Lecture 9: Sept 11

Last time

- Random variables

Today

- Distribution Functions
- Types of Random Variables

Distribution Functions

Distribution Functions are used to describe the behavior of a r.v.

Cumulative distribution function

Definition The *cumulative distribution function* or *cdf* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = \Pr_X(X \leqslant x), \text{ for all } x.$$

Definition The *survival function* of a random variable X, is defined by

$$S_X(x) = 1 - F_X(x) = \Pr_X(X > x).$$

Example Consider the experiment of tossing three fair coins, and let X = number of heads observed. The cdf of X is

$$F_X(x) = \begin{cases} 0 & \text{if } -\infty < x < 0 \\ \frac{1}{8} & \text{if } 0 \leqslant x < 1 \\ \frac{1}{2} & \text{if } 1 \leqslant x < 2 \\ \frac{7}{8} & \text{if } 2 \leqslant x < 3 \\ 1 & \text{if } 3 \leqslant x < \infty \end{cases}$$

Some properties of the cdf:
Let $F(x)$ be a cdf. Then

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

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

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

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

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

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 $S = \{x_1, x_2, \ldots\}$. Then the cdf can be expressed as:

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

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 $X = \{x_1, x_2, \ldots\}$, then

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

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 $x = 1, 2, \ldots,$

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

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*, *sample space* or *support*.

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 $\{x_1, x_2, \ldots\}$ denote the domain of X. Then

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

An obvious consequence is that $f(x) \leqslant 1$ 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 $1, 2, \ldots$ with pmf

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

How would we determine if this is an allowable pmf?
solution: