

## Lecture 6: Sept 2

### Last time

- Random variables

### Today

- Random variables
- Distribution Functions

**Induced probability function** Suppose we have a sample space  $S = \{s_1, s_2, \dots, s_n\}$  with a probability function  $\Pr$  defined on the original sample space. We define a random variable  $X$  with range  $\mathcal{X} = \{x_1, \dots, x_m\}$ . We can define a probability function  $\Pr_X$  on  $\mathcal{X}$  in the following way. We will observe  $X = x_i$  if and only if the outcome of the random experiment is an  $s_j \in S$  such that  $X(s_j) = x_i$ . Therefore,

$$\Pr_X(X = x_i) = \Pr(\{s_j \in S : X(s_j) = x_i\}),$$

defines an *induced* probability function on  $\mathcal{X}$ , defined in terms of the original function  $\Pr$ .

We will write  $\Pr(X = x_i)$  rather than  $\Pr_X(X = x_i)$  for simplicity. Note on notation: random variables will always be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letters.

**Example** Consider the experiment of tossing a fair coin three times. Define the random variable  $X$  to be the number of heads obtained in the three tosses. A complete enumeration of the value of  $X$  for each point in the sample space is

$s$	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(s)$	4	2	2	2	1	1	1	0

What is the range of  $X$ ? What is the induced probability function  $\Pr_X$ ?

*solution:*

So far, we have seen finite  $S$  and finite  $\mathcal{X}$ , and the definition of  $\Pr_X$  is straightforward. If  $\mathcal{X}$  is uncountable, we define the induced probability function,  $\Pr_X$  for anyset  $A \subset \mathcal{X}$ ,

$$\Pr_X(X \in A) = \Pr(\{s \in S : X(s) \in A\}).$$

This defines a legitimate probability function for which the Kolmogorov Axioms can be verified.

## 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

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

**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

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

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  $x$ .