Variances and covariances

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

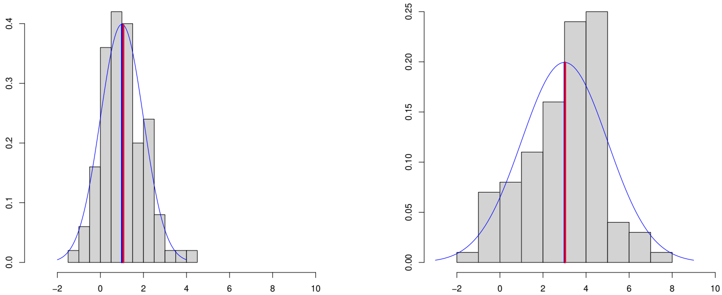
October 15, 2024

Mean and variance

Let X denote random variable (i.e. random outcome of experiment, measurement,...)

The expected value $\mathbb{E}X$ and the variance $\mathbb{V}\mathrm{ar}X$ are summaries that describe the center and the spread of the distribution of X.

Example: samples of normally distributed X:



$\mathbb{E}X = \mathbb{V}\mathrm{ar}X = 1$ $\bar{X} = 1.07$ $\mathbb{E}X = 3$ $\mathbb{V}\mathrm{ar}X = 4$ $\bar{X} = 3.04$

Standard deviation:

$$\text{sd}(X) = \sqrt{\mathbb{V}\text{ar}X}$$

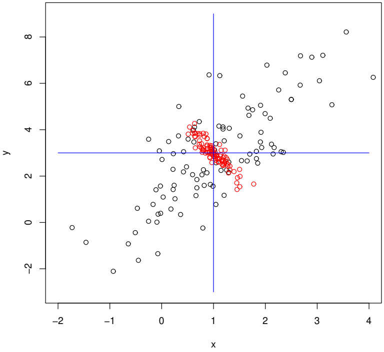
Same unit as X - if X is measurement in m then standard deviation also in m.

Convenient summary for normal distribution - can be directly translated into 'width' of distribution:

- $\mathbb{E}X \pm 1.96\text{sd}(X)$ 95% probability interval.
- $\mathbb{E}X \pm 2\text{sd}(X)$ 95.4% probability interval.
- $\mathbb{E}X \pm 3\text{sd}(X)$ 99.7% probability interval.

Covariance and correlation

Samples of correlated random variables X and Y:



Covariance is the expected value of products of deviations from X and Y from their mean:

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

Black dots: $\mathbb{V}\mathrm{ar}X = 1$
$\mathbb{V}\mathrm{ar}\,Y = 4$, $\mathbb{Cov}[X, Y] = 1.5$

Red dots: $\mathbb{V}\mathrm{ar}X = 0.0625$
$\mathbb{V}\mathrm{ar}\,Y = 0.3$
$\mathbb{Cov}[X, Y] = -0.125$

Covariance measure of association between X and Y but can be large just because variances of X and Y are large.

Correlation is covariance normalized by standard deviations:

$$\rho = \mathbb{C}\mathrm{orr}[X, Y] = \frac{\mathbb{C}\mathrm{ov}[X, Y]}{\mathrm{sd}X\mathrm{sd}Y}$$

Black dots: $\mathbb{C}\mathrm{orr}[X, Y] = 0.75$ Red dots: $\mathbb{C}\mathrm{orr}[X, Y] = -0.91$

Correlation useful measure of 'strength' of association between two random variables.

Correlation always between -1 and 1 and does not depend on scaling (unit) of X and Y.

Results regarding variances and covariances

Let $Y, Y_1, \ldots, Y_n$ and $X_1, \ldots, X_m$ be random variables and $a, a_1, \ldots, a_n$ be fixed coefficients.

$$\mathbb{E}(aY) = a\mathbb{E}Y \qquad \mathbb{V}\mathrm{ar}(aY) = a^2\mathbb{V}\mathrm{ar}\,Y$$

$$\mathbb{C}\mathrm{ov}(a, Y) = 0 \qquad \mathbb{C}\mathrm{ov}(Y, Y) = \mathbb{V}\mathrm{ar}(Y)$$

$$\mathbb{C}\mathrm{ov}(a_1 Y_1, a_2 Y_2) = a_1 a_2 \mathbb{C}\mathrm{ov}(Y_1, Y_2)$$

$$\begin{aligned}\mathbb{C}\mathrm{ov}(Y_1 + Y_2, X_1 + X_2) =& \mathbb{C}\mathrm{ov}(Y_1, X_1) + \mathbb{C}\mathrm{ov}(Y_1, X_2) \\ &+ \mathbb{C}\mathrm{ov}(Y_2, X_1) + \mathbb{C}\mathrm{ov}(Y_2, X_2)\end{aligned}$$

The last rule says that covariance between sums splits into a sum of covariances for each pair where one variable comes from the first sum and the other from the second sum.

More generally,

$$\mathbb{C}\mathrm{ov}\Big(\sum_{i=1}^{n} Y_i, \sum_{j=1}^{m} X_j\Big) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{C}\mathrm{ov}(Y_i, X_j)$$

Results - continued

It follows from the rules on the previous slide that

$$\mathbb{V}\mathrm{ar} \sum_{i=1}^{n} Y_i = \mathbb{C}\mathrm{ov}(\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{C}\mathrm{ov}(Y_i, Y_j)$$

A special case ($n = 2$) is

$$\mathbb{V}\mathrm{ar}(Y_1 + Y_2) = \mathbb{V}\mathrm{ar}\, Y_1 + \mathbb{V}\mathrm{ar}\, Y_2 + 2\mathbb{C}\mathrm{ov}(Y_1, Y_2)$$

If Y_1 and Y_2 are independent then

$$\mathbb{C}\mathrm{ov}(Y_1, Y_2) = 0$$

From this and the previous result it follows that if $Y_1, \ldots, Y_n$ are independent then

$$\mathbb{V}\mathrm{ar} \left(\sum_{i=1}^{n} Y_i \right) = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}\, Y_i$$

Finally, the correlation coefficient does not depend on scaling

$$\mathbb{C}\mathrm{orr}(a_1 Y_1, a_2 Y_2) = \mathbb{C}\mathrm{orr}(Y_1, Y_2)$$

This follows because

$$\begin{aligned}\mathbb{C}\mathrm{orr}(a_1 Y_1, a_2 Y_2) &= \frac{\mathbb{C}\mathrm{ov}(a_1 Y_1, a_2 Y_2)}{\sqrt{\mathbb{V}\mathrm{ar}\, a_1 Y_1}\sqrt{\mathbb{V}\mathrm{ar}\, a_2 Y_2}} \\ &= \frac{a_1 a_2 \mathbb{C}\mathrm{ov}(Y_1, Y_2)}{a_1 a_2 \sqrt{\mathbb{V}\mathrm{ar}\, Y_1}\sqrt{\mathbb{V}\mathrm{ar}\, Y_2}} = \mathbb{C}\mathrm{orr}(Y_1, Y_2)\end{aligned}$$

Subprime loans and the financial crisis

The financial crisis (2007-2009) was caused by wrong assessment (or ignorance) of risk on mortgage loans.

Some experts claim: risk managers failed to take into account that mortage loans are correlated due to dependence on common economic trends

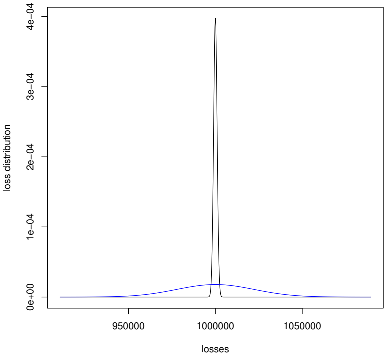
Why does correlation matter ?

Suppose $X_1, \ldots, X_{1000}$ represent losses in portfolio of 1000 loans. Assume common mean $\mathbb{E}X_i = 1000$ and variance $\mathbb{V}\mathrm{ar}X_i = 1000$. Risk manager needs to consider variance of total loss $\sum_{i=1}^{1000} X_i$.

This variance is $1000 \cdot 1000 = 1$mio (sd 1000) if losses independent.

If instead losses correlated, $\mathbb{C}\mathrm{orr}(X_i, X_j) = 0.5$, then variance is $1000 \cdot 1000 + 1000 \cdot 0.5 \cdot 1000 \cdot (1000 - 1) = 500.5$mio (sd 22372).

Loss distribution



95% interval for loss distribution is 22 times wider when taking into account correlation (blue distribution)

Exercises

Assume X has variance 2 and Y has variance 3. Assume X and Y are independent.

- What is the variance of $2X$?
- What is the covariance $\mathbb{C}\mathrm{ov}(X, Y)$?
- What is $\mathbb{V}\mathrm{ar}(X + Y)$?
- What is $\mathbb{C}\mathrm{ov}(X, X + Y)$?
- Assume now that $\mathbb{C}\mathrm{ov}(X, Y) = 2$. What is then $\mathbb{V}\mathrm{ar}(X + Y)$?
- Assume again that $\mathbb{C}\mathrm{ov}(X, Y) = 2$ what is $\mathbb{C}\mathrm{orr}(X, Y)$?

(solutions on next slide)

8, 0, 5, 2, 9, 0.81