Chapter 4: Random variables

Yonghyun Kwon
Department of Mathematics, Korea Military Academy

Defining random variables

Random variables

- A *random variable(RV)* is a numeric quantity whose value depends on the outcome of a random event
  - We use a capital letter, like X, to denote a random variable
  - The values of a random variable are denoted with a lowercase letter, in this case x
  - For example, $P(X = x)$
- There are two types of random variables:
  - *Discrete random variables* often take only integer values
    - Example: Number of credit hours, Difference in number of credit hours this term vs last
  - *Continuous random variables* take real (decimal) values
    - Example: Cost of books this term, Difference in cost of books this term vs last

Discrete random variable

Let a random variable X denote the number of heads in two coin tosses. Find the probability distribution of X.

- Probability distribution function:

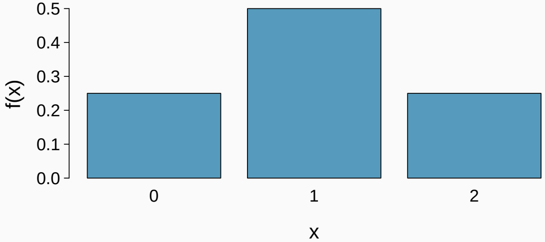
- Probability distribution table:

Probability mass function(pmf), $f(x)$

$$f(x) = \begin{cases} P(X = x_i), & x = x_i \ (i = 1, \cdots, k) \\ 0, & o.w. \end{cases}$$

Properties of probability mass function

- $0 \leq f(x) \leq 1, \quad \sum_{i=1}^{k} f(x_i) = 1.$
- $P(a < X < b) = \sum_{a<x<b} f(x).$



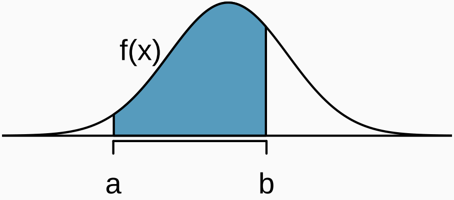
Continuous random variable

Probability density function(pdf), $f(x)$

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

Properties of probability density function

- $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$.
- $P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx$.



Practice

Consider a continuous RV X with a probability density function $f(x) = \begin{cases} cx(1-x), & 0 \leq x \leq 1 \\ 0, & o.w. \end{cases}$, where c is a constant.

- Find a constant c so that $f(x)$ is a probability density function.

- Find $P(X \leq 1/2)$.

Expectation and variance

If you repeat tossing two coins, how many heads would you expect to appear on average?

Expectation

We are often interested in *average* outcome of a random variable.

Expectation(mean) of a RV X

$$\mu = E(X) = \begin{cases} \sum_{i=1}^{k} x_i f(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Expectation of a function of RV $g(X)$

$$E\left(g(X)\right) = \begin{cases} \sum_{i=1}^{k} g(x_i) f(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Practice

Let a random variable X denote the number of heads in two coin tosses. Find the expectation of X.

The probability distribution of X is given as

x	0	1	2
$f(x)$	1/4	2/4	1/4

- $E(X) =$

Find the expectation of $Y = (X - 1)^2$.

- $E(Y) = E\left[(X-1)^2\right] =$

  $=$

Variance and standard deviation

We are also often interested in the *variability* of a random variable.

Variance and standard deviation of a RV X

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_{i=1}^{k}(x_i-\mu)^2 f(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx & X \text{ continuous} \end{cases}$$

$$\sigma = SD(X) = \sqrt{Var(X)}$$

Computing variance

$$Var(X) = E(X^2) - E(X)^2$$

Note:
$Var(X) = E[(X-\mu)^2] = E(X^2-2\mu X+\mu^2) = E(X^2)-2\mu E(X)+\mu^2 = E(X^2)-E(X)^2$

Practice

Tony is waiting for a bus. Suppose that waiting time X (hours) has the following a probability density function $f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & o.w. \end{cases}$

- Find $E(X)$:
- Find $Var(X)$:

Properties of expectation and variance

Let a and b be constants(not random).

Property of expectation

$E(aX+b) = aE(X)+b$

Properties of variance

- $Var(aX+b) = a^2 Var(X)$
- $Var(X) \geq 0$

Note: If X is a continuous random variable(discrete X can be shown similarly),

$$E(aX+b) = \int_{-\infty}^{\infty}(ax+b)f(x)dx = a\int_{-\infty}^{\infty}xf(x)dx + b\int_{-\infty}^{\infty}f(x)dx = aE(X)+b$$

$$\begin{aligned} Var(aX+b) &= E\left(aX+b-E(aX+b)\right)^2 = E\left(aX+b-aE(X)-b\right)^2 \\ &= a^2E\left(X-E(X)\right)^2 = a^2 Var(X) \end{aligned}$$

Practice

Tony is waiting for a bus. Suppose that waiting time X (hours) has the following a probability density function $f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & o.w. \end{cases}$

Let $Y = 2X + 1$.

- Find $E(Y)$:
- Find $Var(Y)$:

Joint distributions

Recall: Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad \begin{array}{l} \leftarrow \text{joint probability} \\ \leftarrow \text{marginal probability} \end{array}$$

Joint probability: Probability for two events(variables).

Marginal probability: Probability for a single event(variable).

Joint probability distribution

Joint pmf of two *discrete* RVs X and Y

$$f(x,y) = P(X = x, Y = y)$$

- $0 \leq f(x,y) \leq 1, \sum_{\forall x} \sum_{\forall y} f(x,y) = 1.$
- $P(a < X \leq b, c < Y \leq d) = \sum_{a<x\leq b} \sum_{c<y\leq d} f(x,y)$

Joint pdf of two *continuous* RVs X and Y

$$\int_a^b \int_c^d f(x,y)dydx = P(a < X < b, c < Y < d)$$

- $0 \leq f(x,y), \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1.$

Marginal probability distribution

Marginal pmf of a *discrete* RV

$$f_X(x) = \sum_{\forall y} f(x, y), \quad f_Y(y) = \sum_{\forall x} f(x, y)$$

Marginal pdf of a *continuous* RV

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Practice

In a certain suburban area, each household reported the number of cars X and the number of television sets Y that they owned. Joint pmf of X and Y is given below.

$X \backslash Y$	1	2	3	4
1	0.1	0.0	0.1	0.0
2	0.3	0.0	0.1	0.2
3	0.0	0.2	0.0	0.0

- Find the marginal pmf of X:

- Find the marginal pmf of Y:

Independence of two variables

Recall: Independence of two events

Two events A and B are independent if

$$P(A \text{ and } B) = P(A) \times P(B).$$

Independence of two random variables

X and Y are (mutually) independent if

$$f(x, y) = f_X(x) f_Y(y).$$

Practice

Joint pdf of X and Y is given by

$$f(x,y) = \begin{cases} \dfrac{3x-y}{9} & 1 < x < 3,\ 1 < y < 2 \\ 0, & o.w. \end{cases}$$

Determine whether X and Y are independent.

Covariance

Expectation of a function of RVs $g(X, Y)$

$$E\left(g(X,Y)\right) = \begin{cases} \sum_{\forall x}\sum_{\forall y} g(x,y)f(x,y) & \text{if } X, Y \text{ discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy & \text{if } X, Y \text{ continuous} \end{cases}$$

Covariance measures the joint variability of two random variables.

Covariance

Covariance of two random variables X and Y is

$$\begin{aligned} Cov(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Note: If X and Y are independent, $Cov(X, Y) = 0$, but not vice versa.

Practice

Joint and marginal pmf of X and Y are given below.

$X \backslash Y$	1	2	3	4	$f(x)$
1	0.1	0.0	0.1	0.0	0.2
2	0.3	0.0	0.1	0.2	0.6
3	0.0	0.2	0.0	0.0	0.2
$f(y)$	0.4	0.2	0.2	0.2	1.0

- Find $E(X)$ and $E(Y)$:
- Find $E(XY)$:
- Find $Cov(X, Y)$:

Note: Are X and Y independent?

Correlation

Correlation

Correlation of two random variables X and Y is

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Properties of correlation

- $-1 \leq \rho(X, Y) \leq 1$
- $\rho(X, Y) > 0$: positive linear relationship between X and Y.
- $\rho(X, Y) < 0$: negative linear relationship between X and Y.
- $\rho(X, Y) = 0$: no linear relationship between X and Y.

Linear combinations

A *linear combination* of random variables X and Y is given by

$$aX + bY$$

where a and b are some fixed constants.

The average value of a linear combination of RVs is given by

Expectation of a linear combination of two RVs

$$E(aX + bY) = a \times E(X) + b \times E(Y)$$

The variance of a linear combination of two *independent* RVs is

Variance of a linear combination of two RVs

$$Var(aX + bY) = a^2 \times Var(X) + b^2 \times Var(Y)$$

Practice

On average a notebook costs \$10 and a pen costs \$15. You plan to buy 5 notebooks (all at the same random price X) and 4 pens (all at the same random price Y). What is the expected total cost $E(5X + 4Y)$?

The standard deviation of the price is \$1.5 for a notebook and \$2 for a pen. Assuming independence between X and Y, what is the standard deviation of the total cost $SD(5X + 4Y)$?

Simplifying random variables

$2X = X + X$ can be understood as a linear combination of two random variables X and X. Thus,

$$E(X + X) = E(X) + E(X) = 2E(X),$$

but

$$Var(X + X) \neq Var(X) + Var(X),$$

because X and X themselves are not independent. Instead,

$$Var(X + X) = Var(2X) =$$

Variance of Linear Combinations: General Case

The variance formula for $aX + bY$ assumed independence. In general, covariance must be included.

Variance of a linear combination (general)

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2\,Cov(X, Y)$$

$$Var(aX \pm bY) = a^2\,Var(X) + b^2\,Var(Y) \pm 2ab\,Cov(X, Y)$$

Note: When X and Y are independent, $Cov(X, Y) = 0$, and the formula reduces to

$$Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y),$$

which is the formula from the previous slide.

Exercises

Exercises in OpenIntro Statistics 4th ed.

- **Conditional probability:**
    - Exercise 3.22, 3.42.
- **Random variables:**
    - Exercise 3.34, 3.44 (a) and (b), 3.47(or 3.45)

1. Suppose that X and Y have a joint probability density function

$$f(x,y) = \begin{cases} \frac{3}{2}y^2, & 0 \le x \le 2, 0 \le y \le 1 \\ 0, & o.w. \end{cases}$$

1. Find the marginal pdf of Y, $f_Y(y)$.
2. Find $E(Y)$ and $Var(Y)$.

2. Suppose that X and Y have a joint probability mass function

$$f(x,y) = \begin{cases} \frac{1}{12}(x+y), & x = 0,2, \text{ and } y = 0,1,2, \\ 0, & o.w. \end{cases}$$

1. Find the marginal pmf of X, $f_X(x)$. (Hint: Use a table in slide 15).
2. Find the marginal pmf of Y, $f_Y(y)$.
3. Find $E(X)$ and $E(Y)$.
4. Find $E(XY)$ and $Cov(X,Y)$.

3. Suppose that X and Y have a joint probability density function

$$f(x, y) = \begin{cases} e^{-x-y}, & x > 0,\ y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

1. Find the marginal pdfs $f_X(x)$ and $f_Y(y)$.
2. Find $E(X)$ and $E(Y)$.
3. Find $E(XY)$ and $\mathrm{Cov}(X, Y)$.
4. Find $P(X < Y)$.