STA286 Lecture 12

Neil Montgomery

Last edited: 2017-02-07 10:55

variance rules

$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

variance rules

$$\mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X)$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2E((X - EX)(Y - EY))$$

variance rules

$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2E((X - EX)(Y - EY))$$

$E((X - EX)(Y - EY))$ is called the *covariance* of X and Y, or $\mathrm{Cov}(X, Y)$, or σ_{XY}

variance rules

$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2E((X - EX)(Y - EY))$$

$E((X - EX)(Y - EY))$ is called the *covariance* of X and Y, or $\mathrm{Cov}(X, Y)$, or σ_{XY}

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)\,E(Y)$$

variance rules

$$\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2E((X - EX)(Y - EY))$$

$E((X - EX)(Y - EY))$ is called the *covariance* of X and Y, or $\mathrm{Cov}(X, Y)$, or σ_{XY}

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)\,E(Y)$$

So, $X \perp Y$ implies Cov(X,Y) $= 0$, in which case the variance of the sum is the sum of the variances.

covariance

$\text{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

covariance

Cov(X, Y) is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

covariance

$\text{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

- with high probability, larger values of X and Y happen at the same time, and

covariance

$\mathrm{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

- with high probability, larger values of X and Y happen at the same time, and
- with high probability, smaller values of X and Y happen at the same time, then:

covariance

$\mathrm{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

- with high probability, larger values of X and Y happen at the same time, and
- with high probability, smaller values of X and Y happen at the same time, then:
  - with high probability $(X - EX)(Y - EY)$ will be **positive**.

covariance

$\mathrm{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

- with high probability, larger values of X and Y happen at the same time, and
- with high probability, smaller values of X and Y happen at the same time, then:
  - with high probability $(X - EX)(Y - EY)$ will be **positive**.

2. If:

covariance

$\text{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

- with high probability, larger values of X and Y happen at the same time, and
- with high probability, smaller values of X and Y happen at the same time, then:
    - with high probability $(X - EX)(Y - EY)$ will be **positive**.

2. If:

- with high probability, larger values of X and smaller values of Y happen at the same time, and

covariance

$\mathrm{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:

   - with high probability, larger values of X and Y happen at the same time, and
   - with high probability, smaller values of X and Y happen at the same time, then:
     - with high probability $(X - EX)(Y - EY)$ will be **positive**.

2. If:

   - with high probability, larger values of X and smaller values of Y happen at the same time, and
   - with high probability, smaller values of X and larger values of Y happen at the same time, then:

covariance

$\text{Cov}(X, Y)$ is a numerical summary of a certain kind of relationship that might exist between two distributions.

It is a measure of *linear* relationship, in the following sense(s):

1. If:
   - with high probability, larger values of X and Y happen at the same time, and
   - with high probability, smaller values of X and Y happen at the same time, then:
     - with high probability $(X - EX)(Y - EY)$ will be **positive**.

2. If:
   - with high probability, larger values of X and smaller values of Y happen at the same time, and
   - with high probability, smaller values of X and larger values of Y happen at the same time, then:
     - with high probability $(X - EX)(Y - EY)$ will be **negative**.

correlation coefficient

Covariance is in the unit of the product of the units of X and Y, so its magnitude is uninformative.

correlation coefficient

Covariance is in the unit of the product of the units of X and Y, so its magnitude is uninformative.

A scale-free version is called *correlation coefficient* (or just "correlation"):

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

correlation coefficient

Covariance is in the unit of the product of the units of X and Y, so its magnitude is uninformative.

A scale-free version is called *correlation coefficient* (or just "correlation"):

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Positive and negative correlation mean the same as positive and negative covariance; in addition, the correlation of different pairs of distributions can be compared. Also:

$$-1 \leqslant \rho \leqslant 1$$

discrete example - positive correlation close to 1

		X		
Y	-1	0	1	Marginal
-1	0.30	0.02	0.01	0.33
0	0.02	0.30	0.02	0.34
1	0.01	0.02	0.30	0.33
Marginal	0.33	0.34	0.33	1.00

discrete example - positive correlation close to 1

	X			
Y	-1	0	1	Marginal
-1	0.30	0.02	0.01	0.33
0	0.02	0.30	0.02	0.34
1	0.01	0.02	0.30	0.33
Marginal	0.33	0.34	0.33	1.00

$E(X) = E(Y) = 0$, so $\text{Cov}(X, Y) = E(XY)$.

discrete example - positive correlation close to 1

		X		
Y	-1	0	1	Marginal
-1	0.30	0.02	0.01	0.33
0	0.02	0.30	0.02	0.34
1	0.01	0.02	0.30	0.33
Marginal	0.33	0.34	0.33	1.00

$E(X) = E(Y) = 0$, so $\text{Cov}(X, Y) = E(XY)$.

In this case (tedious exercise) $\rho = 0.879$.

discrete example - positive correlation close to 1

		X		
Y	-1	0	1	Marginal
-1	0.30	0.02	0.01	0.33
0	0.02	0.30	0.02	0.34
1	0.01	0.02	0.30	0.33
Marginal	0.33	0.34	0.33	1.00

$E(X) = E(Y) = 0$, so $\mathrm{Cov}(X, Y) = E(XY)$.

In this case (tedious exercise) $\rho = 0.879$.

The probability is strongly concentrated along the "$X = Y$ diagonal".

discrete example - negative correlation close to 1

	X			
Y	-1	0	1	Marginal
-1	0.01	0.02	0.30	0.33
0	0.02	0.30	0.02	0.34
1	0.30	0.02	0.01	0.33
Marginal	0.33	0.34	0.33	1.00

discrete example - negative correlation close to 1

		X		
Y	-1	0	1	Marginal
-1	0.01	0.02	0.30	0.33
0	0.02	0.30	0.02	0.34
1	0.30	0.02	0.01	0.33
Marginal	0.33	0.34	0.33	1.00

In this case (tedious exercise) $\rho = -0.879$.

discrete example - negative correlation close to 1

		X		
Y	-1	0	1	Marginal
-1	0.01	0.02	0.30	0.33
0	0.02	0.30	0.02	0.34
1	0.30	0.02	0.01	0.33
Marginal	0.33	0.34	0.33	1.00

In this case (tedious exercise) $\rho = -0.879$.

The probability is strongly concentrated along the "$X = -Y$ diagonal".

discrete exmaple - $X \perp Y$

		X		
Y	-1	0	1	Marginal
-1	0.1089	0.1122	0.1089	0.33
0	0.1122	0.1156	0.1122	0.34
1	0.1089	0.1122	0.1089	0.33
Marginal	0.33	0.34	0.33	1.00

discrete exmaple - $X \perp Y$

	X			
Y	-1	0	1	Marginal
-1	0.1089	0.1122	0.1089	0.33
0	0.1122	0.1156	0.1122	0.34
1	0.1089	0.1122	0.1089	0.33
Marginal	0.33	0.34	0.33	1.00

X and Y are indepedent (can be tediously verified.)

discrete exmaple - $X \perp Y$

		X		
Y	-1	0	1	Marginal
-1	0.1089	0.1122	0.1089	0.33
0	0.1122	0.1156	0.1122	0.34
1	0.1089	0.1122	0.1089	0.33
Marginal	0.33	0.34	0.33	1.00

X and Y are indepedent (can be tediously verified.)

Easy to show $E(XY) = 0$, so that $\rho = 0$.

discrete example - $\rho = 0$ but very much not independent!

	X			
Y	-1	0	1	Marginal
-1	0.00	0.00	0.25	0.25
0	0.50	0.00	0.00	0.50
1	0.00	0.00	0.25	0.25
Marginal	0.50	0.00	0.50	1.00

discrete example - $\rho = 0$ but very much not independent!

	X			
Y	-1	0	1	Marginal
-1	0.00	0.00	0.25	0.25
0	0.50	0.00	0.00	0.50
1	0.00	0.00	0.25	0.25
Marginal	0.50	0.00	0.50	1.00

X and Y are not independent.

discrete example - $\rho = 0$ but very much not independent!

		X		
Y	-1	0	1	Marginal
-1	0.00	0.00	0.25	0.25
0	0.50	0.00	0.00	0.50
1	0.00	0.00	0.25	0.25
Marginal	0.50	0.00	0.50	1.00

X and Y are not independent.

But $E(XY) = 0$, so $\rho = 0$.

standard deviation as an absolute property of a distribution

mean and SD aren't unique, but they do say something

$E(X)$ and $E(X^2)$ provide information about X that limit its values and probabilities to some extent. Two examples are *Markov's* and *Chebyshev's* inequalities.

mean and SD aren't unique, but they do say something

$E(X)$ and $E(X^2)$ provide information about X that limit its values and probabilities to some extent. Two examples are *Markov's* and *Chebyshev's* inequalities.

Theorem (Markov): If $X \geqslant 0$ has expected value $E(X)$, then:

$$P(X \geqslant t) \leqslant \frac{E(X)}{t}.$$

mean and SD aren't unique, but they do say something

$E(X)$ and $E(X^2)$ provide information about X that limit its values and probabilities to some extent. Two examples are *Markov's* and *Chebyshev's* inequalities.

Theorem (Markov): If $X \geqslant 0$ has expected value $E(X)$, then:

$$P(X \geqslant t) \leqslant \frac{E(X)}{t}.$$

Theorem (Chebyshev): If $\mathrm{Var}(X) = \sigma^2$ and $E(X) = \mu$:

$$P(|X - \mu| \geqslant t) \leqslant \frac{\sigma^2}{t^2}$$

Equivalently:

$$P(\mu - k\sigma < X < \mu + k\sigma) \geqslant 1 - \frac{1}{k^2}$$

example - "uniform on (0,1)"

Suppose X has the density $f(x) = 1$ on $0 < x < 1$ and 0 otherwise.

Then $E(X) = \frac{1}{2}$ and $\text{Var}(X) = \frac{1}{12}$

Various applications of Markov's and Chebyshev's inequality for this example show how weak they really are — mainly useful in theory than in practice.