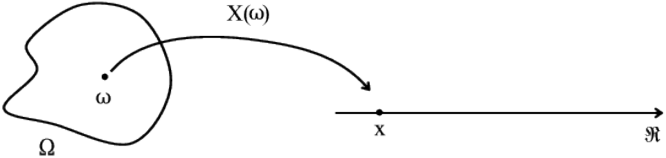
MATH 205: Statistical methods

September 29th, 2021

Lecture 8: Continuous random variables

Random variables and expectations

Random variables



Notations:

- random variables are denoted by uppercase letters (e.g., X);
- the observed values of the random variables are denoted by lowercase letters (e.g., x)

Discrete random variable

Definition

A random variables X is discrete if the set of all possible values of X

- is finite
- is countably infinite

Note: A set A is countably infinite if its elements can be put in one-to-one correspondence with the set of natural numbers, i.e, we can index the element of A as a sequence

$$A = \{x_1, x_2, \dots, x_n, \dots\}$$

Discrete random variable

Definition (Probability Distribution of a Discrete Random Variable)

The probability distribution of a discrete random variable is the set of numbers $P(X = x)$ or each value x that X can take. The distribution takes the value 0 at all other numbers. Notice that the distribution is non-negative. The probability distribution is also known as the **probability mass function**.

Represent the probability mass function

- As a table

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

- As a function:

$$p(x) = \begin{cases} \frac{1}{2}\left(\frac{2}{3}\right)^x & \text{if } x = 1, 2, 3, \dots, \\ 0 & \text{elsewhere} \end{cases}$$

Joint probability

Definition 4.4 (Joint Probability Distribution of Two Discrete Random Variables) Assume we have two random variables X and Y. The probability that X takes the value x and Y takes the value y could be written as $P(\{X = x\} \cap \{Y = y\})$. It is more usual to write it as

$$P(x, y).$$

This is referred to as the **joint probability distribution** of the two random variables (or, quite commonly, the **joint**). You can think of this as a table of probabilities, one for each possible pair of x and y values.

Marginal probability

Definition 4.6 (Marginal Probability of a Random Variable) Write $P(x, y)$ for the joint probability distribution of two random variables X and Y. Then

$$P(x) = \sum_y P(x, y) = \sum_y P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})$$

is referred to as the **marginal probability distribution** of X.

Independent variables

Definition 4.7 (Independent Random Variables) The random variables X and Y are **independent** if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all values x and y. This means that

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})P(\{Y = y\}),$$

which we can rewrite as

$$P(x, y) = P(x)P(y)$$

Example

Example

Measurements for the length and width of rectangular plastic covers for CDs are rounded to the nearest mm. Let X denote the length and Y denote the width. Assume that the joint probability of X and Y is represented by the following table

		x=length		
		129	130	131
y=width	15	0.12	0.42	0.06
	16	0.08	0.28	0.04

What is the probability distribution of X?

Example

Example

Measurements for the length and width of rectangular plastic covers for CDs are rounded to the nearest mm. Let X denote the length and Y denote the width. Assume that the joint probability of X and Y is represented by the following table

		x=length		
		129	130	131
y=width	15	0.12	0.42	0.06
	16	0.08	0.28	0.04

Are X and Y independent?

Example

Example

Assume that the joint probability of X and Y is represented by the following table

		Y		
		1	**2**	**3**
X	**1**	0.32	0.03	0.01
	2	0.06	0.24	0.02
	3	0.02	0.03	0.27

What are the probability distribution of X and Y? Are they independent?

Continuous random variables

Continuous random variables

Definition

Let X be a random variable. Suppose that there exists a nonnegative real-valued function $f : \mathbb{R} \to [0, \infty)$ such that for any subset of real numbers A, we have

$$P(X \in A) = \int_A f(x)dx$$

Then X is called **absolutely continuous** or, for simplicity, **continuous**. The function f is called the **probability density function**, or simply the **density function** of X.

Whenever we say that X is continuous, we mean that it is absolutely continuous and hence satisfies the equation above.

Properties

Let X be a continuous r.v. with density function f, then

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)\ dx = 1$
- For any fixed constant a, b,

$$P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

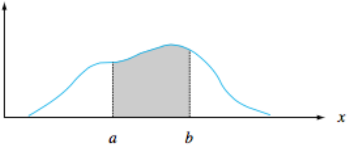


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Example

Problem

Let X be a continuous r.v. with density function

$$f(x) = \begin{cases} 2x & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Compute $P(X \in [0.25, 0.75])$

Distribution function

Definition

If X is a random variable, then the function F defined on $(-\infty, \infty)$ by

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

is called the distribution function of X.

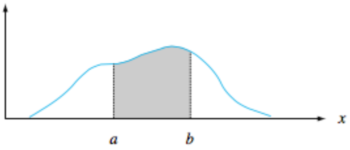


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Distribution function

For continuous random variable:

$$F(t) = P(X \leq t) = \int_{(-\infty,t]} f(x)dx$$
$$= \int_{-\infty}^{t} f(x)dx$$

Distribution function

For continuous random variable:

$$P(a \leq X \leq b) = \int_a^b f(x)\ dx = F(b) - F(a)$$

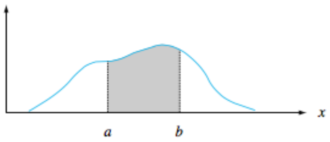


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Moreover:

$$f(x) = F'(x)$$

Example

Problem

The distribution function for the duration of a certain soap opera (in tens of hours) is

$$F(y) = \begin{cases} 1 - \frac{16}{y^2} & \text{if } y \geq 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find $P[4 \leq Y \leq 8]$.