Discrete random variables

Tan Do

Vietnamese-German University

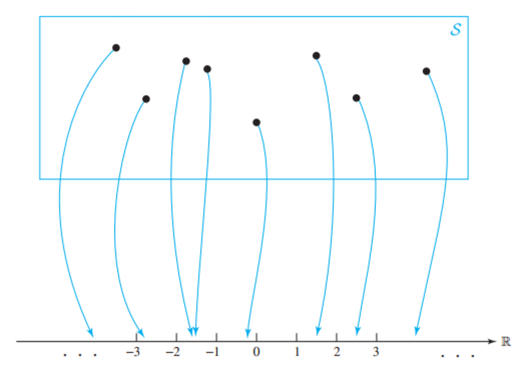
Lecture 7

In this lecture

- Discrete random variables: probability mass function and cumulative distribution function

Definition of a random variable

Definition A random variable is a function that maps each outcome of a particular experiment to a numerical value.



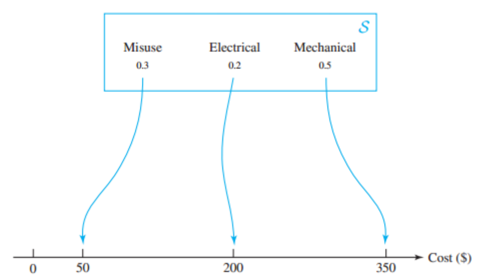
Example 1 (Machine breakdowns) The sample space for the machine breakdown problem is

$$\mathcal{S} = \{\text{electrical, mechanical, misuse}\}.$$

We can associate each outcome with a repair cost:

- Electrical failures cost an average of \$200 to repair.
- Mechanical failures \$350.
- Operator misuse \$50.

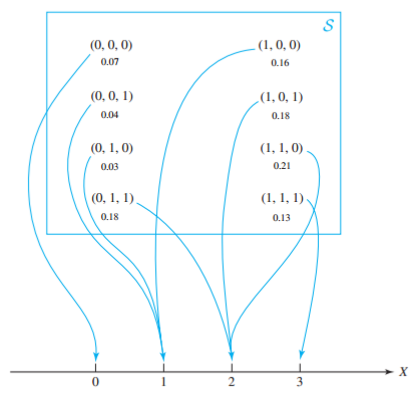
So we have created the random variable *cost* as follows.



Example 2 (Power plant operation) The sample space for the power plant example is given by the figure. Suppose we are interested in the number of plants that are operating. This creates a random variable

$$H = \text{number of power plants generating electricity}$$

which takes the values 0, 1, 2 and 3.



Example 3 (Personnel recruitment) A company has one position available for which 8 applicants have made the short list. The company strategy is to interview the applicants sequentially and to make an offer immediately to anyone they feel outstanding (without interviewing the additional applicants). If none of the first seven applicants interviewed is judged to be outstanding, the 8th applicant is interviewed and then the best of the 8 applicants is offered the job.

The company is interested in how many applicants will need to be interviewed under this strategy. A random variable

$$X = \text{number of applicants interviewed}$$

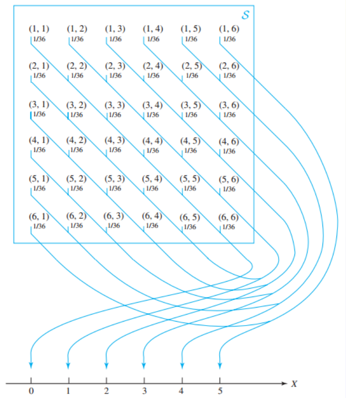
is defined to take the values 1, 2, 3, 4, 5, 6, 7 and 8.

Games of chance

Die rolling The score obtained from the roll of a die can be thought of as a random variable taking the values from 1 to 6.

If 2 dice are rolled, a random variable can be defined to be the sum of the scores, taking the values from 2 to 12.

Or we can define a random variable to be the positive difference between the scores obtained from the 2 dice, taking values from 0 to 5. (See figure)



Convention

We usually use *uppercase* letters such as X, Y, Z, etc., to refer to a random variable.

The values taken by a random variable are then labeled with the corresponding *lowercase* letter.

Example A random variable X takes the values $x = -0.6$, $x = 2.3$ and $x = 4$.

Discrete random variables

The random variables considered in the examples above are all discrete random variables because the values taken by the random variables are all discrete.

The number of values taken by a discrete random variable can be finite (Examples 1, 2, 3) or infinite (Example 4).

Later on, we will study the so-called continuous random variables, as opposed to discrete random variables.
A continuous random variable will take any value within a certain interval.

The distinction between discrete and continuous random variables is necessary since they are handled in 2 different ways.

Probability mass function

Definition The probability mass function (p.m.f) of a random variable X is a set of probability values p_i assigned to each of the values x_i taken by the *discrete* random variable.

These probability values must satisfy

$$0 \leq p_i \leq 1 \quad \text{and} \quad \sum_i p_i = 1.$$

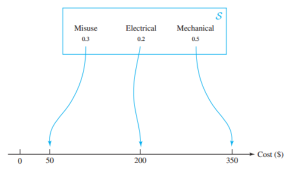
The probability that a random variable takes the value x_i is said to be p_i. This is written as

$$P(X = x_i) = p_i.$$

The probability mass function of a random variable is also called its distribution.

Example (Machine breakdowns) We have

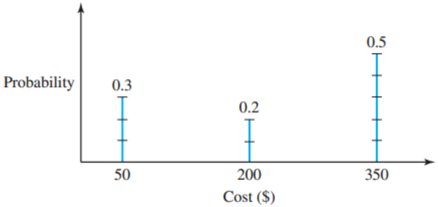
$$P(\text{cost} = 50) = 0.3, \quad \mathrm{P}(\text{cost} = 200) = 0.2 \quad \text{and} \quad \mathrm{P}(\text{cost} = 350) = 0.5.$$



This probability mass function can also be displayed in tabular form and as a line graph:

x_i	50	200	350
p_i	0.3	0.2	0.5

Tabular form



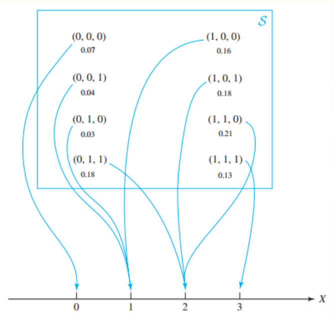
Line graph

Example (Power plant operation)

The probability mass function for the number of plants generating electricity can be inferred from the figure.

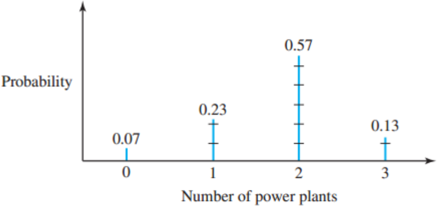
For example, the probability that no plants are generating electricity ($H = 0$) is the probability of the outcome $(0, 0, 0)$ which is 0.07.

$$P(H = 0) = 0.07.$$



x_i	0	1	2	3
p_i	0.07	0.23	0.57	0.13

Tabular form



Line graph

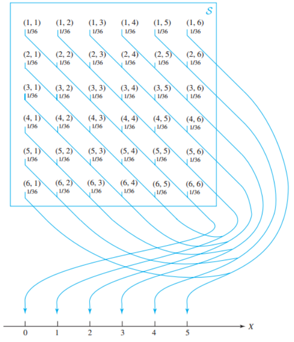
Games of chance

Die rolling

The probability mass function for the positive difference between the scores obtained from the 2 dice can be inferred from the figure.

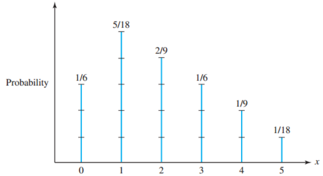
For example, the probability that the difference is 4 is

$$P(X=4) = P((1,5)) + P((5,1)) + P((2,6)) + P((6,2))$$
$$= \frac{4}{36} = \frac{1}{9}.$$



x_i	0	1	2	3	4	5
p_i	1/6	5/18	2/9	1/6	1/9	1/18

Tabular form



Line graph

Cumulative distribution function

An alternative way to specify the probabilistic properties of a random variable X is through the function

$$F(x) = P(X \leq x)$$

which is called the cumulative distribution function (c.d.f).

If the probability mass function is known, then

$$F(x) = \sum_{y \leq x} P(X = y).$$

If the cumulative distribution function is known, then

$$P(X = x) = F(x) - F(x^-),$$

where $F(x^-) = \lim_{t \to x^-} F(t)$.

Example (Machine breakdowns) The cumulative distribution function is given by

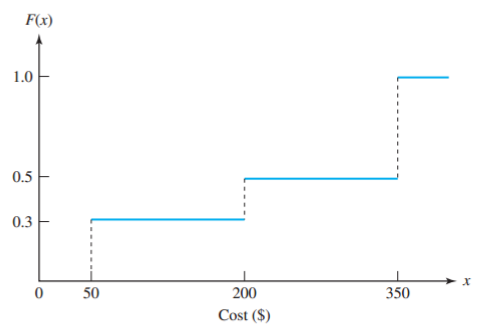
x_i	50	200	350
p_i	0.3	0.2	0.5

$$-\infty < x < 50 \Longrightarrow F(x) = P(\text{cost} \le \text{x}) = 0,$$

$$50 \le x < 200 \Longrightarrow F(x) = P(\text{cost} \le \text{x}) = 0.3,$$

$$200 \le x < 350 \Longrightarrow F(x) = P(\text{cost} \le \text{x}) = 0.3 + 0.2 = 0.5,$$

$$350 \le x < \infty \Longrightarrow F(x) = P(\text{cost} \le \text{x}) = 0.3 + 0.2 + 0.5 = 1.$$



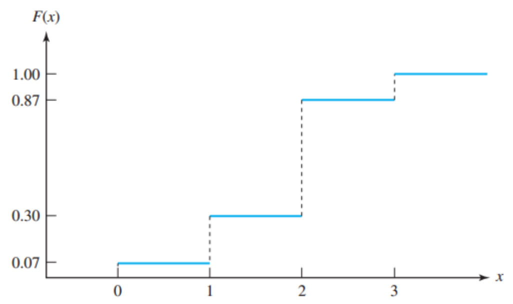
Example (Power plant operation) The cumulative distribution function for the number of plants generating electricity is given by probability mass function obtained before.

For example, the probability that no more than one plant is generating electricity is

$$F(1) = P(X \leq 1) = P(X = 0) + P(X = 1)$$

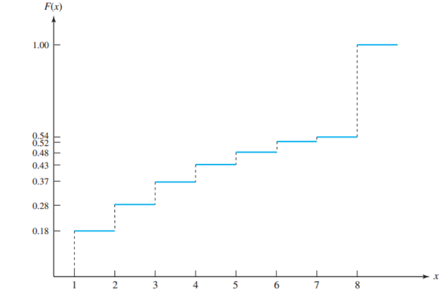
$$= 0.07 + 0.23 = 0.3.$$

x_i	0	1	2	3
p_i	0.07	0.23	0.57	0.13



Example (Personnel recruitment) The following figure provides the cumulative distribution function for the random variable X - the number of applicants interviewed.

	$F(x)$
$-\infty < x < 1$	0.00
$1 \le x < 2$	0.18
$2 \le x < 3$	0.28
$3 \le x < 4$	0.37
$4 \le x < 5$	0.43
$5 \le x < 6$	0.48
$6 \le x < 7$	0.52
$7 \le x < 8$	0.54
$8 \le x < \infty$	1.00

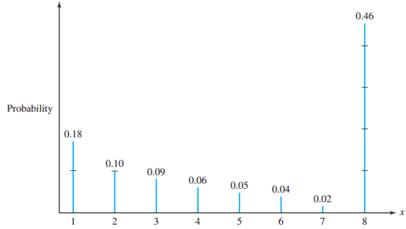


We will use this to find the probability mass function.

$$P(X=1) = F(1) - F(1^-) = 0.18 - 0 = 0.18,$$

$$P(X=2) = F(2) - F(2^-) = 0.28 - 0.18 = 0.1$$

and so on.



Line graph

Games of chance

Die rolling The cumulative distribution function for the positive difference between the scores obtained from the 2 dice can be inferred from the probability mass function derived before.

For example, the probability that the difference is no larger than 2 is

$$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{6} + \frac{5}{18} + \frac{2}{9} = \frac{2}{3}.$$