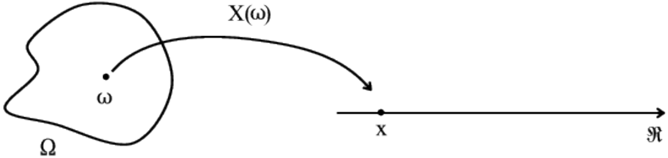
MATH 205: Statistical methods

October 6th, 2021

Lecture 10: Mean, variance, covariance

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

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:

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

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

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,

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

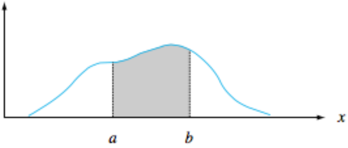


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Joint probability and marginal 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.

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

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

Expectations

Expected value: discrete variables

Definition

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P:

- The expected value of X is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{D}} x P(X = x)$$

- $f(X)$ is a also discrete random variable, which we write F and

$$\mathbb{E}[f(X)] = \sum_{u \in \mathcal{D}_f} u P(F = u) = \sum_{x \in \mathcal{D}} f(x) P(X = x)$$

  which is sometimes referred to as "the expectation of f".

Expected value: continuous variables

Definition

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability density function $p(x)$:

- The expected value of X is defined as

$$\mathbb{E}[X] = \int_{\mathcal{D}} xp(x)\ dx$$

- $f(X)$ is also a continuous random variable, which we write F and

$$\mathbb{E}[f(X)] = \int_{\mathcal{D}} f(x)p(x)\ dx$$

  which is sometimes referred to as "the expectation of f"

Expected value: multivariate

Given two discrete random variable X, Y and f is a function of (X, Y). Then $f(X, Y)$ is a also discrete random variable, and

$$\mathbb{E}[f(X, Y)] = \sum_{x,y} f(x, y)P(x, y)$$

Mean, variance and covariance (of random variables)

Mean and variance

Definition

- The mean or expected value of a random variable X is

$$\mathbb{E}[X]$$

- The variance of a random variable X is

$$var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

- The standard deviation of a random variable X is defined as

$$std(X) = \sqrt{var(X)}$$

Exercise

Problem

A random variable X has the following pmf table

X	0	1	2
probability	0.25	0.5	0.25

Compute $var(X)$.

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 $var(X)$.

Expectations are linear

Useful Facts 4.2 (Expectations Are Linear)

Write f, g for functions of random variables.

- $\mathbb{E}[0] = 0$
- for any constant k, $\mathbb{E}[kf] = k\mathbb{E}[f]$
- $\mathbb{E}[f + g] = \mathbb{E}[f] + \mathbb{E}[g]$.

A useful expression

Lemma

The variance of a random variable X can be computed by

$$var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Properties of variance

Lemma

We have

- *For any constant* k, $var[k] = 0$;
- $var[X] \geq 0$;
- $var[kX] = k^2 var[X]$

Covariance

Definition

The covariance of of two random variables X and Y is

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Covariance: properties

- Note that

$$cov(X, X) = var(X)$$

- The covariance of of two random variables X and Y can be computed as

$$cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Independent variables have zero covariance

Proposition

If X and Y are independent, then

- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
- $cov(X, Y) = 0$

Variance of sum of independent variables

Proposition

If X and Y are independent, then

$$var(X + Y) = var(X) + var(Y)$$

Exercise

Problem

A random variable X has the following pmf table

X	0	1	2
probability	0.25	0.5	0.25

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

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

Assume that X and Y are independent. You can prove that X and $-Y$ are also independent.
Compute

- $E(X+Y)$, $Var(X+Y)$
- $E(X-Y)$, $Var(X-Y)$