Basic Probability Theory

Random Variables & Probability Distributions

Kiranmoy Chatterjee*

1 Motivation: Random Variable & Probability Distribution

So far it is known to us that there are situations in which not all subsets of sample space (generated from a random experiment) can become an event, and that difficulties may arises in case of uncountable sample space.

Example 1. If age of each person from a population is recorded, a complete description of the experiment would involve a probability space with 200 million (population size) points. A convenient way to describe the experiment is to group the data, for example, into 10 years intervals. We may define a function $g(x), x = 5, 10, 15, 25, \ldots$, so that $g(x)$ is the number of individuals, say in million, between the ages $x - 5$ and $x + 5$ years (see Figure 1).

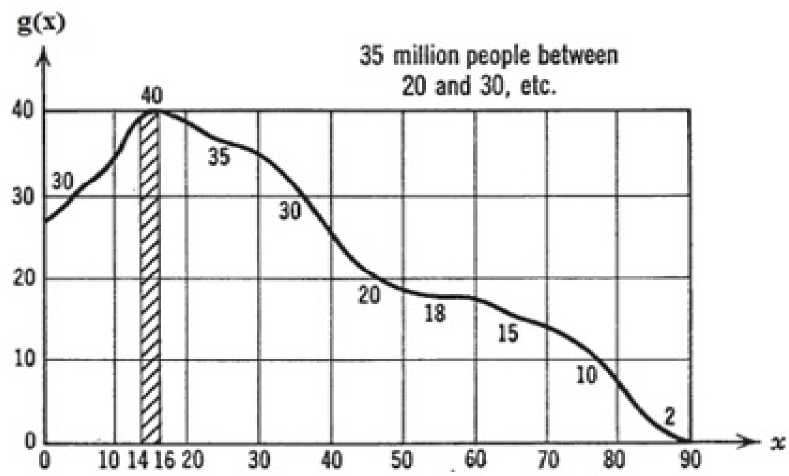


FIGURE 1. Age Statistics.

For example, if $g(15) = 40$, there are 40 million people in the age group $10 - 20$ or, on an average, 4 million people per year over the 10 year span. Note that age of a person, selected at random from the said population, is a random variable which theoretically assumes value in $(0, \infty)$. Now, if we want the probability that age of the randomly picked person will be in between

*Assistant Professor, Department of Statistics, Bidhannagar College, Salt Lake, Kolkata-700064; E-mail: *kiranmoy07@gmail.com*; Target Students: B.Sc.(Hons.) in Statistics; CBCS Paper Code: STSACOR03T

14 and 16 years, we can get a reasonable figure by from the average number of people per year, i.e.,

$$[g(15)/10] \times 2 = 8 \text{ million}.$$

Therefore, dividing by the total population size, hence one can obtain the required probability as (8/200) or 0.04.

If we connect the values of $g(x)$ buy a smooth curve, essentially what we are doing is evaluating $\frac{1}{200}\int_{14}^{16}[g(x)/10]dx$ to find the probability that a person picked at random will be between 14 and 16 years. Hence, in general, the problem is to estimate the probability that a person belongs to the age interval (a, b) and this probability can be estimated by $\int_a^b [g(x)/2000]dx$; thus $g(x)/2000$ is the 'probability density' per unit age at age x. Thus, we are led to the idea of assigning probabilities by means of an integral. We are taking the sample space as a subset of real line $\mathbb{R}$, and assigning the probability of an event A as $P(A) = \int_A f(x)dx$, where f is a non-negative real-valued function representing 'probability density' with domain $\mathbb{R}$, and $A \subseteq \mathbb{R}$. This $f(x)$ actually gives the 'measurement of chance' that age of a randomly picked person will be x year. Here, the 'measurement' is based on 'probability density' which may not lie in $[0, 1]$. If the f is continuous (or, piecewise continuous), then the A should be at least an interval and the integral is Riemann integral.

Otherwise, there may be another type of situation when f has mass only at countable number of discrete points of x. The following is such an example.

Example 2. Suppose in an big amusement carnival, a competitor is entitled for a prize if he throws a ring on a peg from a certain distance. One who meet the target in minimum attempts will be given the prize. Maximum 10 attempts is allowed. Suppose a very large no. of competitors, say 7500 (population size), are contesting in the competition. A convenient way to describe the possible result is to summarize the data into frequency table where frequency (i.e. no. of competitors) against each value of the number of attempts required to get the success (i.e. throwing a ring on a peg). Therefore, we may define a function $g(x), x = 1, 2, 3, 4, \ldots, 10$ so that $g(x)$ is the number of individuals who attained the success in x attempts.

For example, if $g(3) = 1500$, there are 150 people who attained the success. Note that no. of attempts required to attain the success by an individual, selected at random from the said population, is certainly a random variable which theoretically assumes value in $\{1, 2, 3, 4, \ldots, 10\}$. Now, if we want the probability that no. of attempts required to attain the success for a randomly picked person will be in between 3 to 5, we can get it by taking sum of $g(3), g(4), g(5)$ and then dividing by the total population size, i.e.,

$$[g(3) + g(4) + g(5)]/7500 = [1500 + 1900 + 1600]/7500 = 0.667.$$

If we assume the values of $g(x)/7500$ for each of $x = 1, 2, 3, 4, \ldots$, essentially what we are doing is evaluating $\sum_{x=3}^{5} g(x)/7500$ to find the probability that no. of attempts required to attain the success for a randomly picked person will be in between 3 to 5. Hence, in general, the problem is to estimate the probability that for a person, no. of attempts required to attain the success lies in $\{a_1, \ldots, a_2\}$ and this probability can be estimated by $\sum_{x=a_1}^{a_2} g(x)/7500$; thus $g(x)/7500$ itself represents the 'probability' for each value of x and we denote $g(x)/7500$ by $f(x)$ here. We are taking the sample space as a subset of set of positive integers, and assigning the probability of an event A as $P(A) = \sum_{x \in A} f(x)dx$, where f is a non-negative real-valued function representing 'probability'. This $f(x)$ actually gives the 'measurement of chance' that no. of attempts required to attain the success for a randomly picked person will be x. Here, the 'measurement' is based on 'probability' which must lie in $[0, 1]$. Here, the f has mass only at finite or countably infinite number of discrete points of x. In that cases, we assign the probability of an event A as $P(A) = \sum_{x \in A} f(x)$, where the 'measurement of chance' given by $f(x)$ is based on 'probability' and hence, the $f(x)$ at any x must lie in $[0, 1]$.

In both of the above cases, there are several immediate questions, such as, what type of functions f are allowed and on which it is defined, what we mean by $\int_A f(x)dx$ or $\sum_{x \in A} f(x)$ and how do we know that the resulting $P(A)$ is a probability.

Now, let us start with the idea of random variable, its definition and how it is related to our basic aim: "computation of $P(A)$, where A is any arbitrary event". Here, $P(\cdot)$ is the 'probability function' defined according to the Axiomatic definition (*see* the Appendix). This part is discussed in Section 2. Next, in Section 3, we discuss our basic aim: "computation of $P(A)$" based on the assumed probability distribution on the random variable defined in the context.

2 Random Variable

A random variable is a real-valued function defined on sample space associated with some random experiment. That is, it associates a numerical value to each elementary outcome in the sample space.

Example 1. Consider tossing a coin n times. We may be interested on the number of heads appeared when the coin is not known to be unbiased.

The random experiment, mentioned in the above example, yields any one of the 2^n possible outcomes. So, the sample space $\Omega_1 = \{\omega : \omega = (i_1, i_2, ., i_n), i_j \in \{H, T\} \text{ for } j = 1, 2, \ldots, n\}$. If X denotes the number of heads obtained in the experiment. Therefore, the possible values of X are the non-negative integers up to n, i.e. the set $(0, 1, 2, \ldots, n)$. This X is a random variable.

Example 2. Consider a stock price which moves each day either up one unit or down one unit, and suppose its initial value is Rs. 100. One may be interested in the no. of days the stock may take to hit the price Rs. 150.

The daily stock prices in the random experiment, mentioned in Example 2, take the values from finite countable set, i.e. the sample space $\Omega_2 = \{0, 1, 2, 3, \ldots, 150\}$. If Y denotes the first time the value of the stock hits Rs. 150. Therefore, Y takes countably infinite possible values starting from (150-100) or 50, i.e. the set $\{50, 51, 52, \ldots\}$. This Y is a random variable.

Example 3. Consider a new light bulb is switched on until it expires. One is commonly interested in the lifetime T of the light bulb.

In the last example, if we can measure time with infinite precision, then the possible values of T are the non-negative real numbers $[0, \infty)$. This is an uncountable set: there is no way to enumerate $[0, \infty)$ in a sequence. We will have to treat random variables of this type separately from the random variables which take values in a countable set, e.g. the random variables discussed in Examples 1 and 2. While in practice time can only be measured up to finite precision and consequently the possible values of T will in fact be countable, it is still more convenient mathematically to make the idealization that all values in $[0, \infty)$ are possible, and we will do so.

Why we need random variables?

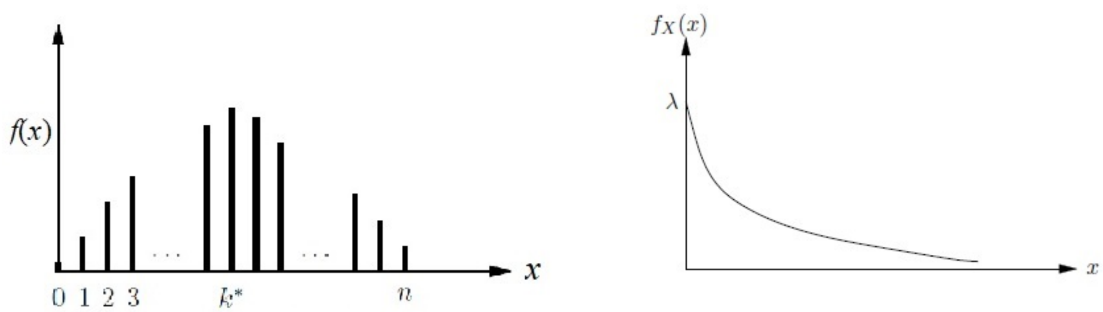
There are several aspects of the introduction of random variables in probability theory.

Firstly, if the possible outcomes of a random experiment are countably infinite, or finite but not equally likely (i.e. violation of the classical probability assumption), then computation of probability of an event A (e.g. the event stated in Example 1 or 2), i.e. $P(A) = \sum_{\omega \in A} P(\omega)$ would be difficult. Therefore, one needs to assume probability structures, say $f(\omega)$, on each of all possible elementary events $\omega \in \Omega$, where f is a real-valued function subject to the condition $\sum_{\omega \in \Omega} f(\omega) = 1$ (Axiom (ii) in Axiomatic Definition of Probability). Here, Ω refers the sample space associated with the random experiment. This f represents the probability distribution of possible outcomes over Ω. In particular, when the possible outcomes are equally likely, $f(\omega) = 1/m$, where $m = |\Omega| < \infty$, the cardinality of finite Ω. However, assuming such a real valued probability function on any Ω, especially for infinite Ω (i.e. $|\Omega| = \infty$), may not be reasonable in most of the cases. This can be done in more convenient way if we take mapping of Ω on $\mathbb{R}$, the real line.

Secondly, in a random experiment, we are always concerned about the chance of occurrence of an event that can be represented by some numerical entity. See all the 3 examples stated above.

Random variable is nothing but such a mapping or function. After defining a random variable, say X, the basic aim of "computing P(A)" changes to "computing $P(X \in A)$".

To compute $P(X \in A)$, at first, one have to reasonably assume a probability structure $\{f(x) : x \in \mathbb{R}\}$ on different possible mapped values $x(\in \mathbb{R})$ of X. This $f(x)$ gives the measurement of chance that the r.v. X takes the value x. This is the function f what we talked about in the Section 1. How to assume a probability structure f on $\mathbb{R}$ (i.e., what type of functions f are allowed as a probability structure on r.v.) and therefore, computation of $P(X \in A)$ will be discussed in Section 3.



This structure represents f(x) shows higher probability densities at the middle values of x and the densities are symmetrically distributed around its middle most value (this structure is relevant to the r.v. stated in Example 1)

This structure represents $f(x)$ shows higher probability densities at the smaller values of x and densities are gradually decaying as x increases (this structure is relevant to the r.v. stated in Example 3)

Figure 2. Left and right panel shows discrete and continuous distributions, respectively

Now, let us define a random variable.

Definition 1. ***Random Variable.*** *A random variable (r.v.) X is a real-valued function $X(\omega)$ over the sample space Ω of a random experiment, i.e., $X : \Omega \to \mathbb{R}$.*

Note that the randomness comes from the fact that outcomes i.e., the functional values are random. Always use upper case letters for random variables $(X, Y, \ldots)$. Further, always use lower

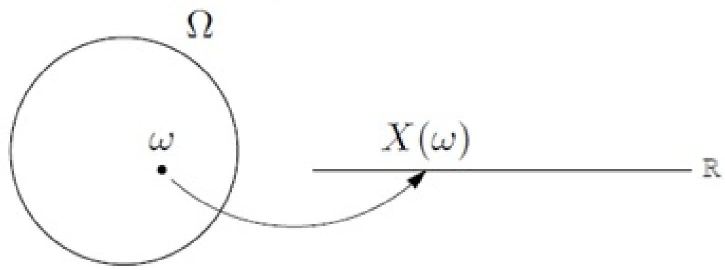
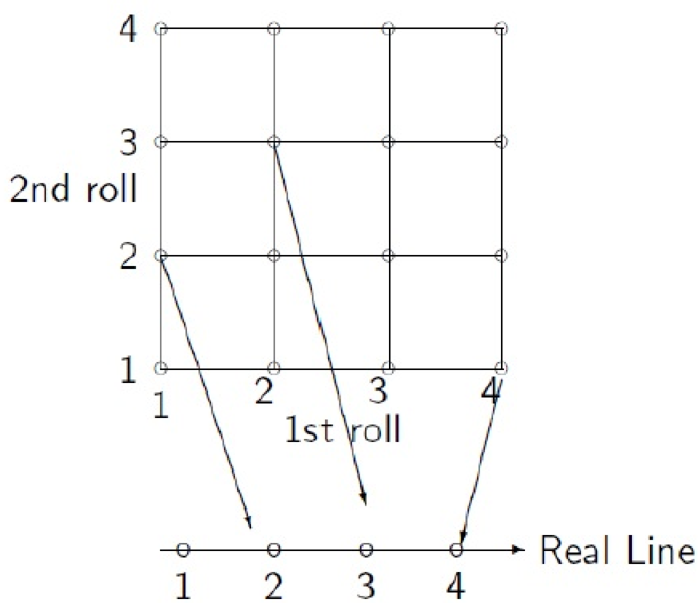


Figure 3. The diagram shows mapping in case of random variable

case letters for values of random variables: $X = x$ means that the random variable X takes the value x.

Next, let us consider some more examples to understand random variable and its variant.

1. Roll a 4-sided die twice.
   (a) Define the random variable X as the maximum of the two rolls,
   (b) Define the random variable Y to be the sum of the outcomes of the two rolls,



   (c) Define the random variable Z to be 0 if the sum of the two rolls is odd and 1 if it is even.

2. Flip coin until first heads shows up. Define the random variable $X = \{1, 2, \ldots\}$ to be the number of flips until the first heads.

3. Let $\Omega = \mathbb{R}$. Define the two random variables as
   $X = \omega$
   $Y = 1$ if $\omega \geq 0$, and otherwise, $Y = -1$.

4. n packets arrive at a node in a communication network. Here Ω is the set of arrival time sequences $(t_1, t_2, \ldots, t_n) \in (0, \infty)^n$.
   (a) Define the random variable N to be the number of packets arriving in interval $(0, 1]$,
   (b) Define the random variable T to be the first inter-arrival time.

Classification of random variables

Based on the nature of $X(\Omega)$, random variables are classified as following:

•*Discrete:* This random variable can assume only a set of finite or countably infinite values. Here, for example, the random variable X satisfies $X(\Omega) \subseteq \mathbb{N} \subset \mathbb{R}$, where $\mathbb{N}$ denotes the set of natural numbers. Examples 1, 2, 3(b), 4(a) are of discrete random variables.

•*Continuous:* This random variable can assume one of a continuum of values and the probability of each value is 0. Here, the random variable satisfies $X(\Omega) = \mathcal{B} \subseteq \mathbb{R}$, where $\mathcal{B}$ is a Borel set (*see Appendix*). Examples 3(a) and 4(b) are of continuous random variables.

•*Mixed:* This random variable is neither solely discrete nor continuous. It assumes one of a continuum of values as well as some set of a countable number of values, i.e. $X(\Omega)$ includes at least one set of real discrete point(s) as well as Borel set(s).

As mentioned earlier, when we define a random variable, say X, the basic aim of "computing $P(A)$" changes to "computing $P(X \in A)$". Irrespective of the nature of random variable X, our prime concern is that how to compute probability of an event involving the r.v. X, i.e. $P(X \in A)$ for any event $A \subseteq \mathbb{R}$.

3 Computation of probability involving random variable

In order to compute $P(X \in A)$ for any event $A \subseteq \mathbb{R}$, at first we have to specify the r.v. X. To do so, we consider the inverse image of the set A under $X(\omega), \{\omega : X(\omega) \in A\}$. So, $X \in A$ iff $\omega \in \{\omega : X(\omega) \in A\}$, thus $P(X \in A) = P(\omega : X(\omega) \in A)$.

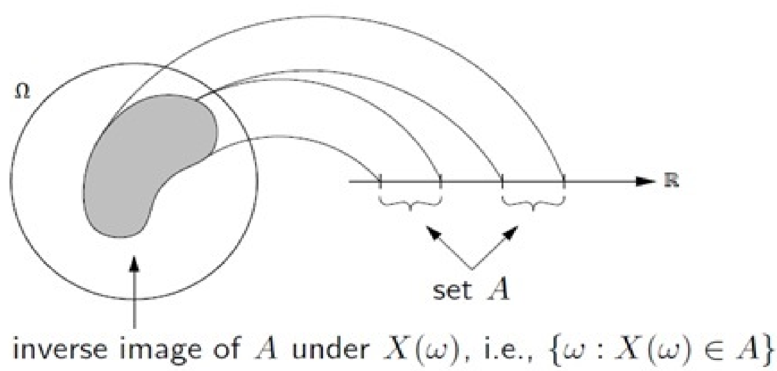


Figure 4. The diagram shows mapping and inverse image in case of random variable

The way in which the probability is computed will depend on the nature of the random variable X. In this context, we require a probability structure $\{f(x) : x \in \mathbb{R}\}$, where $f(x)$ gives the measure of chance that the r.v. X takes the value x. For that, one have to assume the function $f(x)$ on $x \in \mathbb{R}$, and then compute $P(X \in A)$.

If the random variable X is discrete and suppose $x_1, x_2, \ldots$ be the values of X that belong to A, then $P(X \in A) = P(X = x_1 \text{ or } x_2 \text{ or } x_3 \text{ or } \ldots) = \sum_{x \in A} f(x)$ which is same as $\sum_{x \in A} P(X = x)$. For discrete random variable X, 'measurement of chance' at x, denoted by $f(x)$ is exactly a probability value and hence, we commonly denote it as $P(X = x)$.

When X is continuous, $P(X \in A)$ may be computed from $\int_A f(x)dx$, if such a continuous (or, piecewise continuous) function f as well as the integral exist. Existence of such a function f associated with a given r.v. X is a matter of investigation. For continuous random variable X, 'measurement of chance' at x, earlier denoted by $f(x)$, need not be a probability and hence, we commonly denote it by $f_X(x)$.

If the r.v. is mixed type, and the set A consists of some discrete values $\{x_1, x_2, \ldots\}$ as well as some Borel set, then existence of corresponding 'measurement of chance' at each x, say $f(x)$, over the X is not possible.

Hence, we require a general 'probability structure' on random variable that can be used for all situations in order to compute $P(X \in A)$.

Before proceeding towards a general characterization of r.v., at first take a note that any subset A (of $\mathbb{R}$) can be built up based on one or more sets of the form $(-\infty, t], t \in \mathbb{R}$, using set theoretic operations. For examples,

$$(a, b] = (-\infty, b] - (-\infty, a], \qquad (a, b) = \bigcap_{n=1}^{\infty} (a, b + 1/n],$$

$$[b, d) = (a, d) - (a, b), \qquad [b, d] = (a, d] - (a, b).$$

Therefore, we can start with the problem of finding $P(X \in (-\infty, t])$. If we succeed, that will lead us to find $P(X \in A)$ for any given $A \subset \mathbb{R}$.

Distribution Function

Definition 2. ***Distribution Function.*** *For a random variable X, we can associate a function known as distribution function or cumulative distribution function (c.d.f) or simply, distribution function (d.f.), defined by*

$$F_X(t) = P(X \leq t), \qquad t \in \mathbb{R}. \tag{1}$$

Thus, the distribution function determines the probability that the r.v. X falls in an interval, say, $(a, b]$:

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a).$$

Example 4. Suppose a coin with probability p of landing heads is tossed until the first time a heads appears. Let T be the number of tosses required.

For a real number t, let [t] denote the integer part of t. We have $P(T > t) = P(T > [t]) = P(\text{first [t] tosses are tails}) = (1-p)^{[t]}$. Consequently,

$$F_T(t) = P(T \le t) \quad = \quad \begin{cases} 1-(1-p)^{[t]} & \text{if} \quad t \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

We can use this to compute $P(X \in A)$, where $A = (2,5]$.

$$P(2 < T \le 5) = F_T(5) - F_T(2) = 1-(1-p)^5 - 1 + (1-p)^2 = (1-p)^2 - (1-p)^5.$$

We may want to find the probability that a r.v. X falls in a closed interval, e.g., $[a,b]$. To this end, we need to find $P(X < x)$, where $P(X < x)$ can be obtained by

$$P(X < x) \quad = \quad \lim_{s \uparrow x} F_X(s). \tag{2}$$

The value of $\lim_{s\uparrow x} F_X(s)$ is called the left limit of F at x, and is sometime denoted by $F_X(x-)$. Part of the conclusion of Equation (2) is that a distribution function has left limits everywhere. Thus, we can also use the distribution function of X to calculate other probabilities involving X:

$$P(a \le X \le b) = P(X \le b) - P(X < a) = F_X(b) - F_X(a-),$$

$$P(X = a) = P(X \le a) - P(X < a) = F_X(a) - F_X(a-),$$

$$P(a < X < b) = P(X < b) - P(X \le a) = F_X(b-) - F_X(a).$$

Note: A random variable X is said to be proper if $P(-\inf < X < \infty) = 1$. Almost all random variables we will encounter in practice will be proper, but there exist random variables which are not proper. In the *Appendix*, a example of a random variable is given that is not proper.

At this point we fetch the question, already raised in the last but one paragraph before the section Distribution Function that how to compute $P(X \in A)$ for given $A \subseteq \mathbb{R}$. From the knowledge so far, we require the form of F_X, defined in Equation (1), in order to compute $P(X \in A)$. If F_X is assumed to be a function on $\mathbb{R}$, then some questions are naturally arise: (i) how the F_X should be shaped, (ii) what type of function it should be, (iii) what should be its range?

In the following theorem, we summarize the properties that any function F_X associated with a random variable X should possess in order to be treated itself as a distribution function.

Theorem 1. *Let X be a proper random variable with distribution function F_X. Then*

1. *F_X is monotonically non-decreasing, i.e. $F_X(s) \le F_X(t)$ for $-\infty < s < t < \infty$,*
2. *F_X is right-continuous, i.e., $F_X(t) = \lim_{u \downarrow 0} F_X(t+u)$,*
3. *$\lim_{t\to-\infty} F_X(t) = 0$,*
4. *$\lim_{t\to\infty} F_X(t) = 1$.*

Proposition 1. *Two random variables X and Y are identical if and only if their associated distribution functions $F_X(a)$ and $F_Y(a)$ are same for all $a \in \mathbb{R}$.*

Discrete Random Variable

Definition 3. *A random variable is said to be discrete if the set of possible values of the random variable is finite or countably infinite.*

Suppose, $\mathfrak{X} = (x_1, x_2, ..)$ be the set of possible values of a discrete random variable X. Here $X(\omega)$ partitions Ω into the sets $\{\omega : X(\omega) = x_i\}$ for $i = 1, 2, .$ Therefore, to specify a discrete r.v. X, it suffices to know the probability that the X takes the value x_i, for each i. As X has mass only on each of the x_i values, then the probability that the random variable X takes the value $x_i \in \mathfrak{X}$ is denoted by $P(X = x_i)$. So, it is clear that $P(X = x) = 0$ for any $x \in \mathfrak{X}^c \cap \mathbb{R}$. Now, we will specify what type of function $P(X = x)$ should be so that it can characterize the probability distribution of the said discrete random variable X.

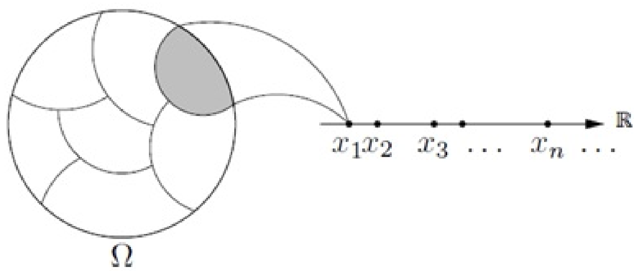


Figure 5. The diagram shows mapping in case of discrete random variable

Definition 4. ***Probability Mass Function.*** *Let X be a discrete random variable which takes values in the set $\mathfrak{X} = (x_1, x_2, \ldots)$. Then a function $P(X = x)$, associated to the random variable X, satisfying*
(i) $P(X = x) \geq 0$ for $x \in \mathbb{R}$,
(ii) $\sum_{x \in \mathfrak{X}} P(X = x) = 1$,
is called a probability mass function (or p.m.f).

A discrete random variable is thus completely specified by its probability mass function (p.m.f.).

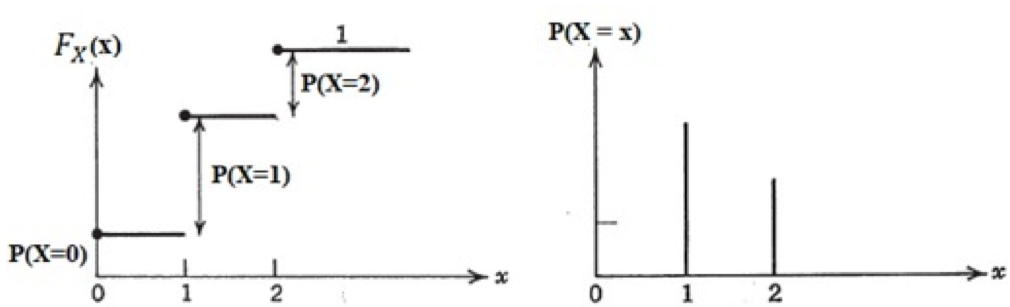


FIGURE 6. Distribution and Probability Functions of a Discrete Random Variable.

Now, let us see how does the distribution function F look like in case of a discrete random variable. As X has mass only on each of the x_i values, therefore, $F_X(x_i) - F_X(x_i-) = P(X \leq$

$x_i) - P(X < x_i) = P(X = x_i) > 0$, for all $i = 1, 2, \ldots$. Thus, we say that F_X has jump of magnitude $P(X = x)$ at x. Further, the F_X here is constant between the jumps. Therefore, the F_X is shown in the following figure.

So clearly, the pmf of X can be determined if the distribution function F_X is known. Thus, if one is asked to know the distribution of X, it is sufficient to provide either the pmf or the distribution function.

Remark 1. *This notation of probability function $P(X = x)$ is same as the density $f(x)$, as discussed earlier. As X has mass only on each of the x values, so the density function is renamed as probability function as we can compute the probability that the r.v. X takes the value x.*

Continuous Random Variables

Definition 5. *A random variable X is said to be continuous if the set of possible values of the random variable, say $\mathfrak{X} \subseteq \mathbb{R}$ is not finite and $P(X \in \mathcal{B}) > 0$ for any non-singleton subset $\mathcal{B} \subseteq \mathfrak{X}$*

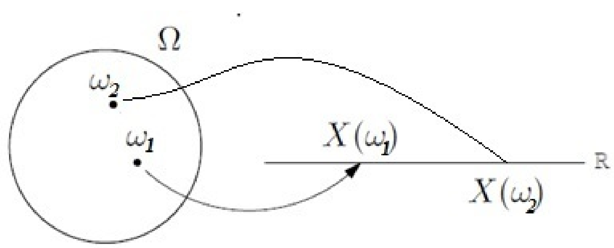


Figure 7. The diagram shows mapping in case of continuous random variable

Unlike discrete random variables, here $X(\omega)$ does not partition Ω into the sets $\{\omega : X(\omega) = x_i\}$ for $i = 1, 2, \ldots$. Therefore, for uncountably infinite possible outcomes of a random experiment, one can have uncountable and infinite possible values of the r.v. X which constitutes a Borel set. Let the set of possible outcomes be $\mathfrak{X}$, a Borel set, so $\mathfrak{X} \subseteq \mathbb{R}$ and $P(X \in \mathfrak{X}) = 1$. As continuous r.v. X does not have mass on any discrete point, then a probability of X, at any discrete point $x \in \mathfrak{X}$ produces

$$P(X = x) = P(X \leq x) - P(X < x) \quad = \quad 0. \tag{3}$$

Therefore, to specify the continuous r.v. X, it suffices to assume the measure of chance in terms of density function (which is similar to probability function but not exactly same). This density, denoted by $f_X(x)$, is defined on $x \in \mathfrak{X}$, such that $f_X(x) > 0$ for any $x \in \mathfrak{X}$ and $f_X(x) = 0$ for any $x \in \mathfrak{X}^c \cap \mathbb{R}$. This $f_X(x)$ is used to compute the probability of an event $X \in \mathcal{B}$, by

$$P(X \in \mathcal{B}) = \int_{\mathcal{B}} f_X(x) dx,$$

for any Borel set $\mathcal{B} \subseteq \mathbb{R}$. Now, we will specify what type of function $f_X(x)$ should be so that it can characterize the probability distribution of the said continuous random variable X.

Definition 6. ***Probability Density Function.*** *Let X be a continuous random variable which takes values in the set $\mathfrak{X} \subseteq \mathbb{R}$. Then a function $f_X(x)$, associated to the random variable X, satisfying*

(i) $f_X(x) \geq 0$ for $x \in \mathbb{R}$,
(ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$,

is called a probability density function (or p.d.f).

A continuous random variable is thus completely specified by its probability density function (p.d.f.).

Now, let us see how does the distribution function F look like in case of continuous random variable. As continuous random variable does not have mass on any single(discrete) real value, therefore, $F_X(x) - F_X(x-) = P(X \leq x) - P(X < x) = P(X = x) = 0$, for any $x \in \mathbb{R}$ (*see* Equation (3)). Thus, we say that F_X has no jump or point of discontinuity. So, in addition to the right continuity, F_X is also left-continuous, i.e., $F_X(t) = \lim_{u \downarrow 0} F_X(t - u)$. Therefore, the F_X, associated with a p.d.f. f_X is shown in the following figure.

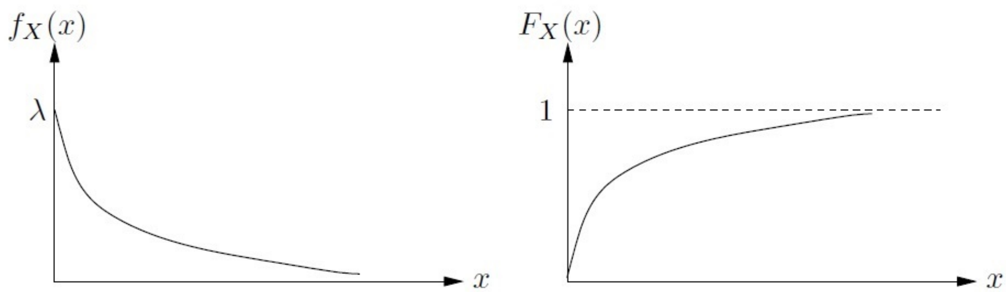


Figure 8: Probability density function and cumulative distribution function of a continuous random variable

If f_X is continuous, it may be obtained from F_X by differentiation; that is,

$$\frac{d}{dx}F_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(t)dt = f_X(x)$$

(the fundamental theorem of calculus). The general proof of this result is not shown here as it is measure-theoretic. So, if f_X is continuous, we have just seen that the F_X is differentiable, its derivative is f_X. So clearly, the p.d.f. of X can be determined if the distribution function F_X is known. Thus, if one is asked to know the distribution of X, it is sufficient to provide either the p.d.f. f_X or the distribution function F_X.

Appendix

▶*Borel Set:*
A Borel set is any set that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement.

▶*Axiomatic Definition of Probability:*
Since an arbitrary event A is nothing but a subset of Ω, so the domain of this probability function, say $P(\cdot)$, is the collection of all subsets of Ω. This collection is known as σ-field or σ-algebra or Borel-field.

Definition 7. *σ-field or Borel-field.* *A collection of all subsets of Ω is called σ-field or Borel-field, denoted by $\mathcal{B}$, if it satisfies the following properties:*
(i) $\phi \in \mathcal{B}$,
(ii) if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ [$\mathcal{B}$ is closed under complementation],
(iii) if $A_1, A_2, \ldots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ [$\mathcal{B}$ is closed under countable union].

Since $\mathcal{B}$ is a σ-field or σ-algebra or Borel-algebra, we see that it necessarily contains all open sets, all closed sets, all unions of open sets, all unions of closed sets, all intersections of closed sets, and all intersections of open sets.

Definition 8. ***Axiomatic Definition of Probability.*** *Given a sample space Ω and an associated σ-field or Borel-field $\mathcal{B}$, a set function $P(\cdot)$ with domain $\mathcal{B}$ is called probability if the following three axioms are satisfied:*
(i) $P(A) \geq 0$ for all $A \in \mathcal{B}$,
(ii) $P(\Omega) = 1$,
(iii) if $A_1, A_2, \ldots \in \mathcal{B}$ are mutually exclusive events, then $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

►*Example of improper random variable:*
Suppose a particle moves on the integer $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ as follows: at each move, it moves up one integer with probability 2/3, and moves down one integer with probability 1/3. The particle starts at 0. Let T be the first time that the particle is at 1. Hence, the event that the particle never hits 1 is $T = \infty$. In this case, one can check that $P(T < \infty) < 1$, which proves that T is not a proper random variable.

References

[1] Ash, R. B. (2008). *Basic Probability Theory*, Dover Publications, Inc., Mineola, N.Y. 11501.