* $E(X) = \mu$

1.3.2 Variance

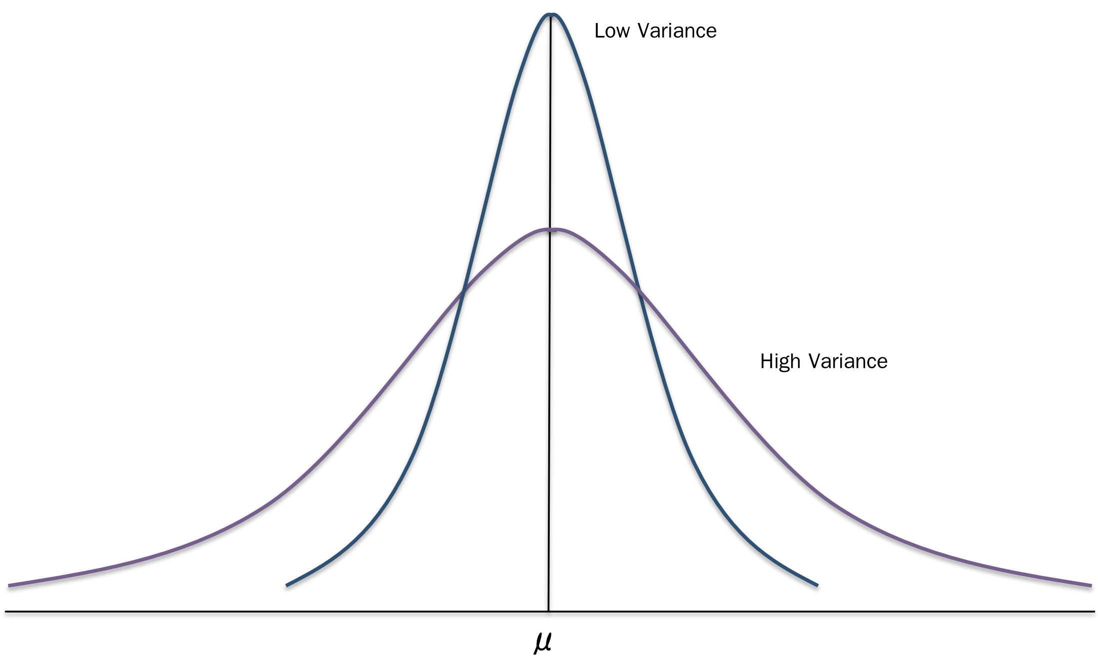
Variance is the measure of dispersion of the value of variable around the expected value. The higher the variance, the more dispersing the random variable (Figure 3). If X is the random variable with expected value μ, we get;

$$Var(X) = \sigma_X^2 = E[X - E(X)]^2 = E(X)^2 - \mu^2 \tag{1.2}$$

From,

$$\begin{aligned} Var(X) &= \sigma_X^2 \\ &= E[X - E(X)]^2 \\ &= E[X^2 - 2XE(X) + (E(X))^2] \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 \\ &= E(X)^2 - \mu^2 \end{aligned}$$

Figure 3: Distribution of Random Variables with Different Variance



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

$$= E[(X - E(X))(X - E(X))]$$

$$= E[X^2 - 2X\,E(X) + E(X)^2]$$

$$= E(X^2) + E(-2X\,E(X)) + E((E(X))^2)$$

$$= E(X^2) - 2E(X\,E(X)) + (E(X))^2$$

$$= E(X^2) - 2E(X)\,E(X) + (E(X))^2$$

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

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

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

$$E(X^2) \neq (E(X))^2$$

$$E(E(X)) = E(X)$$

"The law of iterated expectations"

x	-2	1	2
$f(x)$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{2}{8}$

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

$$E(X^2) = \sum_x x^2 f(x)$$

$$= (-2)^2\left(\frac{5}{8}\right) + 1^2\left(\frac{1}{8}\right) + 2^2\left(\frac{2}{8}\right)$$

$$= \frac{29}{8}$$

$$E(X) = \sum_x x f(x)$$

$$= -2\left(\frac{5}{8}\right) + 1\left(\frac{1}{8}\right) + 2\left(\frac{2}{8}\right)$$

$$= \frac{-5}{8}$$

$$Var(X) = \frac{29}{8} - \left(\frac{-5}{8}\right)^2 = \frac{207}{64}$$

//

Practice

X	-3	0	1
$f(x)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

$E(X) =$

$Var(X) =$

Answers.

X	-3	0	1
$f(x)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

$$E(X) = \sum_x x f(x)$$

$$= -3\left(\frac{2}{5}\right) + 0\left(\frac{1}{5}\right) + 1\left(\frac{2}{5}\right) = -\frac{4}{5}$$

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

$$E(X^2) = -3^2\left(\frac{2}{5}\right) + 0^2\left(\frac{1}{5}\right) + 1^2\left(\frac{2}{5}\right)$$

$$= \frac{20}{5}$$

$$Var(X) = \frac{20}{5} - \left(-\frac{4}{5}\right)^2 = \frac{84}{25}$$

Important properties of expected value include;

1. $Var(b) = 0$

2. $Var(aX+b) = a^2 Var(X)$

3. $Var(X \pm Y) = Var(X) + Var(Y)$; given that X and Y are independent

4. $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$

given that X and Y are independent

where a and b are constant.

1.3.3 Conditional Variance

The conditional variance of X is given Y = y is defined as following:

$$\begin{aligned} var(X|Y=y) &= E\left\{[X - E(X|Y=y)]^2 | Y=y\right\} \\ &= \sum_x [X - E(X|Y=y)]^2 f(x|Y=y) \\ &= \int_{-\infty}^{\infty} [X - E(X|Y=y)]^2 f(x|Y=y) dx \end{aligned} \tag{1.3}$$

Example

Properties of variance

1) The variance of a constant is zero.

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

$$V(aX+b) = E[(ax+b) - E(ax+b)]^2$$

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

$$= E[(ax+\cancel{b}) - aE(X) - \cancel{b}]^2$$

$$= E[ax - aE(x)]^2$$

$$= a^2 \, E[\underbrace{X - E(X)]^2}_{Var(X)}$$

$$= a^2 \, Var(X)$$

3 If X and Y are independent random variables, then

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

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

$$E(X) = \mu$$

$$VAR(X-Y) = E[(X-Y)^2] - [E(X-Y)]^2 \quad ①$$

$$(a-b)^2 = a^2 - 2ab + b^2$$

$$= E(X^2+Y^2-2XY) - (E(X) - E(Y))^2 \quad ②$$

$$= E(X^2) + E(Y^2) - 2E(XY) - (E(X)^2 + E(Y)^2 - 2E(X)E(Y))$$

$$= \underbrace{E(X^2) - E(X)^2}_{Var(X)} + \underbrace{E(Y^2) - E(Y)^2}_{VAR(Y)} - \underbrace{2(E(XY) - E(X)E(Y))}_{2COV(X,Y)}$$

$$2COV(X,Y) = 0$$

when X and Y are independent random variables.

Properties of conditional expectation and conditional variance

① If $f(x)$ is a function of X, then

$$E(f(X)|X) = f(X),$$

that is, the function of X behaves as a constant in computation of its expectation conditional on X.

e.g. $E(X^3|X) = E(X^3)$

② If $f(X)$ and $g(X)$ are functionals of X, then

$$E[f(X)Y + g(X)|X] = f(X)E(Y|X) + g(X)$$

e.g. $E(XY + cX^2|X) = XE(Y|X) + cX^2$

where c is a constant.

③ If X and Y are independent

$$E(Y|X) = E(Y)$$

④ The law of iterated expectations

$$E(Y) = E_x[E(Y|X)]$$

⑤ If X and Y are independent, then

$$Var(Y|X) = Var(Y)$$

⑥ $Var(Y) = E[var(Y|X)] + var[E(Y|X)]$

The unconditional variance of Y is equal to expectation of the conditional variance of Y plus the variance of the conditional expectation of Y

1.3.4 Covariance

Theorem. Let X and Y be two random variables with means μ_x and μ_y, respectively. Then, we can define the covariance between these two variables as following:

$$cov(X,Y) = E\{(X-\mu_x)(Y-\mu_y)\} = E(XY) - \mu_x\mu_y \tag{1.4}$$

If X and Y are continuous random variables we can calculate their cov(X,Y):

$$\begin{aligned} cov(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_x)(Y-\mu_y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}XYf(x,y)dxdy - \mu_x\mu_y \end{aligned} \tag{1.5}$$

Properties of Covariance

1. If X and Y are independent, the covariance between X and Y is zero.

Proof:

$$\begin{aligned} cov(X,Y) &= E(XY) - \mu_x\mu_y \\ &= \mu_x\mu_y - \mu_x\mu_y \\ &= 0 \end{aligned}$$

2. $cov(a+bX, c+dY) = bd * cov(X,Y)$, where a,b,c, and d are constants.

Proof:

Example Suppose the join PDF of random variables X and Y can be represented as in the below table. What is the covariance between X and Y?

		X			
		1	2	3	
Y	1	0.25	0.25	0	$f(Y=1) =$ 0.50
	2	0	0.25	0.25	$f(Y=2) =$ 0.50
		$f(X=1) =$ 0.25	$f(X=2) =$ 0.50	$f(X=3) =$ 0.25	$f(X) =$ 1 $f(Y) =$ 1

$COV = E(XY) - E(X)E(Y)$

$E(X) = \sum_x x f(x)$

$= 1(0.25) + 2(0.50) + 1(0.25) = 1.50$

$E(Y) = \sum_y y f(y)$

$= 1(0.50) + 2(0.50) = 1.50$

$E(XY) = \sum_y \sum_x xy f(x,y)$

$= (1)(1)(0.25) + (1)(2)(0.25) + (1)(3)(0) +$

$(1)(2)(0) + (2)(2)(0.25) + (3)(2)(0.25)$

$= ☺$

$COV = ☺ - (1.50)(1.50)$

$=$

Next, we will turn our attention to seeing how we can apply the covariance to calculate the correlation between the random variables X and Y

1.3.5 Correlation

When we calculate the covariance of X and Y, it reflects the units of both random variables. However, it is useful to have **a dimensionless measure of dependency** by calculating the correlation instead.

Definition Let X and Y be any two random variables (discrete or continuous) with standard deviation σ_X and σ_Y, respectively. The **correlation coefficient** of X and Y, denoted **corr(X,Y)** or ρ_{XY} (the greek letter "rho") is defined as:

$$\rho_{XY} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{cov(x,y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Example Suppose the join PDF of random variables X and Y can be represented as in the below table. What is the correlation between X and Y?

			X		
		1	2	3	
	1	0.25	0.25	0	$f(Y=1)$ =0.5
Y	2	0	0.25	0.25	$f(Y=2)$ =0.5
		$f(X=1)$ = 0.25	$f(X=2)$ = 0.5	$f(X=3)$ =0.25	$f(X)=1$ $f(Y)=1$

cov =

$var(x) = E(x^2) - [E(x)]^2$

$E(X) = 1.50$

$E(X^2) = 1^2(0.25) + 2^2(0.5) + 3^2(0.5)$

$= 0.25 + 2 + 4.5 = 6.75$

$var(X) = 6.75 - (1.5)^2$

$= 4.5$

$$var(X) = \sigma_x^2 = 4.5$$

$$\sigma_x = \sqrt{4.5}$$

$$var(Y) = \sigma_y^2 = E(Y^2) - (E(Y))^2$$

$$E(Y) = 1.50$$

$$E(Y)^2 = 2.25$$

$$E(Y^2) = 1^2(0.5) + 2^2(0.5)$$

$$= 0.5 + 2 = 2.5$$

$$var(Y) = \sigma_y^2 = 2.5 - 2.25 = 0.25$$

$$\sigma_y = \sqrt{0.25}$$

$$\rho = \frac{cov(X,Y)}{\sigma_x \sigma_y} = \frac{\quad}{\sqrt{(4.5)(0.25}}$$

From the definition, ρ_{XY} is measure of linear association between two random variables. The value of ρ lies between -1 and +1, $-1 \leq \rho_{XY} \leq +1$. We can interpret the value of correlation as:

▶ If $\rho_{XY} = 1$, then X and Y are perfectly, positively, linearly correlated.
▶ If $\rho_{XY} = -1$, then X and Y are perfectly, negatively, linearly correlated.
▶ If $\rho_{XY} = 0$, then X and Y are completely,un-linearly correlated. This means that X and Y may correlated in some other manner i.e. a parabolic manner., but NOT in a linear manner
▶ If $\rho_{XY} \leq 0$, then X and Y are positively, linearly correlated, but NOT perfectly.
▶ If $\rho_{XY} \geqq 0$, then X and Y are negatively, linearly correlated, but NOT perfectly.

Theorem. If X and Y are independent random variables, then:

$$corr(X,Y) = cov(X,Y) = 0$$

Example: Let X = the outcome of a fair, black, 6-sided die.
Let Y =outcome of a fair, red, 4-sided die.
What is the covariance of X and Y? What is the correlation of X and Y?

practice

NOTE: The converse of the theorem is NOT NECESSARILY CORRECT!

Example: Let X and Y be two discrete random variables with the following join PDF:

			X		
		0	1	2	
	0	0	0.20	0.10	$f(Y=0)$ =
Y	1	0.20	0.40	0	$f(Y=1)$ =
	2	0.10	0	0	$f(Y=2)$ =
		$f(X=0)$ =	$f(X=1)$ =	$f(X=2)$ =	

What is the correlation between X and Y? And, are X and Y independent?

1.3.6 Variances of Correlated Variables

Let X and Y be two random variables, then

$$\begin{aligned} var(X+Y) &= var(X)+var(Y)+2cov(X,Y) \\ &= var(X)+var(Y)+2\rho\sigma_x\sigma_y \\ var(X-Y) &= var(X)+var(Y)-2cov(X,Y) \\ &= var(X)+var(Y)-2\rho\sigma_x\sigma_y \end{aligned} \tag{1.6}$$

The generalized result:
Let $\sum_{i=1}^{n} X_i = X_1+X_2+\cdots+X_n$, then the variance of the linear combination $\sum X_i$ is:

$$\begin{aligned} var\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} var(X_i)+2\sum_{i<j}\sum cov(X_i,X_j) \\ &= \sum_{i=1}^{n} var(X_i)+2\sum_{i<j}\sum \rho_{ij}\sigma_i\sigma_j \end{aligned} \tag{1.7}$$

Example:
what is the $\mathrm{var}(X_1+X_2+X_3)$?

$Var(X_1+X_2+X_3) = Var(X_1) + Var(X_2) + Var(X_3) + 2Cov(X_1,X_2) + 2Cov(X_2,X_3) + 2Cov(X_1,X_3)$

$= Var(X_1) + Var(X_2) + Var(X_3) + 2\rho_{12}\sigma_1\sigma_2 + 2\rho_{13}\sigma_1\sigma_3 + 2\rho_{23}\sigma_2\sigma_3$

1.3.7 Higher Moments of Probability Distributions

In the previous subsection, we have already discussed about mean, variance, and covariance as the measures of the first and second moments of univariate and multivariate PDFs. Besides the first two moments, we are occasionally interested in the higher moments such as the third and fourth moments which are normally applied in studying the "Shape" of the distribution. In general, the r^{th} moments about the mean is defined as

$$r^{th}\textbf{moment} : E(X-\mu)^r$$

By the definition of r^{th} moments, we can easily define the third and fourth moments as:

Third moment:

$$E(X-\mu)^3$$

Fourth moment:

$$E(X-\mu)^4$$

We can study the shape of the distribution by calculating **skewness** and **kurtosis**.

SKEWNESS is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

One measure of skewness is defined as:

$$\begin{aligned} S &= \frac{E(X-\mu)^3}{\sigma^3} \\ &= \frac{\text{third moment about the mean}}{\text{cube of the standard deviation}} \end{aligned} \tag{1.8}$$

KURTOSIS is a measure of the peakedness of the probability distribution of a real-valued random variable

We can also measure kurtosis as:

$$\begin{aligned} S &= \frac{E(X-\mu)^4}{\sigma^4} \\ &= \frac{\text{fourth moment about the mean}}{\text{square of the second moment}} \end{aligned} \tag{1.9}$$

♣ **Platykurtic (fat or short-tailed)** $\Longrightarrow$ PDFs with Kurtosis < 3
♣ **Leptokurtic (slim or long-tailed)** $\Longrightarrow$ PDFs with Kurtosis > 3
♣ **Mesokurtic (which is the normal distribution)** $\Longrightarrow$ PDFs with Kurtosis $= 3$

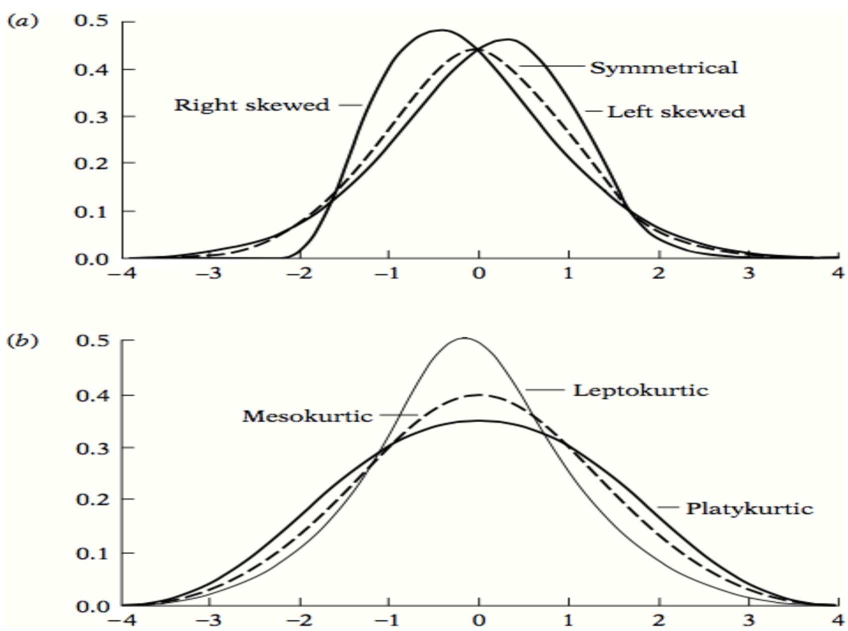


Figure 1.1: (a) Skewness; (b)Kurtosis