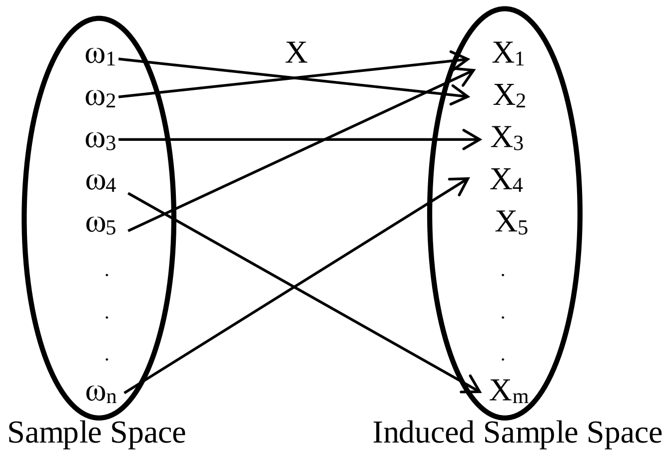
4. Probability Distributions

4.1 Random Variables

Sample spaces can contain quantitative outcomes or qualitative outcomes. It is more convenient to interact with sample spaces which contain numerical outcomes. A function that maps the original sample space in to the real numbers is called a random variable.

Definition

Let Ω be a sample space. Let X be a function from Ω to R (i.e. X: Ω → R). Then X is called a random variable.



There are two types of random variables.

- Discrete Random Variables – Can take distinct or separate values
  Induced sample space is finite or countable
- Continuous Random Variables – Can take any value in a range/ interval
  Induced sample space is not countable

Example 1: Consider the experiment of tossing a coin. Express the following events using a suitably defined random variable.

H = The event of getting a head

T =The event of getting a tail

Example 2: Consider the experiment of tossing a coin 10 times. Let E_i be the event of getting i number of heads (i=0, 1, 2, … 10). Write E_i in terms of a suitably defined random variable.

Example 3: Identify two possible random variables in the below given scenario.

Claim amounts of a certain insurance company are modeled using a normal distribution with an unknown mean and a known standard deviation σ = \$20. For a random sample of 20 claim amounts all that is known is that 5 of them are greater than \$200.

4.2 Probability Mass Function (pmf)

- The function that gives the probability of each possible value of a discrete random variable is called its probability mass function.
- For a discrete random variable X, the probability mass function is defined by,

$$f_X(x) = P(X = x)\,, \qquad\qquad x \in R$$

- A pmf can be an equation, a table, or a graph that shows how probability is assigned to possible values of the random variable.

Properties of a Probability Mass Function

Let X be a discrete random variable with probability mass function f_X. Then,

1. For any $x \in R, 0 \leq f_X(x) \leq 1$
2. Let E be an event and $I = \{X(\omega): \omega \in E\}$. Then, $P(E) = P(X \in I) = \sum_{x \in I} f_X(x)$
3. Let $R = \{X(\omega): \omega \in \Omega\}$. Then, $\sum_{x \in I} f_X(x) = 1$

Example 4: Write down the probability mass function for each of these experiments.

i. Tossing a fair coin
ii. The number of heads, Y, obtained when two fair coins are tossed.

Are these random variables discrete or continuous?

Example 5: Discrete random variable R has pmf, $P(R = r) = c(3 - r)$ for r = 0, 1, 2, 3.

i. Find the value of the constant c. C4+c3+c2+c1=c10
ii. Draw a vertical line graph to illustrate the distribution.
iii. Find $P(1 \leq R < 3)$.

Example 6: Let X be a random variable with the following probability distribution.

x	1	1.5	2	2.5	3	Other
$f_X(x)$	k	2k	4k	2k	k	0

i. Find the value of k
ii. Calculate$P(X \geq 1.75)$

Example 7: An appliance dealer sells three different models of freezers having 13.5, 15.9 and 19.1 cubic feet of storage space, respectively. The percentages of sales of freezers of above sizes were 20%, 50% and 30% respectively.

i. Let W be the amount of storage space of the freezer purchased by the next customer to buy a freezer. Find the probability mass function of W. is W discrete or continuous?
ii. What is the probability that the next customer will buy a freezer having a storage space less than 16 cubic feet?
iii. The profits from each of the above freezers are Rs.1000, 1200 and 1600 respectively. Find the probability mass functions of profits by defining a suitable random variable.
iv. If two freezers are sold per day, what is the probability that the total profit per day will be greater than Rs. 2500?

4.3 Probability Density Function

A continuous random variable takes on an uncountable infinite number of possible values. For a discrete random variable X that takes on a finite or countably infinite number of possible values, we determined $P(X = x)$ for all of the possible values of X, and called it the probability mass function ("pmf"). For continuous random variables, we'll need to find the probability that X falls in some interval (a, b), that is, we'll need to find $P(a < X < b)$ because as discussed later in the lesson finding $P(X = x)$ for a particular value of the continuous random variable will be meaningless.

A probability density function (PDF), or density of a continuous random variable, is a function, whose value at any given sample (or point) in the sample space can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

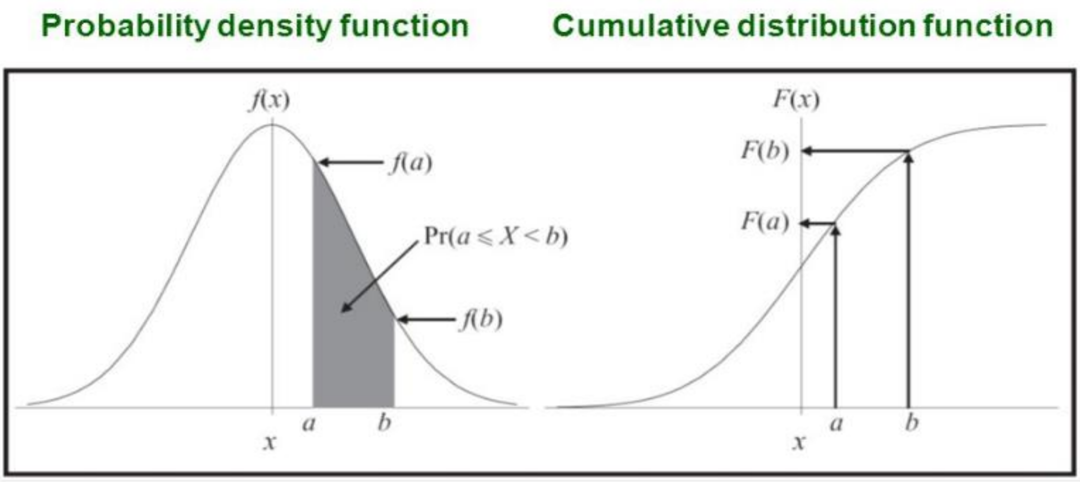
Properties of a pdf

Let X be a continuous random variable with probability density function f_x . Then,

1. For any $x \in R$, $0 \leq f_x(x)$
2. Let E be an event and $I = \{X(\omega): \omega \in E\}$. Then $P(E) = P(X \in I) = \int_I f_X(x)dx$
3. Let $R = \{X(\omega): \omega \in \Omega\}$. Then $\int_R f_X(x)dx = 1$

Calculation of Probability using pdf

Let a, b $\in$R such that $a \leq b$. Then,



$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Figure: $\boldsymbol{P(a \leq X \leq b) = \int_a^b f_X(x)dx}$

Note: If X is a continuous random variable with pdf f_X, then for any k $\in$R,

$$P(X = k) = P(k \leq X \leq k) = \int_k^k f_X(x)dx = 0$$

Therefore, for a continuous random variable X,

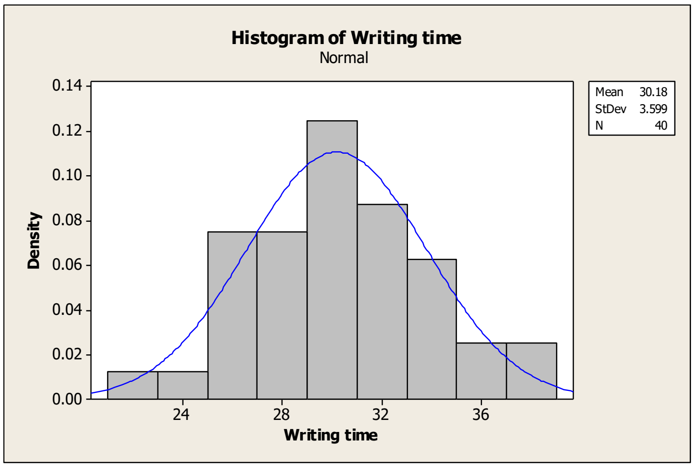
$$P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) = \int_a^b f_X(x)dx$$

There are large number of commonly used mathematical functions that are used as models for probability density functions. Uniform distribution, normal distribution, gamma distribution and exponential distribution are few examples. A suitable model for a given random variable can be chosen based on prior knowledge, or literature on similar situations. You shall learn about few commonly used pdf models in the next chapters. Two examples have been given below.

Example:

A consumer organization wishes to study about lifetime (hors in writing) of a certain brand of ballpoint pen. The organization purchased 40 such pens and inserted each in a specifically constructed machine that causes the pen to write continuously until the ink runs out. Below table list outs the observed lifetimes. Let X be the writing time of a randomly selected pen. Suggest a suitable model for the pdf of X.

31.7	26.2	35.0	34.8	33.1	30.8	34.1	34.4	31.4	29.5
29.0	29.7	32.5	25.3	30.6	37.7	28.7	23.1	25.7	26.7
31.0	28.7	30.7	36.4	27.3	32.1	29.2	30.1	26.3	25.9
27.5	28.6	37.2	31.3	22.7	27.4	31.7	30.0	33.6	29.6



Histogram of writing time data and a possible model for the pdf

Histogram is roughly symmetrical and bell shaped. Therefore, a bell shaped symmetric curve is suitable as the model for pdf of X. there are a number of mathematical functions that has the symmetric bell-shape. One such function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, -\infty < x < \infty$$

This is called the normal distribution. It contains two parameters μ and σ. In order to determine suitable values for these parameters, we use the well-known mathematical properties of this function. This function is symmetric around the value μ and $f(x) \approx 0$ when $x < \mu - 3\sigma$ and $x <$

$\mu + 3\sigma$. Matching these properties with the above histogram and curve, we chose $\mu = 30$ and $\sigma \approx$ (40-20)/6= 3.3. Hence we may take

$$f_X(x) = \frac{1}{3.3\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-30}{3.3}\right)^2\right\}, -\infty < x < \infty$$

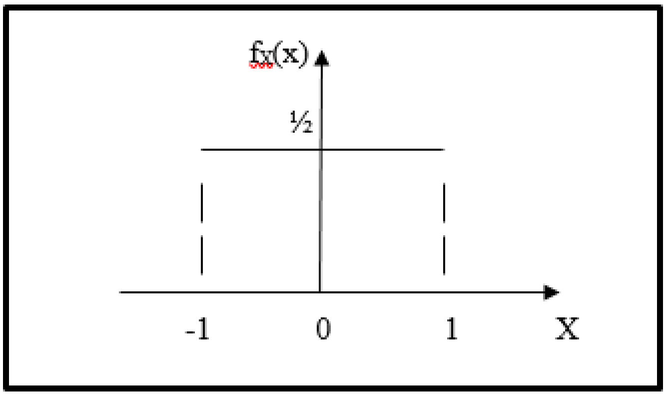
As the pdf of X. this curve is shown in the above figure and it seems a fairly good approximation.

Now, suppose we want to find the proportion of pens that can write more than 25 hours. In other words, we need $P(X > 25)$. Using or model, this probability can be calculated as

$$P(X > 25) = \int_{25}^{\infty} \frac{1}{3.3\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-30}{3.3}\right)^2\right\} dx, \qquad -\infty < x < \infty$$

This calculation is somewhat difficult. There is an easier method to calculate such probabilities of normal distributions. It will be discussed later.

Example 8: Probability mass function of a given random variable is given below.



i. Find the probability that X≤0.5.
ii. Find the probability that -0.7≤X≤0.3.