uniform distribution ($\boxed{\text{unid}}$) with parameter $m \in \mathbb{N}$, if its pdf is

$$X \begin{pmatrix} k \\ \dfrac{1}{m} \end{pmatrix}_{k=\overline{1,m}}. \tag{3.2}$$

The random variable in Example 1.3 (and 1.9) (Lecture 3), the number shown on a die, has a Discrete Uniform distribution $U(6)$.

Binomial Distribution $B(n, p)$

A random variable X has a Binomial distribution ($\boxed{\text{bino}}$) with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$ ($q = 1 - p$), if its pdf is

$$X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}. \tag{3.3}$$

This distribution corresponds to the Binomial model. Given n Bernoulli trials with probability of success p, let X denote the number of successes. Then $X \in B(n, p)$. Also, notice that the Bernoulli distribution is a particular case of the Binomial one, for $n = 1$, $Bern(p) = B(1, p)$.

Hypergeometric Distribution $H(N, n_1, n)$

A random variable X has a Hypergeometric distribution ($\boxed{\text{hyge}}$) with parameters $N, n_1, n \in \mathbb{N}$ ($n, n_1 \leq N$), if its pdf is

$$X \begin{pmatrix} k \\ \dfrac{C_{n_1}^k C_{N-n_1}^{n-k}}{C_N^n} \end{pmatrix}_{k=\overline{0,n}}. \tag{3.4}$$

This distribution corresponds to the Hypergeometric model. If X is the number of successes in a Hpergeometric model, then $X \in H(N, n_1, n)$.

Negative Binomial (Pascal) Distribution $NB(n, p)$

A random variable X has a Negative Binomial (Pascal) ($\boxed{\text{nbin}}$) distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, if its pdf is

$$X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k=0,1,\ldots}. \tag{3.5}$$

This distribution corresponds to the Negative Binomial model. If X denotes the number of failures that occurred before the occurrence of the n^{th} success in a Negative Binomial model, then $X \in NB(n,p)$.

Geometric Distribution $Geo(p)$

As before (probabilistic models), we have an important special case for the Negative Binomial distribution; if $n = 1$ in the previous distribution, then we have a *Geometric distribution*. A random variable X has a Geometric distribution ($\boxed{\text{geo}}$) with parameter $p \in (0,1)$, if its pdf is given by

$$X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k=0,1,\dots}. \tag{3.6}$$

If X denotes the number of failures that occurred before the occurrence of the 1^{st} success in a Geometric model, then $X \in Geo(p)$. Also, $Geo(p) = NB(1,p)$.

Poisson Distribution $\mathcal{P}(\lambda)$

A random variable X has a Poisson distribution ($\boxed{\text{poiss}}$) with parameter $\lambda > 0$, if its pdf is

$$X \begin{pmatrix} k \\ \dfrac{\lambda^k}{k!}e^{-\lambda} \end{pmatrix}_{k=0,1,\dots} \tag{3.7}$$

A Poisson r.v. **<u>does not</u>** come from the Poisson model! Poisson random variables arise in connection with so-called Poisson *processes,* processes that involve observing discrete events in a continuous interval of <u>time</u>, length, space, etc. The variable of interest in a Poisson process, X, represents the number of occurrences of the discrete event in a fixed interval of time, length, space. For instance, the number of gas emissions taking place at a nuclear plant in a 3-month period, the number of earthquakes hitting a certain area in a year, the number of white blood cells in a drop of blood, all these are modeled by Poisson random variables. The parameter λ of a Poisson distribution represents the *average* number of occurrences of the event in that interval of time or other continuous medium (this will be discussed in more detail in the next chapter).

Poisson's distribution is also known as the "law of rare events", the name coming from the fact that

$$\lim_{k\to\infty} \frac{\lambda^k}{k!}e^{-\lambda} = 0,$$

i.e. as k gets larger, the event $(X = k)$ becomes less probable, more "rare". The discrete events that are counted in a Poisson process are also called "rare events".

4 Discrete Random Vectors; Joint Probability Distribution Function; Operations with Discrete Random Variables and Independent Discrete Random Variables

We will restrict our discussion to a two-dimensional discrete random vector $(X, Y) : S \to \mathbb{R}^2$.

Definition 4.1. *Let $(S, \mathcal{K}, P)$ be a probability space. A* ***discrete random vector*** *is a function $(X, Y) : S \to \mathbb{R}^2$ satisfying the following two conditions:*

(i) *for all $(x, y) \in \mathbb{R}^2$,*

$$(X \le x, Y \le y) = \{e \in S \mid X(e) \le x, Y(e) \le y\} \in \mathcal{K}$$

(ii) *the set of values that it takes $(X, Y)(S)$ is at most countable in $\mathbb{R}^2$;*

Definition 4.2. *Let $(X, Y) : S \to \mathbb{R}^2$ be a two-dimensional discrete random vector. The* ***joint probability distribution (function)*** *of (X, Y) is a two-dimensional array of the form*

$$\begin{array}{c|ccccc|c} X \setminus Y & y_1 & \dots & y_j & \dots & & \\ \hline x_1 & & & & & & \\ \vdots & & & \vdots & & & \\ x_i & & \dots & p_{ij} & \dots & & p_i \\ \vdots & & & \vdots & & & \\ & & & & & & \\ \hline & & & q_j & & & \end{array} \tag{4.1}$$

where $(x_i, y_j) \in \mathbb{R}^2,\ (i, j) \in I \times J$ are the values that (X, Y) takes and $p_{ij} = P(X = x_i, Y = y_j)$ is the probability that (X, Y) takes the value (x_i, y_j).

Proposition 4.3. *Let (X, Y) be a random vector with joint probability distribution given by (4.1). Then*

$$\sum_{j \in J} p_{ij} = p_i \ \text{ and } \ \sum_{i \in I} p_{ij} = q_j,$$

where $p_i = P(X = x_i),\ i \in I$ and $q_j = P(Y = y_j),\ j \in J$. The probabilities p_i and q_j are called ***marginal*** *pdf's.*

Let $X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$ and $Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j\in J}$ be two discrete random variables and let $\alpha \in \mathbb{R}$. As before, denote by $p_{ij} = P(X = x_i, Y = y_j)$. We can define the following operations:

Sum. The sum of X and Y is the random variable with pdf given by

$$X + Y \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}. \tag{4.2}$$

Product. The product of X and Y is the random variable with pdf given by

$$X \cdot Y \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}. \tag{4.3}$$

Scalar Multiple. The random variable αX, $\alpha \in \mathbb{R}$, with pdf given by

$$\alpha X \begin{pmatrix} \alpha x_i \\ p_i \end{pmatrix}_{i\in I}. \tag{4.4}$$

Quotient. The quotient of X and Y is the random variable with pdf given by

$$X/Y \begin{pmatrix} x_i/y_j \\ p_{ij} \end{pmatrix}_{(i,j)\in I\times J}, \tag{4.5}$$

provided that $y_j \neq 0$, for all $j \in J$.
In general, if $h : \mathbb{R} \to \mathbb{R}$ is a function, then we can define the random variable $h(X)$, with pdf given by

$$h(X) \begin{pmatrix} h(x_i) \\ p_i \end{pmatrix}_{i\in I}. \tag{4.6}$$

Definition 4.4. *Two discrete random variables X and Y with probability distribution functions*

$$X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I} \quad \textit{and} \quad Y \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j\in J}$$

*are said to be **independent** if*

$$p_{ij} = P\left(X = x_i, Y = y_j\right) = P\left(X = x_i\right) P\left(Y = y_j\right) = p_i q_j, \tag{4.7}$$

for all $(i, j) \in I \times J$.

Remark 4.5. If X and Y are independent discrete random variables, then in (4.2), (4.3) and (4.5), $p_{ij} = p_i q_j$, for all $(i, j) \in I \times J$.

Example 4.6. Let X be a random variable with pdf

$$X \left(\begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{array} \right).$$

Find the pdf of $Y = 3X^2 - 1$.

Solution. Remember, we operate on the *values*, **never** on the probabilities!
If X takes the values $-1, 0$ and 1, then Y takes the values -1 (when $X = 0$) and 2 (when $X = -1$ or $X = 1$).

Now, we compute (carefully!) the probability for each value.

$$\begin{aligned} P(Y = -1) &= P(X = 0) \\ &= \frac{1}{4}, \\ P(Y = 2) &= P\Big((X = -1) \cup (X = 1)\Big) \\ &= P(X = -1) + P(X = 1) \\ &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4}, \end{aligned}$$

since the events $(X = -1)$ and $(X = 1)$ are mutually exclusive.
Thus, the pdf of Y is

$$Y \left(\begin{array}{cc} -1 & 2 \\ \frac{1}{4} & \frac{3}{4} \end{array} \right).$$

■

Example 4.7. Let X and Y be two independent random variables with pdf's

$$X \left(\begin{array}{cc} -1 & 0 \\ 0.2 & 0.8 \end{array} \right) \text{ and } Y \left(\begin{array}{cc} 1 & 2 \\ 0.6 & 0.4 \end{array} \right),$$

respectively. Find the pdf of $X + Y$.

Solution. First, let's find all the possible values of $X + Y$, by taking all the combinations of $x_i + y_j, i, j = 1, 2$. So, $X + Y$ can take the values $0, 1$ and 2.
Then compute their corresponding probabilities:

$$
\begin{aligned}
P(X+Y=0) &= P(X=-1, Y=1) \\
&\overset{\text{ind}}{=} P(X=-1)P(Y=1) \\
&= 0.2 \cdot 0.6 \;=\; 0.12, \\
P(X+Y=1) &= P\Big((X=-1, Y=2) \cup (X=0, Y=1)\Big) \\
&\overset{\text{m.e.}}{=} P(X=-1, Y=2) + P(X=0, Y=1) \\
&\overset{\text{ind}}{=} P(X=-1)P(Y=2) + P(X=0)P(Y=1) \\
&= 0.2 \cdot 0.4 + 0.8 \cdot 0.6 \;=\; 0.56, \\
P(X+Y=2) &= P(X=0, Y=2) \\
&\overset{\text{ind}}{=} 0.8 \cdot 0.4 \;=\; 0.32.
\end{aligned}
$$

Alternatively, we could have computed the first and the third (which are easier) and found the second one by

$$P(X+Y=1) = 1 - \Big(P(X+Y=0) + P(X+Y=2)\Big) = 1 - 0.44 = 0.56.$$

■

Remark 4.8.
1. The sum of n independent $Bern(p)$ random variables is a $B(n, p)$ variable.
2. The sum of n independent $Geo(p)$ random variables is a $NB(n, p)$ variable.