Chapter 1. Review of Probability Theory Notions

1 Random Variables

1.1 Types of Random Variables

Random variables, variables whose observed values are determined by chance, give a comprehensive quantitative overlook of random phenomena. Random variables are the fundamentals of modern Statistics.

To any experiment we assign its **sample space**, denoted by S, consisting of all its possible outcomes, called **elementary events** and denoted by $e_i,\ i \in \mathbb{N}$. An **event** is a subset of S (events are denoted by capital letters, $A, B, A_i,\ i \in \mathbb{N}$). A random variable is then a function of an outcome. The domain of a random variable is the sample space S. Its range can be the set of all real numbers $\mathbb{R}$, or only the positive numbers $(0, \infty)$, or the integers $\mathbb{Z}$, or the interval $(0, 1)$, or the set $\{0, 1, 2\}$, etc., depending on what possible values the random variable can potentially take. Once an experiment is completed, and the outcome e is known, the value of the random variable $X(e)$ becomes determined.

Definition 1.1. *Let S be the sample space of some experiment. A **random variable** is a function $X : S \to \mathbb{R}$ satisfying the property that for every $x \in \mathbb{R}$, the probability of the event*

$$(X \le x) \ := \ \{e \in S \mid X(e) \le x\} \ \subseteq \ S \tag{1.1}$$

exists.

- *if the set of values that it takes, $X(S)$, is at most countable in $\mathbb{R}$, then X is a **discrete random variable** (quantities that are counted);*
- *if $X(S)$ is a continuous subset of $\mathbb{R}$ (an interval), then X is a **continuous random variable** (quantities that are measured).*

Example 1.2. Consider the experiment of tossing 3 fair coins and counting the number of heads. Obviously, the same model suits the number of girls in a family with 3 children, the number of 1's in a random binary code consisting of 3 characters, the number of 6's obtained when 3 dice are rolled, etc.

Let X be the number of heads (girls, 1's, 6's). Prior to an experiment, its value is not known. But it is clear that X has to be an integer between 0 and 3. Since assuming each value is an event, we can compute its probabilities (we denote by $H-$"heads" and by $T-$"tails").

$$\begin{aligned}
P(X=0) &= P(TTT) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8},\\
P(X=1) &= P(HTT)+P(THT)+P(TTH) = \frac{3}{8},\\
P(X=2) &= P(HHT)+P(HTH)+P(THH) = \frac{3}{8},\\
P(X=3) &= P(HHH) = \frac{1}{8}.
\end{aligned}$$

Summarizing,

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

This table contains everything that is known about random variable X prior to the experiment. Before we know the outcome, we cannot tell what X equals to. However, we can list all the possible values of X and determine the corresponding probabilities.

Also, we notice that the sum of all the probabilities on the second row is 1. This is no coincidence. For every outcome, the variable X takes one and only one value x. This makes events $\{X=x\}_{x=\overline{0,3}}$ disjoint and exhaustive (i.e., they form a *partition* of the sample space) and therefore,

$$\sum_{x=0}^{3} P(x) = \sum_{x=0}^{3} P(X=x) = 1.$$

Example 1.3. A fair die is rolled. Let X denote the number that shows on the die and Y the number of 6's rolled.
Then X can take the values $1, 2, 3, 4, 5$ and 6 and they are equally probable, i.e.,

$$P(X=i) = \frac{1}{6},\ i=1,2,\ldots,6.$$

The variable Y can only take the values 0 or 1, with probabilities

$$\begin{aligned}
P(Y=0) &= \frac{5}{6},\\
P(Y=1) &= \frac{1}{6}.
\end{aligned}$$

Example 1.4. A plane can land at *any* time between 7 and 8 a.m. Let X denote the landing time of the plane. Since time is a quantity that varies *continuously*, we cannot pinpoint any single value on the interval $[7, 8]$, i.e.

$$P(X = t) \;=\; 0, \forall t \in [7, 8] \text{ (or any } t \in \mathbb{R}, \text{ for that matter).}$$

However, since it's equally likely for the plane to land at *any* time in that interval, we can definitely say that, for instance,

$$\begin{aligned}
P(7 \le X \le 7{:}30) &= P(7 < X < 7{:}30) = \frac{1}{2},\\
P(7 \le X \le 7{:}20) &= P(7{:}20 \le X \le 7{:}40) = P(7{:}40 \le X \le 8) = \frac{1}{3},\\
P(7 \le X \le 7{:}15) &= P(7{:}45 \le X \le 8) = \frac{1}{4}, \text{ etc.}
\end{aligned}$$

1.2 PDF and CDF

For each random variable, discrete or continuous, there are two important functions associated with it. The first basically defines or describes the random variable, and the second one helps us compute probabilities about the random variable.

Definition 1.5. *Let $X : S \to \mathbb{R}$ be a discrete random variable. The* ***probability distribution/density function (pdf)****, or* ***probability mass function (pmf)*** *of X is an array of the form*

$$X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}, \tag{1.2}$$

where $x_i \in \mathbb{R},\ i \in I$, are the values that X takes and $p_i = P(X = x_i)$ are the probabilities that X takes each value x_i.

Definition 1.6. *The* ***cumulative distribution function (cdf)*** *of a (discrete or continuous) random variable X, is the function $F = F_X : \mathbb{R} \to \mathbb{R}$, defined by*

$$F(x) \;=\; P(X \le x). \tag{1.3}$$

Example 1.7. For the random variable X in Example 1.2, the pdf is

$$X\begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix} = \begin{pmatrix} k \\ C_3^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{3-k} \end{pmatrix}_{k=\overline{0,3}}.$$

Its cdf is the piecewise-defined function

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1/8 = 0.125, & \text{if } 0 \le x < 1 \\ 1/2 = 0.5, & \text{if } 1 \le x < 2 \\ 7/8 = 0.875, & \text{if } 2 \le x < 3 \\ 1, & \text{if } x \ge 3 \end{cases}$$

Below see the graphs of the two functions (Figure 1). White circles denote excluded points.

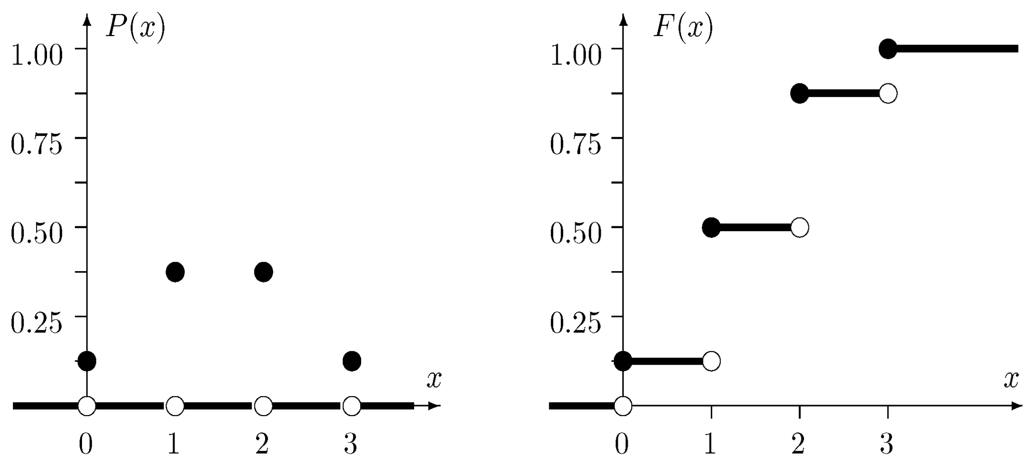


Fig. 1: PDF and CDF, Example 1.2

The pdf of a discrete random variable has the following properties:

- all values $x_i, i \in I$, are distinct and listed in increasing order and all probabilities $p_i > 0, i \in I$;
- $\sum_{i \in I} p_i = 1$.

The cdf has the following properties:

- if $a < b$ are real numbers, then $P(a < X \le b) = F(b) - F(a)$;

- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$;
- $P(X < x) = F(x - 0) = \lim_{y \nearrow x} F(y)$ and $P(X = x) = F(x) - F(x - 0)$;
- If X is discrete, then $F(x) \;=\; \sum_{x_i \le x} p_i$.

Proposition 1.8. *Let X be a continuous random variable with cdf $F : \mathbb{R} \to \mathbb{R}$. Then F is absolutely continuous, i.e. there exists a real function $f : \mathbb{R} \to \mathbb{R}$, such that*

$$F(x) \;=\; \int_{-\infty}^{x} f(t)\, dt, \tag{1.4}$$

for all $x \in \mathbb{R}$.

Definition 1.9. *Let X be a continuous random variable. Then the function f from Proposition 1.8 is called the* ***probability density function (pdf)*** *of X.*

The pdf of a continuous random variable has the following properties:

- $f(x) \ge 0$, for all $x \in \mathbb{R}$;
- $\int_{\mathbb{R}} f(t)dt = 1$.

The cdf of a continuous random variable has the following properties:

- $P(X = x) = 0, P(X < x) = P(X \le x) = F(x)$ and

  $P(a < X \le b) = P(a < X \le b) \;=\; P(a < X < b) \;=\; P(a \le X \le b) = \int_{a}^{b} f(t)\, dt$;
- $F'(x) = f(x)$, for all $x \in \mathbb{R}$.

Remark 1.10. So, probabilities involving continuous random variables can be computed by integrating the density function over the given sets. Furthermore, recall from Calculus that the integral $\int_a^b f(x)dx$ of a non-negative function f equals the area below the density curve f between the points $x = a$ and $x = b$. Therefore, geometrically, probabilities are represented by areas (see Figure 2). This aspect will be important later on. Also, the *total* area under the graph of a density function is equal to 1.

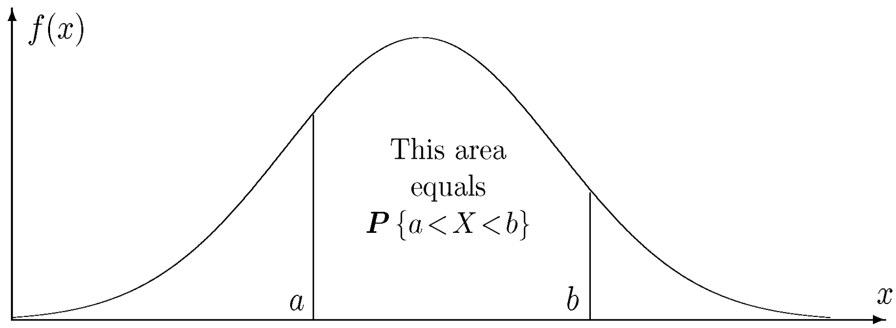


Fig. 2: Probability as area

2 Some Common Distributions (Probability Laws)

2.1 Some 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

$$X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}. \tag{2.1}$$

Notice that the random variable Y from Example 1.3 has a $Bern(1/6)$ distribution. 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 (unid) with parameter $m \in \mathbb{N}$, if its pdf is

$$X \begin{pmatrix} k \\ \frac{1}{m} \end{pmatrix}_{k=\overline{1,m}}. \tag{2.2}$$

The random variable X in Example 1.3, 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

$$X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}. \tag{2.3}$$

This distribution corresponds to the *Binomial model*. Given n Bernoulli trials with probability of success p, let X denote the number of successes. Then $X \in B(n, p)$. The random variable X in Example 1.2 has a Binomial $B(3, 1/2)$ distribution. Also, notice that the Bernoulli distribution is a particular case of the Binomial one, for $n = 1$, $Bern(p) = B(1, p)$.

Geometric Distribution $Geo(p)$

A random variable X has a Geometric distribution ($\boxed{\text{geo}}$) with parameter $p \in (0, 1)$, if its pdf is given by

$$X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k=0,1,\dots}. \tag{2.4}$$

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

Remark 2.1. In a Geometric model setup, one might count the number of *trials* (not *failures*) needed to get the 1^{st} success. Of course, if X is the number of failures and Y the number of trials, then we simply have $Y = X + 1$ (the number of failures plus the one success). The variable Y is said to have a Shifted Geometric distribution with parameter $p \in (0, 1)$ ($Y \in SGeo(p)$). Its pdf is

$$X \begin{pmatrix} k \\ pq^{k-1} \end{pmatrix}_{k=1,2,\dots}. \tag{2.5}$$

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

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

$$X \begin{pmatrix} k \\ \dfrac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k=0,1,\dots} \tag{2.6}$$

Poisson random variables arise in connection with so-called Poisson *processes,* processes that involve observing discrete events in a continuous interval of time, 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.

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".

2.2 Some Common Continuous Distributions

Uniform Distribution $U(a, b)$

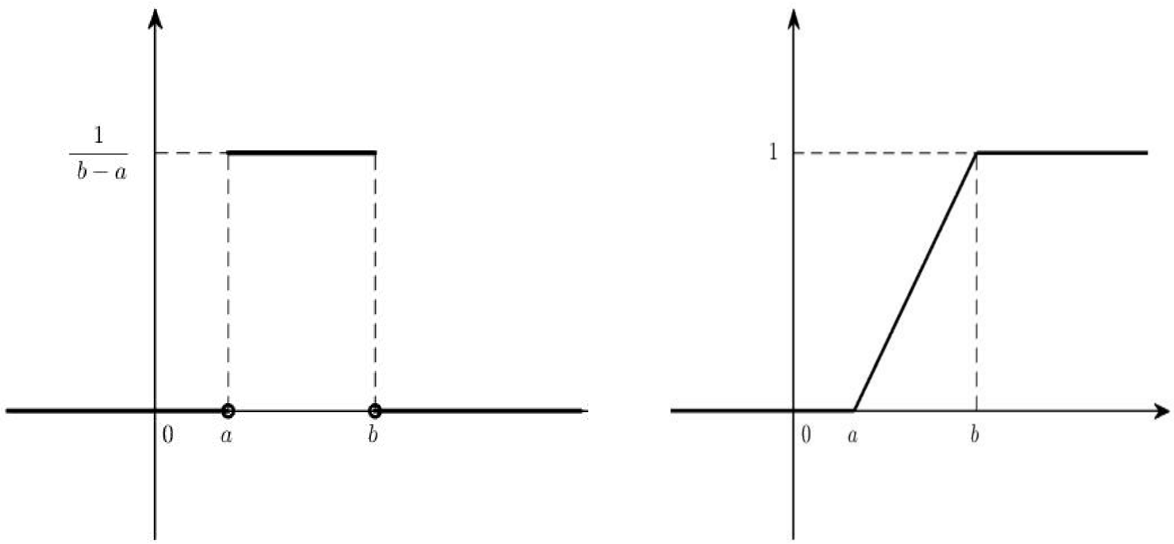
A random variable X has a Uniform distribution ($\boxed{\text{unif}}$) with parameters $a, b \in \mathbb{R},\ a < b$, if its pdf is

$$f(x) \quad = \quad \begin{cases} \dfrac{1}{b-a}, & \text{if } \ x \in [a, b] \\ 0, & \text{if } \ x \notin [a, b]. \end{cases} \tag{2.7}$$

Then, by (1.4), its cdf is

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad = \quad \begin{cases} 0, & \text{if } \ x \le a \\ \dfrac{x-a}{b-a}, & \text{if } \ a < x \le b \\ 1, & \text{if } \ x \ge b\,. \end{cases} \tag{2.8}$$

Remark 2.2.
1. The Uniform distribution is used when a variable can take *any* value in a given interval, equally probable. For example, locations of syntax errors in a program, birthdays throughout a year, etc. The random variable X in Example 1.4 has a Uniform $U(7, 8)$ distribution.
2. A special case is that of a **Standard Uniform Distribution**, where $a = 0$ and $b = 1$. The pdf and



(a) Density Function (pdf) (b) Cumulative Distribution Function (cdf)

Fig. 3: Uniform Distribution

cdf are given by

$$f_U(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases} \quad , \quad F_U(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & x \geq 1\,. \end{cases} \tag{2.9}$$

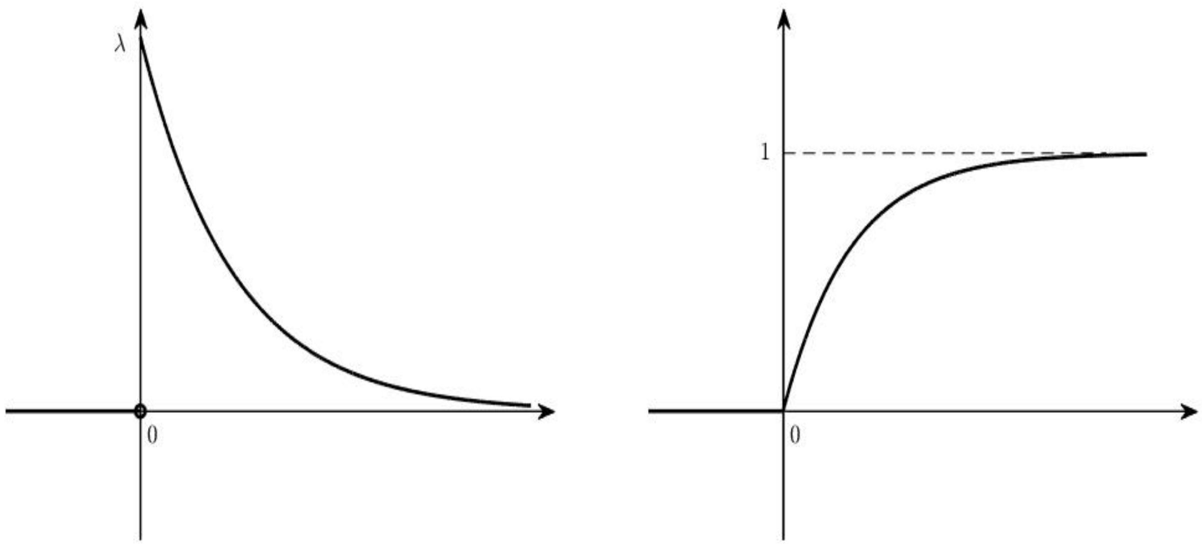
Standard Uniform variables play an important role in stochastic modeling; in fact, *any* random variable, with any thinkable distribution (discrete or continuous) can be generated from Standard Uniform variables.

Exponential Distribution $Exp(\lambda)$

A random variable X has an Exponential distribution (exp) with parameter $\lambda > 0$, if its density function and cdf are given by

$$f(x) \;=\; \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad \text{and} \quad F(x) \;=\; \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \;, \tag{2.10}$$

respectively.



(a) Density Function (pdf) (b) Cumulative Distribution Function (cdf)

Fig. 4: Exponential Distribution

Remark 2.3.
1. The Exponential distribution is often used to model *time*: lifetime, waiting time, halftime, interarrival time, failure time, time between rare events, etc. In a sequence of rare events (where the number of rare events has a Poisson distribution), the time between two consecutive rare events (as well as the time of the occurrence of the first rare event) is Exponential. The parameter λ represents the frequency of rare events, measured in time^{-1}.
2. A word of **caution** here: The parameter μ in Matlab (where the Exponential pdf is defined as $\frac{1}{\mu}e^{-\frac{1}{\mu}x}, x \geq 0$) is actually $\mu = 1/\lambda$. It all comes from the different interpretation of the "frequency". For instance, if the frequency is "2 per hour", then $\lambda = 2/\mathrm{hr}$, but this is equivalent to " one every half an hour", so $\mu = 1/2$ hours. The parameter μ is measured in time units.
3. The Exponential distribution is a special case of a more general distribution, the $Gamma(a, b)$, $a, b > 0$, distribution (gam). The Gamma distribution models the *total* time of a multistage scheme.
4. If $\alpha \in \mathbb{N}$, then the sum of α independent $Exp(\lambda)$ variables has a $Gamma(\alpha, 1/\lambda)$ distribution.

3 Numerical Characteristics of Random Variables

The distribution of a random variable or a random vector, the full collection of related probabilities, contains the entire information about its behavior. This detailed information can be summarized in a few vital numerical characteristics describing the average value, the most likely value of a random variable, its spread, variability, etc. These are numbers that will provide some information about a random variable or about the relationship between random variables.

3.1 Expectation

Definition 3.1.

(i) *If* $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$ *is a discrete random variable, then the* ***expectation (expected value, mean value)*** *of* X *is the real number*

$$E(X) \;=\; \sum_{i\in I} x_i P(X=x_i) \;=\; \sum_{i\in I} x_i p_i, \tag{3.1}$$

if it exists (i.e., the series is absolutely convergent).

(ii) *If* X *is a continuous random variable with density function* $f:\mathbb{R}\to\mathbb{R}$, *then its* ***expectation (expected value, mean value)*** *is the real number*

$$E(X) \;=\; \int_{\mathbb{R}} x f(x)dx, \tag{3.2}$$

if it exists (i.e., the integral is absolutely convergent).

Remark 3.2.

1. The expected value can be thought of as a "long term" average value, a number that we *expect* the values of a random variable to stabilize on.

2. It can also be interpreted as a point of equilibrium, a center of gravity. In the discrete case, if we imagine the probabilities p_i to be weights distributed at the points x_i, then $E(X)$ would be the point that holds the whole thing in equilibrium. In fact, notice that formula (3.1) is *actually* a weighted mean. Consider a random variable with pdf

$$X\begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}.$$

Observing this variable many times, we shall see $X = 0$ about 50% of times and $X = 1$ about 50% of times. The average value of X will then be close to 0.5, so it is reasonable to have $E(X) = 0.5$, which we get by formula (3.1):

$$E(X) \;=\; 0 \cdot 0.5 + 1 \cdot 0.5 \;=\; 0.5.$$

Now, suppose that $P(X = 0) = 0.75$ and $P(X = 1) = 0.25$, i.e its pdf is now

$$X \begin{pmatrix} 0 & 1 \\ 0.75 & 0.25 \end{pmatrix}.$$

Then, in a long run, X is equal to 1 only $1/4$ of times, otherwise it equals 0. Therefore, in this case, $E(X) = 0.25$, which is what we obtain by formula (3.1):

$$E(X) \;=\; 0 \cdot 0.75 + 1 \cdot 0.25 \;=\; 0.25.$$

The expected value as a center of gravity is illustrated in Figure 5.

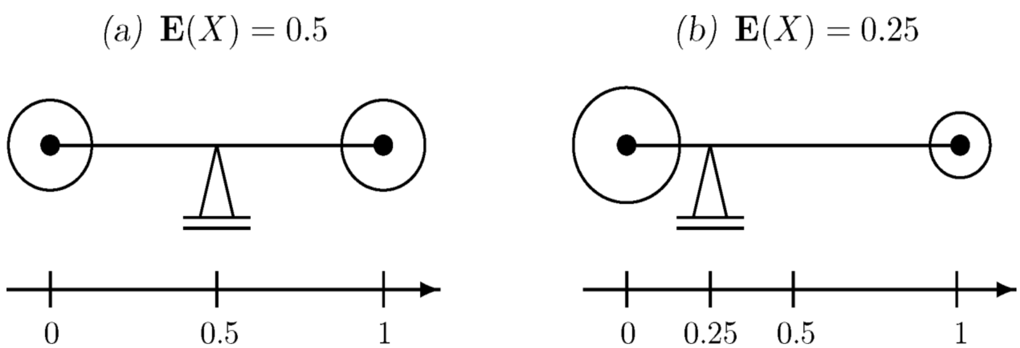


Fig. 5: Expectation as a center of gravity

The same interpretation would go for the continuous case, only there the "weight" would be continuously distributed, according to the density function f.

3. If $f : \mathbb{R} \to \mathbb{R}$ is a measurable function, then

$$E(h(X)) \;=\; \sum_{i \in I} h(x_i) p_i, \tag{3.3}$$

if X is discrete and

$$E(h(X)) \;=\; \int_{\mathbb{R}} h(x)f(x)\,dx, \tag{3.4}$$

if X is continuous.

Example 3.3. Consider the random variable X in Example 1.3, the number that shows on a die when it is rolled. What is its expected value?

Solution.

$$E(X) \;=\; \sum_{i\in I} x_i p_i \;=\; \sum_{i=1}^{6} i\cdot\frac{1}{6} \;=\; \frac{1}{6}\sum_{i=1}^{6} i \;=\; \frac{1}{6}\cdot\frac{6\cdot 7}{2} \;=\; \frac{7}{2}.$$

■

Remark 3.4. In general, the expected value of a Discrete Uniform distribution $U(m),\ m\in\mathbb{N}$, with pdf (2.2) is

$$E(X) \;=\; \sum_{i=1}^{m} i\cdot\frac{1}{m} \;=\; \frac{1}{m}\sum_{i=1}^{m} i \;=\; \frac{1}{m}\cdot\frac{m(m+1)}{2} \;=\; \frac{m+1}{2}.$$

Example 3.5. Find the expected value of a (continuous) Uniform distribution $U(a,b),\ a<b$.

Solution. If X can take *any* value in the interval $[a,b]$, equally probable, then in the long run, it is just as likely to take values at the beginning of the interval, as it is to take the ones towards the end of $[a,b]$. So they would average out at the value right in the middle, i.e. the midpoint of the interval, $\dfrac{a+b}{2}$.

Indeed, if X has a pdf given by (2.7), then by formula (3.2), we have

$$\begin{aligned} E(X) \;&=\; \int_{\mathbb{R}} x f(x)dx \;=\; \int_a^b x\cdot\frac{1}{b-a}dx \;=\; \frac{1}{b-a}\int_a^b x dx \\ &=\; \frac{1}{b-a}\cdot\frac{1}{2}x^2\Big|_a^b \;=\; \frac{1}{b-a}\cdot\frac{b^2-a^2}{2} \;=\; \frac{a+b}{2}. \end{aligned}$$

■

The expected value of a (discrete or continuous) random variable has the following properties:

- $E(aX+b) = aE(X)+b,$ for all $a,b\in\mathbb{R}$;

- $E(X+Y) = E(X) + E(Y)$;
- If X and Y are independent, then $E(X \cdot Y) = E(X)E(Y)$;
- If $X(e) \leq Y(e)$ for all $e \in S$, then $E(X) \leq E(Y)$.

3.2 Variance (Dispersion) and Standard Deviation

Knowledge of the mean value of a random variable is important, but that knowledge *alone* can be misleading. Suppose two patients in a hospital, X and Y, have their pulse (number of heartbeats per minute) checked every day. Over the course of time, they each have a mean pulse of 75, which is considered healthy. But, for patient X the pulse ranges between 70 and 80, while for patient Y, it oscillates between 40 and 110. Obviously, the second patient might have some serious health problems, which the *expected value alone* would not show. So, next, we define some measures of variability.

Definition 3.6. *Let X be a random variable. The* ***variance (dispersion)*** *of X is the number*

$$V(X) \;=\; E\Big[\Big(X - E(X)\Big)^2\Big], \tag{3.5}$$

if it exists. The value $\sigma(X) = \mathrm{Std}(X) = \sqrt{V(X)}$ is called the ***standard deviation*** *of X.*

Variance (and standard deviation) measure the amount of variability (spread) in the values that a random variable takes, with large values indicating a wide spread of values and small values meaning more closely knit values. The standard deviation brings the numbers to the same "level" (e.g., measurement units), while the variance gives the squares of those numbers.

Properties of the variance:

- $V(X) = E(X^2) - (E(X))^2$ (a more efficient computational formula);
- $V(aX+b) = a^2 V(X)$, for all $a, b \in \mathbb{R}$;
- If X and Y are independent, then $V(X+Y) = V(X) + V(Y)$;
- If $X = b$ is a constant random variable (i.e. it only takes that one value with probability 1), then $V(X) = 0$, which is to be expected (the variable X does not vary *at all*).

3.3 Median

Definition 3.7. *The **median** of a random variable X is a real number M that is exceeded with probability no greater than 0.5 and is preceded with probability no greater than 0.5. That is, M is such that*

$$\begin{aligned} P(X > M) &\leq 1/2 \\ P(X < M) &\leq 1/2, \end{aligned}$$

or, equivalently,

$$P(X < M) \leq 1/2 \leq P(X \leq M).$$

Comparing the mean $E(X)$ and the median M, one can tell whether the distribution of X is right-skewed ($M < E(X)$), left-skewed ($M > E(X)$), or symmetric ($M = E(X)$).

For *continuous* distributions, since $P(X < M) = P(X \leq M) = F(M)$, computing a population median reduces to solving one equation:

$$\left\{ \begin{array}{lcccc} P(X > M) & = & 1 - F(M) & \leq & 0.5 \\ P(X < M) & = & F(M) & \leq & 0.5 \end{array} \right. \Rightarrow F(M) = 0.5.$$

The Uniform distribution $U(a,b)$ has cdf

$$F(x) = \frac{x-a}{b-a},\ x \in [a,b].$$

Solving the equation $F(M) = (M-a)/(b-a) = 0.5$, we find the median

$$M = \frac{a+b}{2},$$

which is also the expected value $E(X)$. That should be no surprise, knowing that the Uniform distribution is symmetric.

For the Exponential distribution $\mathrm{Exp}(\lambda)$, the cdf is

$$F(x) = 1 - e^{-\lambda x},\ x > 0.$$

Solving $F(M) = 1 - e^{-\lambda M} = 0.5$, we get

$$M = \frac{\ln 2}{\lambda} \approx \frac{0.6931}{\lambda} < \frac{1}{\lambda} = E(X),$$

since the Exponential distribution is right-skewed.

These two cases are depicted in Figure 6.

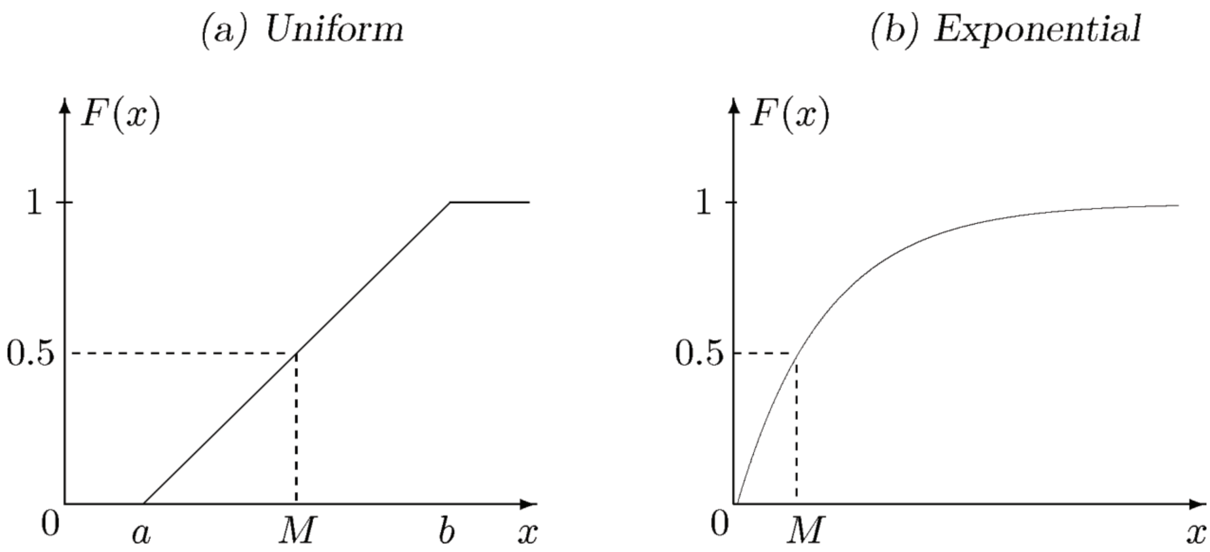


Fig. 6: Median for Continuous Distributions

For *discrete* distributions, the equation $F(x) = 0.5$ has either a whole interval of roots or no roots at all (see Figure 7).

In the first case, the Binomial distribution $B(5, 0.5)$, with $p = 0.5$, successes and failures are equally likely. Pick, for example, $x = 2.2$ in the interval $(2, 3)$. Having fewer than 2.2 successes (i.e., at most 2) has the same chance as having more than 2.2 successes (i.e., at least 3 successes). Therefore, $X < 2.2$ with the same probability as $X > 2.2$, which makes $x = 2.2$ a central value, a median. We can say that $x = 2.2$ (and *any other* value $x \in (2, 3)$) splits the distribution into two equal parts. So, it is a median.

In the other case, the Binomial distribution $B(5, 0.4)$ with $p = 0.4$, we have

$$\begin{aligned} F(x) < 0.5 \quad &\text{for} \quad x < 2, \\ F(x) > 0.5 \quad &\text{for} \quad x \geq 2, \end{aligned}$$

but there is no value of x with $F(x) = 0.5$. Then, $M = 2$ is the median. Seeing a value on either side of $x = 2$ has probability less than 0.5, which makes $x = 2$ a center value.

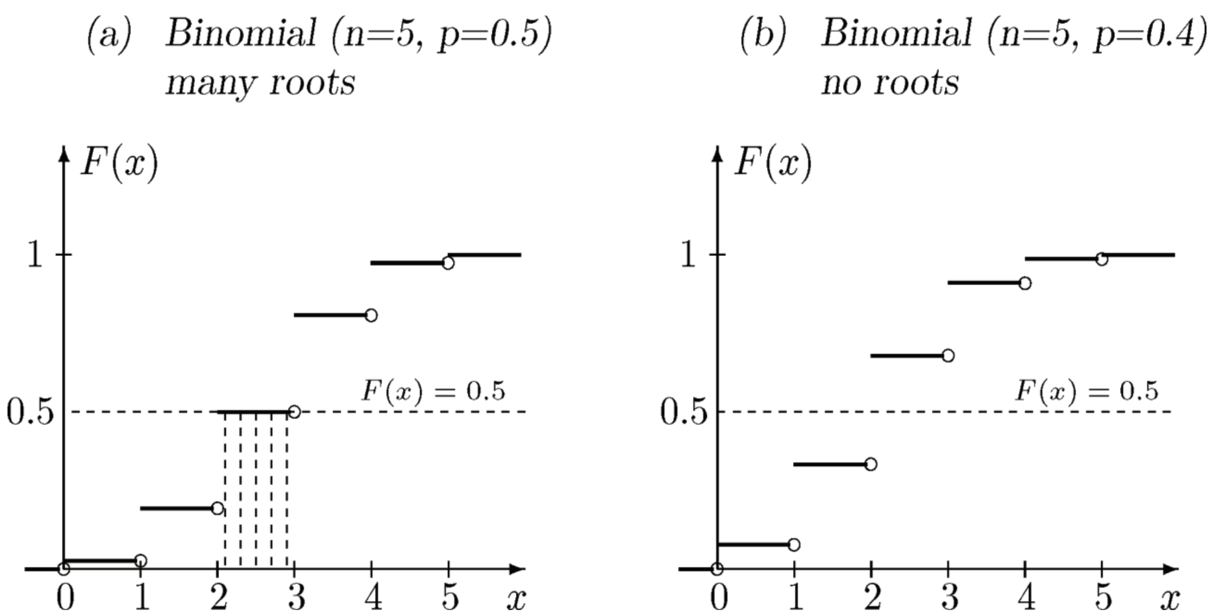


Fig. 7: Median for Discrete Distributions