STAT 511 fa 2019 Lec 04 slides

Random variables and the cdf

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

A random variable is a numeric encoding of the outcome of an experiment.

Random variable

A *random variable* is a function from a sample space $\mathcal{S}$ to the real numbers.

That is, a *random variable* X is a function $X \colon \mathcal{S} \to \mathbb{R}$.

Denote by $\mathcal{X}$ the range of X, the set of values X may take.

We often call $\mathcal{X}$ the *support* of X.

Examples:

1. Flip a coin and let $X = 1$ if heads, $X = 0$ otherwise.
2. Flip a coin three times and let $X =$ the number of heads.
3. Let $X =$ time until you drop your new phone.
4. Let $X =$ number on up-face of rolled die.

Probabilities about a random variable X with finite support

Let X be an rv on a finite $\mathcal{S} = \{s_1, \dots, s_n\}$ with support $\mathcal{X} = \{x_1, \dots, x_m\}$. Then for any $x \in \mathcal{X}$, we write

$$P_X(X = x) = P(\{s \in \mathcal{S} : X(s) = x\}).$$

So $P_X(X = x)$ is the probability that X takes the value x.

Same for countable $\mathcal{S} = \{s_1, s_2, \dots\}$ and countable $\mathcal{X} = \{x_1, x_2, \dots\}$.

Exercise: Tabulate $P_X(X = x)$ for all $x \in \mathcal{X}$ when $X = \#$ heads in 3 coin flips.

Probabilities about a random variable X

Let X be an rv on *any* sample space $\mathcal{S}$ with support $\mathcal{X}$. Then for any $A \in \mathcal{E}_{\mathcal{X}}$, we write

$$P_X(X \in A) = P(\{s \in \mathcal{S} : X(s) \in A\}).$$

In the above, $\mathcal{E}_{\mathcal{X}}$ is a collection of sets of interest in $\mathcal{X}$.

So $P_X(X \in A)$ is the probability that X takes a value in A.

We will call P_X the *probability distribution* of X.

Example: If $X =$ time until you drop your new phone and $A = [1, 2)$, then

$$P_X(X \in [1, 2)) = P(\{s \in [0, \infty) : X(s) \in [1, 2)\}) = P([1, 2)),$$

since $X(s) = s$.

Cumulative distribution function

The *cumulative distribution function (cdf)* F_X of an rv X is the function given by

$$F_X(x) = P_X(X \leq x) \text{ for all } x \in \mathbb{R}.$$

Exercise: Let $X = \#$ heads in 3 coin flips.

1. Give the cdf $F_X(x)$ for all $x \in \mathbb{R}$.
2. Draw a detailed picture of F_X.
3. Discuss interpretation of $F_X(1/2)$, say.
4. Discuss interpretation of jump sizes.
5. Discuss $F_X(x)$ for $x < 0$ and for $X \geq 3$.

Theorem (Properties of a cdf)

The function F_x is a cdf if and only if

-
-
-

Exercise: Let $X = \#$ free throw attempts required for you make one. Assume the attempts are independent with success probability $p \in (0,1)$.

1. Give the support $\mathcal{X}$ of X.
2. Begin tabulating the values $P_X(X = x)$ for each $x \in \mathcal{X}$.
3. Find an expression which gives $P_X(X = x)$ for any $x \in \mathcal{X}$.
4. Find an expression for $F_X(x) = P_X(X \leq x)$ for any $x \in \mathcal{X}$.
5. Draw a picture of F_X when $p = 1/2$.
6. Verify that F_X has the three properties of a cdf.

Exercise: Let X = time (months) until you drop your new phone and suppose X has the cdf given by

$$F_X(x) = \begin{cases} 1 - e^{-x/10}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

1. Verify that F_X has the three properties of a cdf.
2. Use F_X to obtain $P(X \leq 1)$
3. Use F_X to obtain $P(2 < X)$
4. Use F_X to obtain $P(1 < X \leq 3)$

Continuous and discrete random variables

A random variable X with cdf F_X is called a

1. *continuous random variable* if $F_X(x)$ is a continuous function of x
2. *discrete random variable* if $F_X(x)$ is a step function of x.

For a continuous rv X, for any $a, b \in \mathbb{R}$, $a < b$, we have

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

Explain above result.

Identically distributed-ness of two random variables

Two random variables X and Y on the same sample space with the same support $\mathcal{X}$ are called *identically distributed* if for every $A \in \mathcal{E}_{\mathcal{X}}$

$$P_X(X \in A) = P_Y(Y \in A).$$

In the above, $\mathcal{E}_{\mathcal{X}}$ is a collection of sets of interest in $\mathcal{X}$.

Theorem (Identically distributed-ness result)

The following two statements are equivalent:

- *The random variables X and Y are identically distributed.*
- $F_X(x) = F_Y(x)$ *for every* $x \in \mathbb{R}$.

So, two random variables are identically distributed if they have the same cdf.

Example: Let

$$\begin{aligned} X &= \# \text{ coin flips to get a head on Monday} \\ Y &= \# \text{ coin flips to get a head on Tuesday} \end{aligned}$$

Then $F_X(x) = F_Y(x)$ for every x.