

## Lecture 9: Sept 12

### Last time

- Presentations

### Today

- HW2 posted (due: Sept 15th)
- No lecture, but reviews on Fridays with two presentations
- Random variables
- Distribution Functions
- Types of Random Variables

Some properties of the cdf:

Let  $F(x)$  be a cdf. Then

1.  $0 \leq F(x) \leq 1$
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$
3.  $\lim_{x \rightarrow \infty} F(x) = 1$
4.  $F$  is nondecreasing: if  $a < b$ , then  $F(a) \leq F(b)$
5.  $F$  is right-continuous:  $\lim_{x \downarrow b} F(x) = F(b)$ , or  $\lim_{x \rightarrow b^+} F(x) = F(b)$
6.  $\Pr(a < X \leq B) = F(b) - F(a)$

**Theorem** The function  $F(x)$  is a cdf if and only if the following three conditions hold:

1.  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$
2.  $F$  is nondecreasing: if  $a < b$ , then  $F(a) \leq F(b)$
3.  $F$  is right-continuous:  $\lim_{x \downarrow b} F(x) = F(b)$ , or  $\lim_{x \rightarrow b^+} F(x) = F(b)$

The cdf does not contain information about the original sample space.

**Definition** Two random variables  $X$  and  $Y$  are identically distributed if, for every Borel set  $A \subset \mathbb{R}$ ,  $\Pr(X \in A) = \Pr(Y \in A)$ .

**Example** Toss a fair coin  $n$  times. The number of heads and the number of tails have the same distribution.

**Theorem** The following two statements are equivalent:

1. The random variables  $X$  and  $Y$  are *identically distributed*.
2.  $F_X(x) = F_Y(x)$  for every  $x$ .

## Types of Random Variables

**Definition** A random variable  $X$  can be

- *discrete*:
  - $X$  takes on a finite or countably infinite number of values
  - $F_X(x)$  is step-wise constant
- *continuous*:
  - the range of  $X$  consists of subsets of the real line
  - $F_X(x)$  is continuous.
- *mixed*:  $F_X(x)$  is piecewise continuous.

**Example** A random variable has cdf

$$F(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x < 1 \\ 2/3 & 1 \leq x < 2 \\ 11/12 & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

Is this a valid cdf? Is it a discrete random variable or continuous random variable or mixed?  
*solution:*

## Discrete Random Variables

Suppose a random variable  $X$  takes only a finite or countable number of values. Let the sample space of  $X$  be  $S = \{x_1, x_2, \dots\}$ . Then the cdf can be expressed as:

$$F(x) = \sum_{x_i \leq x} \Pr(X = x_i).$$

**Definition** The *probability mass function* (pmf) of a discrete random variable  $X$  is given by

$$f_X(x) = \Pr(X = x) \text{ for all } x.$$

If the sample space of  $X$  is  $X = \{x_1, x_2, \dots\}$ , then

$$f(x_i) = \Pr(X = x_i) = \Pr(x_{i-1} < X \leq x_i) = F(x_i) - F(x_{i-1}).$$

**Example** (Geometric probabilities) Suppose we do an experiment that consists of tossing a coin until a head appears. Let  $p$  = probability of a head on any given toss, and define a random variable  $X$  = number of tosses required to get a head. Then for any  $x = 1, 2, \dots$ ,

$$\Pr(X = x) = (1 - p)^{x-1}p,$$

since we must get  $x - 1$  tails followed by a head for the event to occur and all trials are independent. What is the pmf of the above Geometric distribution? What is the cdf?

*solution:*

**Definition** The *domain* of a random variable  $X$  is the set of all values of  $x$  for which  $f(x) > 0$ . This is also called *range* or *sample space*.

Properties of the pmf:

1.  $f(x) > 0$  for at most a countable number of values  $x$ . For all other values  $x$ ,  $f(x) = 0$ .
2. Let  $\{x_1, x_2, \dots\}$  denote the domain of  $X$ . Then

$$\sum_{i=1}^{\infty} f(x_i) = 1.$$

An obvious consequence is that  $f(x) \leq 1$  over the domain.

**Example** What is the pmf of a deterministic random variable (a constant)?

*solution:*

**Example** In many applications, a formula can be used to represent the pmf of a random variable. Suppose  $X$  can take values  $1, 2, \dots$  with pmf

$$f(x) = \begin{cases} \frac{1}{x(x+1)} & \text{for } x = 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases}$$

How would we determine if this is an allowable pmf?

*solution:*

## Continuous Random Variables

**Definition** A random variable  $X$  is *continuous* if  $F_X(x)$  is a continuous function of  $x$ .

**Definition** A random variable  $X$  is *absolutely continuous* if  $F_X(x)$  is an absolutely continuous function of  $x$ .

**Definition** A function  $F(x)$  is *absolutely continuous* if it can be written

$$F(x) = \int_{-\infty}^x f(x)dx.$$

Absolute continuity is stronger than continuity but weaker than differentiability. An example of an absolutely continuous function is one that is:

- continuous everywhere
- differentiable everywhere, except possibly for a countable number of points.

**Definition** The *probability density function* or pdf,  $f_X(x)$ , of a continuous random variable  $X$  is the function that satisfies

$$F_X(x) = \int_{-\infty}^x f_X(t)dt \quad \text{for all } x.$$

**Notation:** We write  $X \sim F_X(x)$  for the expression “ $X$  has a distribution given by  $F_X(x)$ ” where we read the symbol “ $\sim$ ” as “is distributed as”. Similarly, we can write  $X \sim f_X(x)$  or, if  $X$  and  $Y$  have the same distribution,  $X \sim Y$ .

**Theorem** A function  $f_X(x)$  is a pdf (or pmf) of a random variable  $X$  if and only if

1.  $f_X(x) \geq 0$  for all  $x$ .
2.  $\int_{-\infty}^{\infty} f_X(x)dx = 1$  (pdf)    or     $\sum_x f_X(x) = 1$  (pmf).

**Example** Suppose  $F(x) = 1 - e^{-\lambda x}$  for  $x > 0$  and  $F(x) = 0$  otherwise. Is  $F(x)$  a cdf? What is the associated pdf?

*solution:*

Notes

- If  $X$  is a continuous random variable, then  $f(x)$  is not the probability that  $X = x$ . In fact, if  $X$  is an absolutely continuous random variable with density function  $f(x)$ , then  $\Pr(X = x) = 0$ . (Why?)

*proof*

- Because  $\Pr(X = a) = 0$ , all the following are equivalent:

$$\Pr(a \leq X \leq b), \quad \Pr(a \leq X < b) \quad , \quad \Pr(a < X \leq b) \quad \text{and} \quad \Pr(a < X < b)$$

- $f(x)$  can exceed one!