

## LECTURE NOTE 2 \*

# RANDOM VARIABLES, PROBABILITY MASS/DENSITY FUNCTION, AND CUMULATIVE DISTRIBUTION FUNCTION (UNIVARIATE MODEL)

MIT 14.30 SPRING 2006

HERMAN BENNETT

## 3 Random Variables

### 3.1 Intuitive Definition

A **random variable** is a variable with unknown numerical value that can *take on*, or represent, any possible element from a sample space. The elements of a sample space have probabilities associated (probability function). Thus, it is said that a **random variable** *takes on* various possible values with certain probability.

There are two kinds of random variables (RV): i) *discrete* RV, where  $X$  can take only a finite (or countably infinite) number of values, and ii) *continuous* RV, where  $X$  can take any value in the real line within a bounded or unbounded interval.

### 3.2 Mathematical Definition

A random variable  $X(s)$ , where  $s \in S$ , is a function from a sample space  $S$  (domain) into the real numbers.

The specification of a random variable can also imply a redefinition of the sample space. *E.g.*: In the experiment of tossing a coin 3 times, we could define as a random variable “the total number of heads.”

---

\*Caution: These notes are not necessarily self-explanatory notes. They are to be used as a complement to (and not as a substitute for) the lectures.

**Example 3.1.** Experiment: Fair coin tossed 3 times,

$s$	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(s)$ (“outcome”)	1	2	3	4	5	6	7	8
$X(s)$ (“number of H”)	3	2	2	2	1	1	1	0

Notation: RV will always be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letter. Thus, the random variable  $X$  can take the value  $x$ .

- Diagrams:

The probability function defined on the sample space determines the (probability) distribution of the random variable. The distribution of a discrete (continuous) RV can be fully characterized by its *cumulative distribution function* or its *probability mass (density) function* (except in pathological cases).

## 4 Discrete Model

### 4.1 Probability Mass Function

The probability mass function (pmf) of a discrete RV  $X$ , denoted  $f_X(x)$ , is given by:

$$f_X(x) = P(X = x) , \quad \text{for all } x; \quad (4)$$

and satisfies the following properties:

- i)*  $f(x) \geq 0$  , for all  $x$ .
- ii)*  $\sum_{i=1}^{\infty} f(x_i) = 1$  .

Notation: If it is necessary to stress the fact that  $f(x)$  is a pmf of the random variable  $X$ , the notation  $f_X(x)$  will be used. If it is clear from the context, the subindex will be avoided:  $f(x)$ . The same is true for the functions we will see subsequently.

**Example 4.1.** Following Example 3.1, list all possible values for the random variable “number of H”, write its pmf, check its properties, and graph it.

**Example 4.2.** Uniform Distribution:  $f(x) = 1/k$  for  $x = 1, 2, \dots, k$ , and 0 otherwise. Graph  $f(x)$ . Assume  $k = 10$ , find  $P(X = 3)$  and  $P(2 \leq X)$ .

**Example 4.3.** Binomial Distribution. A sequence of  $n$  independent and identical trials is performed, where each trial can have only two possible outcomes: “success” or “failure.” This is called a binomial experiment. The RV that describes the number of “successes” in the sequence has a binomial distribution, characterized as follows:

$$f(x) = P(X = x) = p^x(1 - p)^{n-x} \binom{n}{x} \quad \text{for } x = 0, 1, \dots, n \quad (5)$$
$$X \sim \text{bin}(n, p)$$

Where  $n$  is the total number of trials and  $p$  the probability of “success” in each trial. If  $n = 1$ , the distribution is called Bernoulli Distribution. Application: A die is rolled 4 times. Define the RV  $X$  as the ‘total number of 6’s in 4 rolls’. Compute the probability of generating two (and only two) 6’s.

## 4.2 Cumulative Distribution Function

The cumulative distribution function (cdf) of a discrete RV  $X$ , denoted  $F_X(x)$ , is given by:

$$F_X(x) = P(X \leq x) = \sum_{i=1}^x f_X(i), \quad \text{for all } x; \quad (6)$$

and satisfies the following properties:

- i)*  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$ .
- ii)*  $F(x)$  is a nondecreasing function of  $x$ .
- iii)*  $F(x)$  is right-continuous; for every number  $x_0$ ,  $\lim_{x \downarrow x_0} F(x) = F(x_0)$ .

- $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$ .
- If  $x_1 < x_2$ , then  $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$ .
- If  $X \sim F_X(x)$ ,  $Y \sim F_Y(y)$ , then  $X \sim Y \iff F_X(x) = F_Y(y)$  for all  $x$ .
- Why right-continuous? (See Example 4.4.)

**Keep in mind:** All that we have said in this section (4.2) about cdf's of discrete RV, is also true for cdf's of continuous RV (details later).

**Example 4.4.** Following Example 4.1, write the cdf, graph it, and check its properties.

## 5 Continuous Model

### 5.1 Probability Density Function

The probability density function (pdf) of a continuous RV  $X$ , denoted  $f_X(x)$ , is given by:

$$\int_{x_1}^{x_2} f_X(x) dx = P(x_1 \leq X \leq x_2), \quad \text{for all } x_1 \text{ and } x_2; \quad (7)$$

and satisfies the following properties:

- i)  $f(x) \geq 0$  for all  $x$ .
- ii)  $\int_{-\infty}^{\infty} f(x) dx = 1$ .

- If  $X$  is a continuous RV, then  $P(X = x_0) = ?$
- $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$
- $\frac{P(x_0 - \varepsilon < X < x_0 + \varepsilon)}{P(x_1 - \varepsilon < X < x_1 + \varepsilon)} \rightarrow \frac{f_X(x_0)}{f_X(x_1)}$  as  $\varepsilon \rightarrow 0$ .

### 5.2 Cumulative Distribution Function

As was mentioned in section 4.2, all the definitions and results about cdf's of discrete RV, are also true for cdf's of continuous RV. The particular formula for this case is:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du, \quad \text{for all } x. \quad (8)$$

- Mixed distributions.
- $\frac{\partial F_X(x)}{\partial x} = f_X(x)$ , provided that  $F'$  exists and  $X$  is a continuous RV.

**Example 5.1.** Uniform Distribution:  $f(x) = \frac{1}{b-a}$  for  $a \leq x \leq b$ , and 0 otherwise. Graph the pdf and compute the cdf.

**Example 5.2.**  $X$  is a continuous RV with pdf:

$$f(x) = \begin{cases} ax^2 & \text{if } 0 < x < 3 \\ 0 & \text{otherwise.} \end{cases}$$

Find  $a$  and  $P(1 < x < 2)$ .

**Example 5.3.** (HOMEWORK)  $X$  is a continuous RV with cdf:

$$F(x) = \frac{1}{1 + e^{-x}} \quad \text{for all } x \in \mathfrak{R}.$$

Compute  $f(x)$  and check its properties. Find  $P(-1 < x < 2)$  using both the pdf and the cdf, and check that you get the same result in both cases.