LECTURE – 17

Agenda:

1. Continuous random variables
2. Probability density functions
3. Probability distribution function

CONTINUOUS RANDOM VARIABLES

Till now we were dealing with discrete random variables, i.e., random quantities which can possibly take finite or countably infinite values. But there are many random quantities that we observe, which can possibly take uncountably many values. For example,

(i) Proportion of people infected by a pandemic. The set of possible values is $[0,1]$, i.e., any real number in the interval $[0,1]$.

(ii) Height of a randomly chosen individual in a country

(iii) The time it takes to get served in a French restaurant

and lots more of them.

For discrete random variables, we had the probability mass function where we assign $P(X=x)$ for every $x \in \mathcal{X} = \text{Range}(X)$, such that

$$\sum_{x \in \mathcal{X}} P(X=x) = 1.$$

HOWEVER, WE CANNOT DO THIS FOR CONTINUOUS RANDOM VARIABLES, IF WE WANT TO BE CONSISTENT WITH THE THREE AXIOMS OF PROBABILITY.

Mathematical fact: For a continuous random variable, if we insist on assigning a positive probability to each single outcome of the experiment i.e., if we insist that $P(X=x) > 0$ for every $x \in \mathcal{X} = \text{Range}(X)$, then $P(S) > 1$, which is in deviation from the second axiom of probability (and also intuition)

WHAT IS A WAY OUT?

Instead of working with the probability mass function, we work with what is known as a probability density function which is supposed to indicate the relative proportion of each value in the range of the random variable under consideration.

PROBABILITY DENSITY FUNCTION FOR A CONTINUOUS RANDOM VARIABLE

<u>Definition</u>: A continuous random variable X is said to have a probability density function $f_X : \mathbb{R} \to \mathbb{R}$ if,

(i) $f_X(x) \geq 0$ for every $x \in \mathbb{R}$
($f_X(x) = 0$ for $x \notin \mathcal{X}$)

(ii) $\int_{-\infty}^{\infty} f_X(x) = 1$, or equivalently $\int_{\mathcal{X}} f_X(x)\,dx = 1$.

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

for all $a < b \in \mathbb{R}$

It turns out that this definition is consistent with the axioms of probability

Recall that for any random variable X, the probability distribution function $F_X : \mathbb{R} \to \mathbb{R}$ is defined as

$$F_X(b) = P(X \leq b) \text{ for every } b \in \mathbb{R}$$

Then, by the definition of the probability density function,

$$\boxed{F_X(b) = P(-\infty \leq X \leq b) = \int_{-\infty}^{b} f_X(x)\,dx}$$

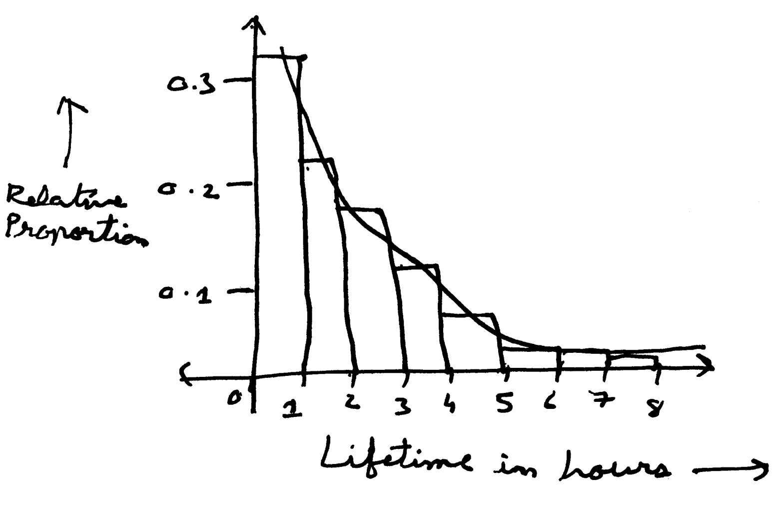
By the fundamental theorem of calculus,

$$\frac{d}{db} F_X(b) = \frac{d}{db}\left(\int_{-\infty}^{b} f_X(x)\,dx\right) = f_X(b).$$

Hence, THE DENSITY FUNCTION f_X IS THE DERIVATIVE OF THE DISTRIBUTION FUNCTION OF A RANDOM VARIABLE

Let us look at a practical example to see how people decide what is an appropriate density function for a random quantity which takes a continuous set of values.

<u>Example:</u> Suppose that we are interested in the battery life of a transistor randomly chosen from a large collection of transistors. Suppose that in the past, somebody chose 50 of these transistors randomly and measured their lifetimes (see Table 5.1 in the textbook). They found that 32% of the 50 observations fall into $[0,1]$, 22% fall into $[1,2]$, and so on.



This histogram gives us a good insight into the possible probabilistic model for the random variable X, where

X = Lifetime of a randomly chosen transistor.

The histogram ~~[illegible]~~ seems to be very well approximated by a negative exponential curve, in particular the function

$$f_X(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Note that $f_X(x) \geq 0$ and $\int_0^\infty f_X(x) = 1$.

Hence, after looking at the historical data, it seems reasonable to assume that X is a random variable with probability density function f_X.

This is a standard way of coming up with probabilistic models for continuous random quantities. Divide the range into small subintervals. From available historical data, find the proportion of observations in each subinterval. Find a function f_X which approximates the resulting histogram, and satisfies $f_X(x) \geq 0$ for every $x \in \mathbb{R}$, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.