Probability and Statistics (EPRST)

Lecture 3

A very convenient concept to describe and study randomness is the concept of **a random variable**.

Definition

Random variable *is a real-valued function, defined on a sample space (on the set of elementary events of some random experiment).*

Interpretation: a random variable is a *function* whose value depends on the result of a random experiment - if the result of the experiment is an elementary event $\omega \in \Omega$, then $X(\omega) \in \mathbb{R}$ is the value of X.

Random variables - examples

Example

We flip a coin. On sample space $\Omega = \{H, T\}$ *we define the following random variable* X*, that is a function* $X : \Omega \to \mathbb{R}$*:*

$$X(\omega) = \begin{cases} -1, & \omega = H, \\ 2, & \omega = T. \end{cases}$$

Example

Another example of a random variable, this time related to the random experiment of rolling a die:

X *- the result of the roll.*

What is the formal definition of X*?*

Random variables - examples

Random variables can take infinitely many values:

Example

We keep tossing a coin until we get heads. A random variable related to this experiment:

$$X \text{ - number of tosses.}$$

Example

Alice and Bob agree to meet between 11:00 and 12:00 at a specified place. A random variable related to this situation:

$$X \text{ - the waiting time of a person who arrives first for the person who arrives second.}$$

Random variables and related random events

We will often consider some random events related to random variables. For example, if X is a random variable, then we might be interested in the probability of the random event that

- X takes a specified value, or
- X takes a value from a given subset of $\mathbb{R}$.

Example

Say X means the outcome of a roll of a fair die. Give the following probabilities:

- $\mathbb{P}(X = 2)$,
- $\mathbb{P}(X > 3)$,
- $\mathbb{P}(X \in (3, \infty))$,
- $\mathbb{P}(X \in (3, 6))$,
- $\mathbb{P}(X < 0)$.

Recall that

random events are subset of sample space Ω.

So

- if $a \in \mathbb{R}$, then the event $X = a$ means in fact subset

$$\{\omega \in \Omega : X(\omega) = a\}$$

  of Ω,
- if $A \subset \mathbb{R}$, then the event „$X \in A$" (X *takes values from* A) is subset

$$\{\omega \in \Omega : X(\omega) \in A\}$$

  of Ω.

Distribution of a random variable

Definition

Distribution of a random variable *X is a function that assigns a number $\mathbb{P}(X \in A)$ to any subset $A \subset \mathbb{R}$.*

Take-home message: the distribution of a random variable X is the set of all possible values of X and any way of specifying how probability is distributed over those values.

Example

Toss of a symmetric coin. Find the distribution of random variable

$$X(\omega) = \begin{cases} -1, & \omega = H, \\ 2, & \omega = T. \end{cases}$$

Two types of probability distributions

We will be dealing with two basic types of probability distributions:

- discrete distributions,
- continuous distributions.

Discrete probability distributions

Definition

*A random variable X has a **discrete distribution**, if there exists a finite or infinite but countable set $S \subset \mathbb{R}$ (called **the support of the distribution of** X) such that*

1. $\mathbb{P}(X = x_n) > 0 \ \forall \ x_n \in S$;
2. $\sum_{x_n \in S} \mathbb{P}(X = x_n) = 1$.

Some examples of discrete random variables:

- toss of a coin, $X = -1$, when H, $X = 2$, when T;
- roll of a die, X - the outcome;
- keep tossing a coin until you get an H, X - the number of flips.

Discrete probability distributions - cont'd

Example

We roll a fair die. Let X be the outcome. What is the distribution of X?

Example

We keep tossing a coin until we get heads. What is the distribution of the number of tosses?

Discrete probability distributions - cont'd

In order to define a discrete probability distributions one needs to provide the support (=set of possible values) $S = \{x_1, x_2, \ldots\}$ and the probabilities

$$\mathbb{P}(X = x_1), \mathbb{P}(X = x_2), \ldots .$$

This implies that for a discretely distributed random variable X and for any subset $A \subset \mathbb{R}$

$$\mathbb{P}(X \in A) = \sum_{i : x_i \in A} \mathbb{P}(X = x_i).$$

Continuous probability distributions

Definition

If there exists a function $f : \mathbb{R} \to [0, \infty)$ such that for any $-\infty \leq a < b \leq +\infty$

$$\mathbb{P}(X \in (a, b)) = \int_a^b f(x)dx,$$

then we say that random variable X has ***a continuous distribution with the density*** *f.*

A function $f : \mathbb{R} \to \mathbb{R}$ is a probability density function iff $f \geq 0$ and $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$.

Example of a continuously distributed random variable

Example

We pick a random point from $[0, 1]$ (=geometric probability). Let X denote the chosen point. What is the density of the distribution of X?

Convenient notation - **an indicator function of a set**: if A is a set then

$$\mathbb{1}_A(x) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Examples of probability density functions

- $f(x) = \mathbb{1}_{(5,6)}(x)$, $f(x) = \mathbb{1}_{(-10,-9)}(x)$;
- $f(x) = (1/100) \cdot \mathbb{1}_{[0,100]}(x)$, $f(x) = (1/100) \cdot \mathbb{1}_{[150,250]}(x)$;
- $f(x) = (1/x^2) \cdot \mathbb{1}_{(1,\infty)}(x)$;
- $f(x) = (1/2)\exp(-|x|)$.

Example

Suppose X is a random variable with a continuous distribution given by the density

$$f(x) = (1/x^2) \cdot \mathbb{1}_{(1,\infty)}(x).$$

Compute: $\mathbb{P}(X \in (3,5))$, $\mathbb{P}(X > 2)$, $\mathbb{P}(X \leq 7)$, $\mathbb{P}(X \in (-1,\infty))$, $\mathbb{P}(X < 1)$, $\mathbb{P}(X \leq 1)$, $\mathbb{P}(X = 3)$.

Discrete vs. continuous probability distributions

If X has a discrete distribution:

- X takes values from a finite or infinitely countable set $S = \{x_1, x_2, \ldots\}$, $\sum_i \mathbb{P}(X = x_i) = 1$;
- distribution of X is determined by S along with $\mathbb{P}(X = x_1)$, $\mathbb{P}(X = x_2)$,...;
- for any $A \subset \mathbb{R}$,

$$\mathbb{P}(X \in A) = \sum_{i: x_i \in A} \mathbb{P}(X = x_i).$$

If X has a continuous distribution with the density f:

- X takes values from an uncountably infinite set S, $\int_S f(x)\mathrm{d}x = 1$;
- for any $x \in \mathbb{R}$, $\mathbb{P}(X = x) = 0$;
- the distribution of X is determined by the density f;
- for any $A \subset \mathbb{R}$,

$$\mathbb{P}(X \in A) = \int_A f(x)\mathrm{d}x.$$

Cumulative distribution function

A universal (applied both to discrete and continuous distributions) tool to describe a probability distribution of a random variable is **(cumulative) distribution function**.

Definition

(Cumulative) distribution function *of the probability distribution of a random variable X is a function $F_X : \mathbb{R} \to \mathbb{R}$ defined for all $t \in \mathbb{R}$ by*

$$F_X(t) = \mathbb{P}(X \leq t).$$

Example

We flip a fair coin. Find the CDF of

$$X(\omega) = \begin{cases} -1, & \omega = H, \\ 2, & \omega = T. \end{cases}$$

CDFs for discrete random variables

If X has a discrete distribution then for any $A \subset \mathbb{R}$

$$\mathbb{P}(X \in A) = \sum_{i: x_i \in A} \mathbb{P}(X = x_i).$$

This means that for any discretely distributed rv X, its CDF is **a staircase function**

$$F_X(t) = \mathbb{P}(X \leq t) = \sum_{i: x_i \leq t} \mathbb{P}(X = x_i).$$

In particular, F_X is discontinuous at points $x_1, x_2, \ldots$. The height of jump of F_X at x_i equals $\mathbb{P}(X = x_i)$.

CDFs for continuous random variables

If f is the density of X, then

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}\left(X \in (-\infty, t]\right) = \int_{-\infty}^{t} f(u)\mathrm{d}u \quad \forall\ t \in \mathbb{R}.$$

This implies (by the Fundamental Theorem of Calculus) that

- F_X is a continuous function,
- $F'_X(x) = f(x)$ as long as the derivative exists.

Example

Let X be a randomly picked number from $[0, 1]$. Find the CDF of the distribution of X.

The distribution of any random variable is uniquely determined by its cumulative distribution function, that is for random variables X and Y

the distribution of X is the same as the distribution of Y

iff

$$F_X(t) = F_Y(t) \text{ for all } t \in \mathbb{R}.$$

This means that if one needs to find the distribution of a random variable, then one can equivalently find the cumulative distribution function.