Advanced Statistics

Dr. Syed Faisal Bukhari
Associate Professor
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab

Textbooks

- **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer

Variance of a Random Variable (Discrete and Continuous)

Let X be a random variable with probability distribution f(x) and mean μ.

$\sigma^2 = E[(X - \mu)^2] = \Sigma (x - \mu)^2 f(x)$, if X is discrete, and

$\sigma^2 = E[(X - \mu)^2] = \int_{x=-\infty}^{x=+\infty} (x - \mu)^2 f(x)\, dx$, if X is continous

The positive square root of the variance, *σ*, is called the **standard deviation** of *X*.

Example: Let the random variable X represent the number of automobiles used for official business purposes on a workday. Probability distribution for Company A is

x	1	2	3
f(x)	0.3	0.4	0.3

and that for company *B* is

x	0	1	2	3	4
f(x)	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company *B* is greater than that for company *A*.

$$\sigma^2 = E[(X - \mu)^2] = \Sigma (x - \mu)^2 f(x)$$

x	f(x)	xf(x)	x^2f(x)	x - μ	$(x - \mu)^2$	$(x - \mu)^2$ f(x)
1	0.30	0.30	0.30	1 - 2 = -1	1	0.30
2	0.40	0.80	1.60	2 - 2 = 0	0	0.00
3	0.30	0.90	2.70	3 - 2 = 1	1	0.30
Total	**1.0**	**2.0**	**4.6**			**0.60**

$\sigma^2 = E[(X - \mu)^2] = \Sigma\ (x - \mu)^2\ f(x)$

x	f(x)	xf(x)	x^2f(x)	x - μ	$(x - \mu)^2$	$(x - \mu)^2$ f(x)
0	0.20	0.00	0	0-2 = -2	4	0.80
1	0.10	0.10	0.10	1 - 2 = -1	1	0.10
2	0.30	0.60	1.20	2 - 2 = 0	0	0.00
3	0.30	0.90	2.70	3 - 2 = 1	1	0.30
4	0.10	0.40	1.60	4 – 2 = 2	4	0.40
Total	**1.0**	**2.0**	**5.6**			**1.60**

Alternative Method

The variance of a random variable *X* is

$\sigma^2 = E(X^2) - \mu^2$

Or

$\sigma^2 = E(X^2) - \{E(X)\}^2$

$\sigma^2 = E(X^2) - \{E(X)\}^2$

$\boldsymbol{\sigma^2}_{\boldsymbol{A}} = 4.6 - (2.0)^2$

$\boldsymbol{\sigma^2}_{\boldsymbol{A}} = 0.60$

$\sigma^2_B = 5.6 - (2.0)^2$

$\boldsymbol{\sigma^2}_{\boldsymbol{B}} = 1.60$

$\sigma^2_A = 0.60$

$\sigma^2_B = 1.60$

Clearly, the variance of the number of automobiles that are used for official business purposes is greater for company *B* than for company *A*.

Example: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the

x	0	1	2	3
f(x)	0.51	0.38	0.10	0.01

Calculate σ^2.

x	f(x)	xf(x)	$x^2f(x)$
0	0.51	0	0
1	0.38	0.38	0.38
2	0.10	0.20	0.40
3	0.01	0.03	0.09
Total		**0.61**	**0.87**

$\sigma^2 = E(X^2) - \mu^2$

Or

$\sigma^2 = E(X^2) - \{E(X)\}^2$

$= 0.87 – (0.61)^2$

$= 0.4979.$

Example: The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

$$f(x) = \begin{cases} 2(x - 1), & \text{for } 1 < x < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the **mean** and **variance of *X***.

$$\boldsymbol{\mu = E(X) = \int_{x=-\infty}^{x=+\infty} x f(x)\,dx}$$

$$= 2\int_1^2 x(x-1)\,dx$$

$$\boldsymbol{\mu = \frac{5}{3}}$$

$$E(X^2) = \int_{x=-\infty}^{x=+\infty} x^2 f(x)\,dx$$

$$= 2\int_1^2 x^2(x-1)\,dx$$

$$\boldsymbol{E(X^2) = \frac{17}{6}}$$

$$\boldsymbol{\sigma^2 = E(X^2) - \{E(X)\}^2}$$

$$= \frac{17}{6} - \left(\frac{5}{3}\right)^2$$

$$\boldsymbol{\sigma^2 = \frac{1}{18}}$$

Mean of Linear Combinations of Random Variable

If a and b are constants, then $E(aX + b) = aE(X) + b$.

Setting $a = 0$, we see that $E(b) = b$.

Setting $b = 0$, we see that $E(aX) = aE(X)$.

Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$\mu_g(X) = E[g(X)] = \sum_x g(x)f(x)$, if X is discrete random variable

$\mu_g(X) = E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)$, if X is continuous random variable

Variance

The **variance is a measure of the dispersion of a set of data points around their mean**. It quantifies how spread out the values are. Here are the key properties of variance:

1. Non-Negative

Variance is always non-negative:

$\mathrm{Var}(X) \geq 0$

2. Zero Variance

If all values are the same, variance is zero:

$\mathrm{Var}(X) = 0 \Leftrightarrow X$ is constant

3. Variance of a Constant

Variance of a constant is always zero:

$Var(c) = 0$

4. Variance of a Linear Transformation

For $Y = aX + b$, the variance is scaled:

$Var(Y) = a^2 \times Var(X)$

5. Variance of Sum of Independent Variables

For independent variables X and Y:

$Var(X + Y) = Var(X) + Var(Y)$

6. Variance of Difference of Independent Variables

For independent variables X and Y:

$Var(X - Y) = Var(X) + Var(Y)$

7. Scaling Effect

If all data points are multiplied by a constant a:

$Var(aX) = a^2 \times Var(X)$

8. Variance and Expectation

Variance can be expressed as:

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

9. Additivity for Independent Variables

The variance of the **sum of independent variables** is the **sum of their variances**:

$$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$$

Example: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	f(x)
0	$\frac{1}{4}$
1	$\frac{1}{8}$
2	$\frac{1}{2}$
3	$\frac{1}{8}$

Example: Calculate the **variance** of ***g*(*X*) = 2*X* + 3**, where *X* is a random variable with probability distribution

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	$\frac{2}{8}$	0	0
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{4}{8}$	$\frac{8}{8}$	$\frac{16}{8}$
3	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{9}{8}$
Total	$\frac{8}{8} = 1$	$\frac{12}{8} = 1.5$	$\frac{26}{8} = 3.25$

$$Var(g(X)) = Var(2X + 3)$$

$$\Rightarrow Var(g(X)) = \mathbf{2^2}\, Var(X)$$

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

$$= 3.25 - (1.5)^2$$

$$\mathbf{Var(X) = 1}$$

$$Var(g(X)) = Var(2X + 3)$$

$$= \mathbf{2^2}\, Var(X)$$

$$= 2^2 \times 1$$

$$= 4$$

Example: Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, -1 < x < 2, \\ 0, eslewhere \end{cases}$$

Find the **expected value** of $\boldsymbol{g(X) = 4X + 3}$. Also find the **variance** of the random **variable** $\boldsymbol{g(X) = 4X + 3}$.

$E[g(X)] = E(4X + 3)$

$E[g(X)] = 4E(X) + 3$

$E(X) = \frac{15}{12}$

$= \frac{5}{4}$

$Var[g(X)] = Var(4X+3)$

$= 4^2 Var(X)$

$E[g(X)] = 4E(X) + 3$

$= 4(\frac{15}{12}) + 3$

$= 8$

$E(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{dx}$

$E(X) = \int_{x=-1}^{x=2} x \times \frac{x^2}{3}\,\mathrm{dx}$

$= \left| \frac{x^4}{12} \right|_{\mathrm{x}=-1}^{\mathrm{x}=2}$

$E(X) = \frac{2^4}{12} - \frac{(-1)^4}{12}$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx$$

$$= \int_{-1}^{2} x^2 \times \frac{x^2}{3} dx$$

$$= \int_{-1}^{2} \frac{x^4}{3} dx$$

$$= \left| \frac{x^5}{15} \right|_{x=-1}^{x=2}$$

$$= \frac{2^5}{15} - \frac{(-1)^5}{15}$$

$$E(X^2) = \frac{33}{15}$$

$\text{Var}(X) = E(X^2) - \{E(X)\}^2$

$$= \frac{33}{15} - \left(\frac{5}{4}\right)^2$$

$$= \frac{51}{80}$$

$$\text{Var}[g(X)] = \text{Var}(4X+3)$$

$$= \mathbf{4^2}\ \mathbf{\text{Var}(X)}$$

$$= 16 \times \frac{51}{80}$$

$\text{Var}[g(X)] = \frac{51}{5}$

What is Covariance?

Covariance is a measure of the **relationship between two random variables**. It shows how **changes in one variable are associated with changes in another**.

Key Points about Covariance

- **Positive Covariance:** Variables increase or decrease together.

- **Negative Covariance:** One variable increases when the other decreases.

- **Zero Covariance:** No consistent relationship between the variables.

Mathematical Definition of Covariance

Covariance between two random variables X and Y is:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)\, f(x, y)$$

Covariance for Discrete and Continuous Variables

For discrete variables:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)\, f(x, y)$$

if X and Y are discrete random variables.

For continuous variables:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y)\, f(x, y)\, dx\, dy$$

if X and Y are continuous random variables.

Covariance of X and Y (Discrete Case)

Let X and Y be random variables with joint probability distribution f(x, y).

The covariance of X and Y is given by:

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)\, f(x, y)$$

if X and Y are discrete random variables.

Covariance of X and Y (Continuous Case)

The covariance of X and Y is given by:

$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] =$
$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty}$ **(x – μX)(y – μY) f(x, y) dx dy** if X and Y are continuous random variables.

Example : Two ballpoint pens are selected at random from a box that contains **3 blue pens**, **2 red pens**, and **3 green pens**. If ***X*** is the **number of blue pens** selected and ***Y*** is the **number of red pens** selected,

(a) Find the joint probability function $f(x, y)$,

(b) Find the expected value of $g(X, Y) = XY$.

(c) Find the covariance of X and Y

f(x, y)		x			Row totals
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{6}{28}$	$\frac{6}{28}$	0	$\frac{12}{28}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column totals		$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$	$\frac{28}{28} = 1$

Marginal Distribution of x

x = 0	0	1	2	Total
g(x)	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$	$\frac{28}{28} = 1$

Marginal Distribution of y

y = 0	0	1	2	Total
h(y)	$\frac{15}{28}$	$\frac{12}{28}$	$\frac{1}{28}$	$\frac{28}{28} = 1$

$\mu_g(X,Y) = E[g(X, Y)] = \sum_x \sum_y g(x, y)f(x, y)$

$\mu_g(X,Y) = \sum_{x=0}^{x=2} \sum_{y=0}^{y=2} xyf(x,y)$

$E(XY) = (0)(0) \times f(0,0) + (0)(1) \times f(0,1) + (0)(2) \times f(0,2) + (1)(0) \times f(1,0) + (1)(1) \times f(1,1) + (1)(2) \times f(1,2) + (2)(0) \times f(2,0) + (2)(1) \times f(2,1) + (2)(2) \times f(2,2)$

$E(XY) = (0)(0) \times \frac{3}{28} + (0)(1) \times \frac{6}{28} + (0)(2) \times \frac{1}{28} + (1)(0) \times \frac{9}{28} + (1)(1) \times \frac{6}{28} + (1)(2) \times 0 + (2)(0) \times \frac{3}{28} + (2)(1) \times 0 + (2)(2) \times 0 = \frac{6}{28}$

$E(XY) = \frac{3}{14}$

$\mu_X = \sum_{x=0}^{2} x g(x)$

$\mu_X = (0)(\frac{10}{28}) + (1)(\frac{15}{28}) + (2)(\frac{3}{28})$
$= \frac{3}{4}$

$\mu_y = \sum_{y=0}^{2} y g(y)$

$= (0)(\frac{15}{28}) + (1)(\frac{3}{7}) + (2)(\frac{1}{28})$
$= \frac{1}{2}$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y$$

$$= \frac{3}{14} - \frac{3}{4} \times \frac{1}{2}$$

$$= -\frac{9}{56}$$