

1.4 Distribution Functions

Definition 1.8. The probability of the event $(X \leq x)$ expressed as a function of $x \in \mathbb{R}$:

$$F_X(x) = P_X(X \leq x)$$

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

Example 1.7. The cdf of the rv defined in Example 1.5 can be written as

$$F_X(x) = \begin{cases} 0, & \text{for } x \in (-\infty, 0); \\ q, & \text{for } x \in [0, 1); \\ q + p = 1, & \text{for } x \in [1, \infty). \end{cases}$$

□

Properties of cumulative distribution functions are given in the following theorem.

Theorem 1.4. The function $F(x)$ is a cdf iff the following conditions hold:

- (i) The function is nondecreasing, that is, if $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$;
- (ii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
- (iii) $\lim_{x \rightarrow \infty} F_X(x) = 1$;
- (iv) $F(x)$ is right-continuous.

Proof. Note that a cdf can be equivalently written as

$$F_X(x) = P_X(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(A_x),$$

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

(i) For any $x_i < x_j$ we have $A_{x_i} \subseteq A_{x_j}$. Thus by 1.1 part (f), we have

$$P(A_{x_i}) \leq P(A_{x_j}) \quad \text{and so} \quad F_X(x_i) \leq F_X(x_j).$$

(ii) Let $\{x_n : n = 1, 2, \dots\}$ be a sequence of decreasing real numbers such that $x_n \rightarrow -\infty$ as $n \rightarrow \infty$. Then, for $x_n \geq x_{n+1}$ we have $A_{x_n} \supseteq A_{x_{n+1}}$ and

$$\bigcap_{n=1}^{\infty} A_{x_n} = \emptyset.$$

Hence, by Theorem 1.3 (b), we get

$$\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(A_{x_n}) = P\left(\bigcap_{n=1}^{\infty} A_{x_n}\right) = P(\emptyset) = 0.$$

Since the above holds for any sequence $\{x_n\}$ such that $\{x_n\} \rightarrow -\infty$, we conclude that $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

(iii) Can be proved by similar reasoning as in (ii).

(iv) A function $g : \mathbb{R} \rightarrow \mathbb{R}$ is right continuous if $\lim_{\delta \rightarrow 0} g(x + \delta) = g(x)$.

Let x_n be a decreasing sequence such that $x_n \rightarrow x$ as $n \rightarrow \infty$. Then, by definition, $A_x \subseteq A_{x_n}$ for all n and A_x is the largest set for which it is true. This gives

$$\bigcap_{n=1}^{\infty} A_{x_n} = A_x$$

and by Theorem 1.3 (b), we get

$$\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(A_{x_n}) = P\left(\bigcap_{n=1}^{\infty} A_{x_n}\right) = P(A_x) = F_X(x).$$

Since the above holds for any sequence $\{x_n\}$ such that $\{x_n\} \rightarrow x$, we conclude that $\lim_{\delta \rightarrow 0} F_X(x + \delta) = F_X(x)$ for all x and so F_X is right continuous. □

Example 1.8. We will show that

$$F_X(x) = \frac{1}{1 + e^{-x}}$$

is the cdf of a rv X . It is enough to show that the function meets the requirements of Theorem 1.4. We have

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= 0 \quad \text{since} \quad \lim_{x \rightarrow -\infty} e^{-x} = \infty; \\ \lim_{x \rightarrow \infty} F_X(x) &= 1 \quad \text{since} \quad \lim_{x \rightarrow \infty} e^{-x} = 0. \end{aligned}$$

Also, the derivative of $F(x)$ is

$$F'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0,$$

showing that $F(x)$ is increasing. Furthermore, it is a continuous function, not only right-continuous. □

Now we can define discrete and continuous rvs more formally.

Definition 1.9. A random variable X is continuous if $F_X(x)$ is a continuous function of x . A random variable X is discrete if $F_X(x)$ is a step function of x . \square

1.5 Density and Mass Functions

1.5.1 Discrete Random Variables

Values of a discrete rv are elements of a countable set $\{x_1, x_2, \dots\}$. We associate a number $p_X(x_i) = P_X(X = x_i)$ with each value x_i , $i = 1, 2, \dots$, such that:

1. $p_X(x_i) \geq 0$ for all i ;
2. $\sum_{i=1}^{\infty} p_X(x_i) = 1$.

Note that

$$F_X(x_i) = P_X(X \leq x_i) = \sum_{x \leq x_i} p_X(x), \quad (1.1)$$

$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1}). \quad (1.2)$$

The function p_X is called the *probability mass function* (pmf) of the random variable X , and the collection of pairs

$$\{(x_i, p_X(x_i)), i = 1, 2, \dots\} \quad (1.3)$$

is called the *probability distribution* of X . The distribution is usually presented in either tabular, graphical or mathematical form.

Example 1.9. Consider Example 1.5, but now, we have n mice and we observe efficacy or no efficacy for each mouse independently. We are interested in the number of mice which respond positively to the applied drug candidate. If X_i is a random variable as defined in Example 1.5 for each mouse, and we may assume that the probability of a positive response is the same for all mice, then we may create a new random variable X as the sum of all X_i , that is,

$$X = \sum_{i=1}^n X_i.$$

X denotes k successes in n independent trials and it has a binomial distribution, which we denote by

$$X \sim \text{Bin}(n, p),$$

where p is the probability of success. The pmf of a binomially distributed rv X with parameters n and p is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n,$$

where n is a positive integer and $0 \leq p \leq 1$.

Exercise 1.4. For $n = 8$ and the probability of success $p = 0.4$ obtain mathematical, tabular and graphical form (pmf and cdf) of $X \sim \text{Bin}(n, p)$. □

1.5.2 Continuous Random Variables

Values of a continuous rv are elements of an uncountable set, for example a real interval. A cdf of a continuous rv is a continuous, nondecreasing, differentiable function. An interesting difference from a discrete rv is that for a $\delta > 0$

$$P_X(X = x) = \lim_{\delta \rightarrow 0} (F_X(x + \delta) - F_X(x)) = 0.$$

We define the *probability density function* (pdf) of a continuous rv as:

$$f_X(x) = \frac{d}{dx} F_X(x). \quad (1.4)$$

Hence,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt. \quad (1.5)$$

Similarly to the properties of the probability distribution of a discrete rv we have the following properties of the density function:

1. $f_X(x) \geq 0$ for all $x \in \mathcal{X}$;
2. $\int_{\mathcal{X}} f_X(x) dx = 1$.

Probability of an event that $X \in (-\infty, a)$, is expressed as an integral

$$P_X(-\infty < X < a) = \int_{-\infty}^a f_X(x) dx = F_X(a) \quad (1.6)$$

or for a bounded interval (b, c) as

$$P_X(b < X < c) = \int_b^c f_X(x) dx = F_X(c) - F_X(b). \quad (1.7)$$

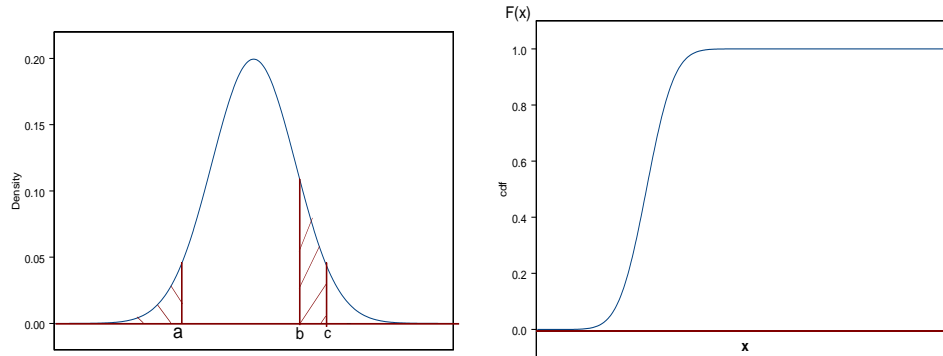


Figure 1.1: Distribution Function and Cumulative Distribution Function for $N(4.5, 2)$

Exercise 1.5. A certain river floods every year. Suppose that the low-water mark is set at 1 and a high-water mark X has distribution function

$$F_X(x) = \begin{cases} 0, & \text{for } x < 1; \\ 1 - \frac{1}{x^2}, & \text{for } x \geq 1. \end{cases}$$

1. Verify that $F_X(x)$ is a cdf.
2. Find the pdf of X .
3. Calculate the probability that the high-water mark is between 3 and 4.

1.6 Families of Distributions

Example 1.10. Normal distribution $\mathcal{N}(\mu, \sigma^2)$

The density function is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (1.8)$$

There are two parameters which tell us about the position and the shape of the density curve. □

There is an extensive theory of statistical analysis for data which are realizations of normally distributed random variables. This distribution is most common in applications, but sometimes it is not feasible to assume that what we observe can indeed be a sample from such a population. In Example 1.6 we observe time to a specific response to a drug candidate. Such a variable can only take nonnegative values, while the normal rv's domain is \mathbb{R} . A lognormal distribution is often used in such cases. X has a lognormal distribution if $\log X$ is normally distributed.

Exercise 1.6. Recall the following discrete distributions:

1. Uniform $\mathcal{U}(n)$;
2. Bern(p);
3. Geom(p);
4. Hypogeometric(n, M, N);
5. Poisson(λ).

Exercise 1.7. Recall the following continuous distributions:

1. Uniform $\mathcal{U}(a, b)$;
2. Exp(λ);
3. Gamma(α, λ);
4. Beta(α, β);

Note that all the distributions depend on some parameters, like p, λ, μ, σ or other. These values are usually unknown, so their estimation is one of the important problems in statistical analyzes. These parameters determine some characteristics of the shape of the pdf/pmf of a random variable. It can be location, spread, skewness etc.

We denote by P_θ the distribution function of a rv X depending on the parameter θ (a scalar or an s -dimensional vector).

Definition 1.10. *The **distribution family** is a set*

$$\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^s\}$$

defined on a common measurable space (Ω, \mathcal{A}) .

□

Example 1.11. Assume that in the efficacy Example 1.9 we have 3 mice. Then the sample space Ω consists of triples of “efficacious” or “non-efficacious” responses,

$$\Omega = \{(\omega_1, \omega_1, \omega_1), (\omega_1, \omega_1, \omega_2), \dots, (\omega_2, \omega_2, \omega_2)\}.$$

Take the σ -algebra \mathcal{A} as the power set on Ω .

The random variable defined as the number of successes in $n = 3$ trials has Binomial distribution with probability of success p .

$$X : \Omega \rightarrow \{0, 1, 2, 3\}.$$

Here p is the parameter which together with the distribution function defines the family on the common measurable space (Ω, \mathcal{A}) .

For $p = \frac{1}{2}$ we have the symmetric mass function

| | | | | |
|------------|---------------|---------------|---------------|---------------|
| $X = x$ | 0 | 1 | 2 | 3 |
| $P(X = x)$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |

while for $p = \frac{3}{4}$ the asymmetric mass function

| | | | | |
|------------|----------------|----------------|-----------------|-----------------|
| $X = x$ | 0 | 1 | 2 | 3 |
| $P(X = x)$ | $\frac{1}{64}$ | $\frac{9}{64}$ | $\frac{27}{64}$ | $\frac{27}{64}$ |

Both functions belong to the same family of binomial distributions with $n = 3$.

□

Exercise 1.8. Draw graphs of probability density functions of random variables having Exponential and Gamma distributions, each for several different parameter values.

□

1.7 Expected Values

Definition 1.11. The expected value of a function $g(X)$ is defined by

$$E(g(X)) = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{for a continuous r.v.,} \\ \sum_{j=0}^{\infty} g(x_j)p(x_j) & \text{for a discrete r.v.,} \end{cases}$$

and g is any function such that $E|g(X)| < \infty$.

□

Two important special cases of g are:

Mean $E(X)$, also denoted by EX , when $g(X) = X$,

Variance $E(X - E X)^2$, when $g(X) = (X - E X)^2$. The following relation is very useful while calculating the variance

$$E(X - E X)^2 = E X^2 - (E X)^2. \quad (1.9)$$

Example 1.12. Let X be a random variable such that

$$f(x) = \begin{cases} \frac{1}{2} \sin x, & \text{for } x \in [0, \pi], \\ 0 & \text{otherwise.} \end{cases}$$

Then the expectation and variance are following

- Expectation

$$E X = \frac{1}{2} \int_0^\pi x \sin x dx = \frac{\pi}{2}.$$

- Variance

$$\begin{aligned} \text{var}(X) &= E X^2 - (E X)^2 \\ &= \frac{1}{2} \int_0^\pi x^2 \sin x dx - \left(\frac{\pi}{2}\right)^2 \\ &= \frac{\pi^2}{4}. \end{aligned}$$

□

Example 1.13. Normal distribution, $X \sim \mathcal{N}(\mu, \sigma^2)$

We will show that the parameter μ is the expectation of X and the parameter σ^2 is the variance of X .

First we show that $E(X - \mu) = 0$.

$$\begin{aligned} E(X - \mu) &= \int_{-\infty}^{\infty} (x - \mu) f(x) dx \\ &= \int_{-\infty}^{\infty} (x - \mu) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= -\sigma^2 \int_{-\infty}^{\infty} f'(x) dx = 0. \end{aligned}$$

Hence $E X = \mu$.

Similar arguments give

$$\begin{aligned} E(X - \mu)^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\ &= -\sigma^2 \int_{-\infty}^{\infty} (x - \mu) f'(x) dx = \sigma^2. \end{aligned}$$

□

The following useful properties of the expectation follow from properties of integration (summation).

Theorem 1.5. *Let X be a random variable and let a , b and c be constants. Then for any functions $g(x)$ and $h(x)$ whose expectations exist we have:*

- (a) $E[ag(X) + bh(X) + c] = a E[g(X)] + b E[h(X)] + c$;
- (b) If $g(x) \geq h(x)$ for all x , then $E(g(X)) \geq E(h(X))$;
- (c) If $g(x) \geq 0$ for all x , then $E(g(X)) \geq 0$;
- (d) If $a \geq g(x) \geq b$ for all x , then $a \geq E(g(X)) \geq b$.

□

Exercise 1.9. Show that $E(X - b)^2$ is minimized by $b = E(X)$.

□

Variance of a random variable together with the mean are the most important parameters used in the theory of statistics. The following theorem is a result of the properties of the expectation function.

Theorem 1.6. *If X is a random variable with a finite variance, then for any constants a and b ,*

$$\text{var}(aX + b) = a^2 \text{var } X.$$

□