Covariance and Correlation
Class 7, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Understand the meaning of covariance and correlation.

2. Be able to compute the covariance and correlation of two random variables.

2 Covariance

Covariance is a measure of how much two random variables vary together. For example, height and weight of giraffes have positive covariance because when one is big the other tends also to be big.

Definition: Suppose X and Y are random variables with means μ_X and μ_Y. The *covariance* of X and Y is defined as

$$\text{Cov}(X,Y) = E((X - \mu_X)(Y - \mu_Y)).$$

2.1 Properties of covariance

1. $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X,Y)$ for constants a, b, c, d.

2. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.

3. $\text{Cov}(X,X) = \text{Var}(X)$

4. $\text{Cov}(X,Y) = E(XY) - \mu_X\mu_Y$.

5. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$ for any X and Y.

6. If X and Y are independent then $\text{Cov}(X,Y) = 0$.
   Warning: The converse is false. Zero covariance does not always imply independence.

Note that by Property 5, the formula in Property 6 reduces to the earlier formula $\text{Var}(X + Y) = \text{Var}(X) = \text{Var}(Y)$ when X and Y are independent.

We give the proofs below. However, understanding and using these properties is more important than memorizing their proofs.

2.2 Sums and integrals for computing covariance

Since covariance is defined as an expected value we compute it in the usual way as a sum or integral.

Discrete case: If X and Y have joint pmf $p(x_i, y_j)$ then

$$\mathrm{Cov}(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)(x_i - \mu_X)(y_j - \mu_Y) = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)x_i y_j\right) - \mu_X\mu_Y.$$

Continuous case: If X and Y have joint pdf $f(x,y)$ over range $[a,b]\times[c,d]$ then

$$\mathrm{Cov}(X,Y) = \int_c^d \int_a^b (x-\mu_x)(y-\mu_y)f(x,y)\,dx\,dy = \left(\int_c^d \int_a^b xyf(x,y)\,dx\,dy\right) - \mu_x\mu_y.$$

2.3 Examples

Example 1. Flip a fair coin 3 times. Let X be the number of heads in the first 2 flips and let Y be the number of heads on the last 2 flips (so there is overlap on the middle flip). Compute $\mathrm{Cov}(X,Y)$.

answer: We'll do this twice, first using the joint probability table so you can see how that works, and then using the properties of covariance.

With 3 tosses there are only 8 outcomes {HHH, HHT,...}, so we can create the joint probability table directly.

$X\backslash Y$	0	1	2	$p(x_i)$
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
$p(y_j)$	1/4	1/2	1/4	1

From the marginals we compute $E(X) = 1 = E(Y)$. From the full table we compute

$$E(XY) = 1\cdot\frac{2}{8} + 2\frac{1}{8} + 2\frac{1}{8} + 4\frac{1}{8} = \frac{5}{4}.$$

So $\mathrm{Cov}(X,Y) = \frac{5}{4} - 1 = \boxed{\frac{1}{4}}$.

Next we compute $\mathrm{Cov}(X,Y)$ using the properties of covariance. As usual, let X_i be the result of the i^{th} flip, so $X_i \sim \text{Bernoulli}(.5)$. We have

$$X = X_1 + X_2 \quad \text{and} \quad Y = X_2 + X_3.$$

We know $E(X_i) = 1/2$ and $\mathrm{Var}(X_i) = 1/4$. Therefore $\mu_X = 1 = \mu_Y$. Using Property 2 of covariance, we have

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X_1+X_2, X_2+X_3) = \mathrm{Cov}(X_1,X_2)+\mathrm{Cov}(X_1,X_3)+\mathrm{Cov}(X_2,X_2)+\mathrm{Cov}(X_2,X_3).$$

Since the different tosses are independent we know

$$\mathrm{Cov}(X_1,X_2) = \mathrm{Cov}(X_1,X_3) = \mathrm{Cov}(X_2,X_3) = 0.$$

Looking at the expression for $\text{Cov}(X, Y)$ there is only one non-zero term

$$\text{Cov}(X, Y) = \text{Cov}(X_2, X_2) = \text{Var}(X_2) = \boxed{\frac{1}{4}}.$$

Example 2. (Zero covariance does not imply independence.) Let X be a random variable that takes values $-2, -1, 0, 1, 2$; each with probability $1/5$. Let $Y = X^2$. Show that $\text{Cov}(X, Y) = 0$ but X and Y are not independent.

<u>**answer:**</u> We make a joint probability table:

$Y \backslash X$	-2	-1	0	1	2	$p(y_j)$
0	0	0	1/5	0	0	1/5
1	0	1/5	0	1/5	0	2/5
4	1/5	0	0	0	1/5	1/5
$p(x_i)$	1/5	1/5	1/5	1/5	1/5	1

Using the marginals we compute means $E(X) = 0$ and $E(Y) = 2$.

Next we show that X and Y are not independent by finding one place where $p(x_i, y_j) \neq p(x_i)p(x_j)$:

$$P(X = -2,\ Y = 0) = 0 \neq 1/25 = P(X = -2) \cdot P(Y = 0).$$

Finally we compute covariance:

$$\text{Cov}(X, Y) = \frac{1}{5}(-8 - 1 + 1 + 8) - \mu_X \mu_y = 0.$$

Discussion: This example shows that $\text{Cov}(X, Y) = 0$ does not imply that X and Y are independent. In fact, X and X^2 are as dependent as random variables can be: if you know the value of X then you know the value of X^2 with 100% certainty.

The key point is that $\text{Cov}(X, Y)$ measures the *linear relationship* between X and Y. In the above example X and X^2 have a quadratic relationship that is completely missed by $\text{Cov}(X, Y)$.

2.4 Proofs of the properties of covariance

1 and 2 follow from similar properties for expected value.

3. This is the definition of variance:

$$\text{Cov}(X, X) = E((X - \mu_X)(X - \mu_X)) = E((X - \mu_X)^2) = \text{Var}(X).$$

4. Recall that $E(X - \mu_x) = 0$. So

$$\begin{aligned}
\text{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\
&= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\
&= E(XY) - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\
&= E(XY) - \mu_X \mu_Y.
\end{aligned}$$

5. Using properties 3 and 2 we get

$\text{Var}(X+Y) = \text{Cov}(X+Y, X+Y) = \text{Cov}(X,X)+2\text{Cov}(X,Y)+\text{Cov}(Y,Y) = \text{Var}(X)+\text{Var}(Y)+2\text{Cov}(X,Y).$

6. If X and Y are independent then $f(x,y) = f_X(x)f_Y(y)$. Therefore

$$\begin{aligned}
\text{Cov}(X,Y) &= \int\int (x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y)\,dx\,dy \\
&= \int (x-\mu_X)f_X(x)\,dx \int (y-\mu_Y)f_Y(y)\,dy \\
&= E(X-\mu_X)E(Y-\mu_Y) \\
&= 0.
\end{aligned}$$

3 Correlation

The units of covariance $\text{Cov}(X,Y)$ are 'units of X times units of Y'. This makes it hard to compare covariances: if we change scales then the covariance changes as well. Correlation is a way to remove the scale from the covariance.

Definition: The *correlation coefficient* between X and Y is defined by

$$\text{Cor}(X,Y) = \rho = \frac{\text{Cov}(X,Y)}{\sigma_X\,\sigma_Y}.$$

3.1 Properties of correlation

1. ρ is the covariance of the standardizations of X and Y.

2. ρ is dimensionless (it's a ratio).

3. $-1 \le \rho \le 1$. Furthermore,
   $\rho = +1$ if and only if $Y = aX + b$ with $a > 0$,
   $\rho = -1$ if and only if $Y = aX + b$ with $a < 0$.

Property 3 shows that ρ measures the *linear* relationship between variables. If the correlation is positive then when X is large, Y will tend to large as well. If the correlation is negative then when X is large, Y will tend to be small.

Example 2 shows that correlation can completely miss higher order relationships.

3.2 Proof of Property 3 of correlation

(*This is for the mathematically interested.*)

$0 \le \text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \text{Var}\left(\frac{X}{\sigma_X}\right) + \text{Var}\left(\frac{Y}{\sigma_Y}\right) - 2\text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) = 2 - 2\rho$

$\Rightarrow\ \rho \le 1$

Likewise $0 \le \text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \ \Rightarrow\ -1 \le \rho.$

If $\rho = 1$ then $0 = \text{Var}\left(\dfrac{X}{\sigma_X} - \dfrac{Y}{\sigma_Y}\right) \Rightarrow \dfrac{X}{\sigma_X} - \dfrac{Y}{\sigma_Y} = c.$ ■

Example. We continue Example 1. To compute the correlation we divide the covariance by the standard deviations. In Example 1 we found $\text{Cov}(X,Y) = 1/4$ and $\text{Var}(X) = 2\text{Var}(X_j) = 1/2$. So, $\sigma_X = 1/\sqrt{2}$. Likewise $\sigma_Y = 1/\sqrt{2}$. Thus

$$\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{1/4}{1/2} = \frac{1}{2}.$$

We see a positive correlation, which means that larger X tend to go with larger Y and smaller X with smaller Y. In Example 1 this happens because toss 2 is included in both X and Y, so it contributes to the size of both.

3.3 Bivariate normal distributions

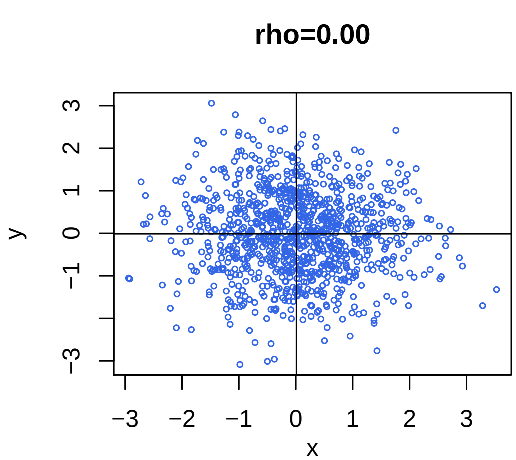
The *bivariate normal distribution* has density

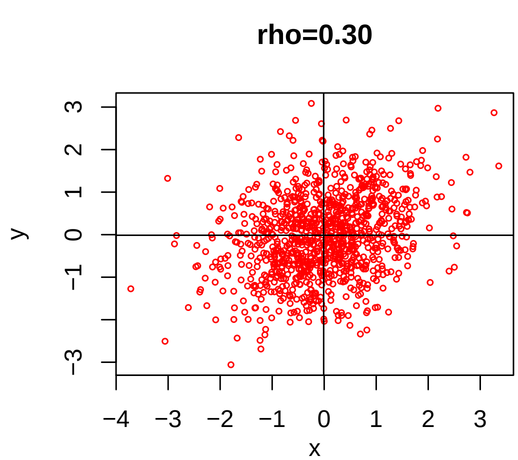
$$f(x,y) = \frac{e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

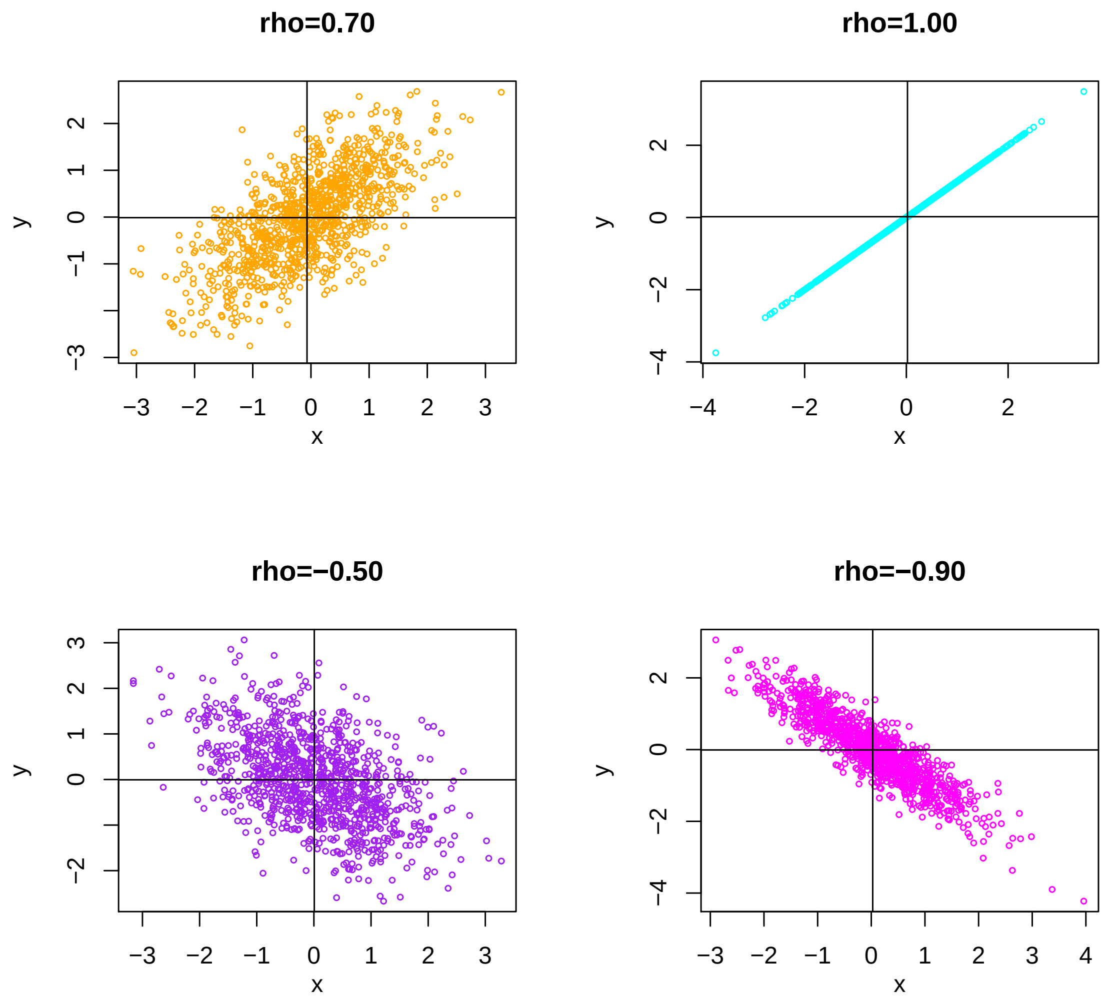
For this distribution, the marginal distributions for X and Y are normal and the correlation between X and Y is ρ.

In the figures below we used R to simulate the distribution for various values of ρ. Individually X and Y are standard normal, i.e. $\mu_X = \mu_Y = 0$ and $\sigma_X = \sigma_Y = 1$. The figures show scatter plots of the results.

These plots and the next set show an important feature of correlation. We divide the data into quadrants by drawing a horizontal and a verticle line at the means of the y data and x data respectively. A positive correlation corresponds to the data tending to lie in the 1st and 3rd quadrants. A negative correlation corresponds to data tending to lie in the 2nd and 4th quadrants. You can see the data gathering about a line as ρ becomes closer to ± 1.



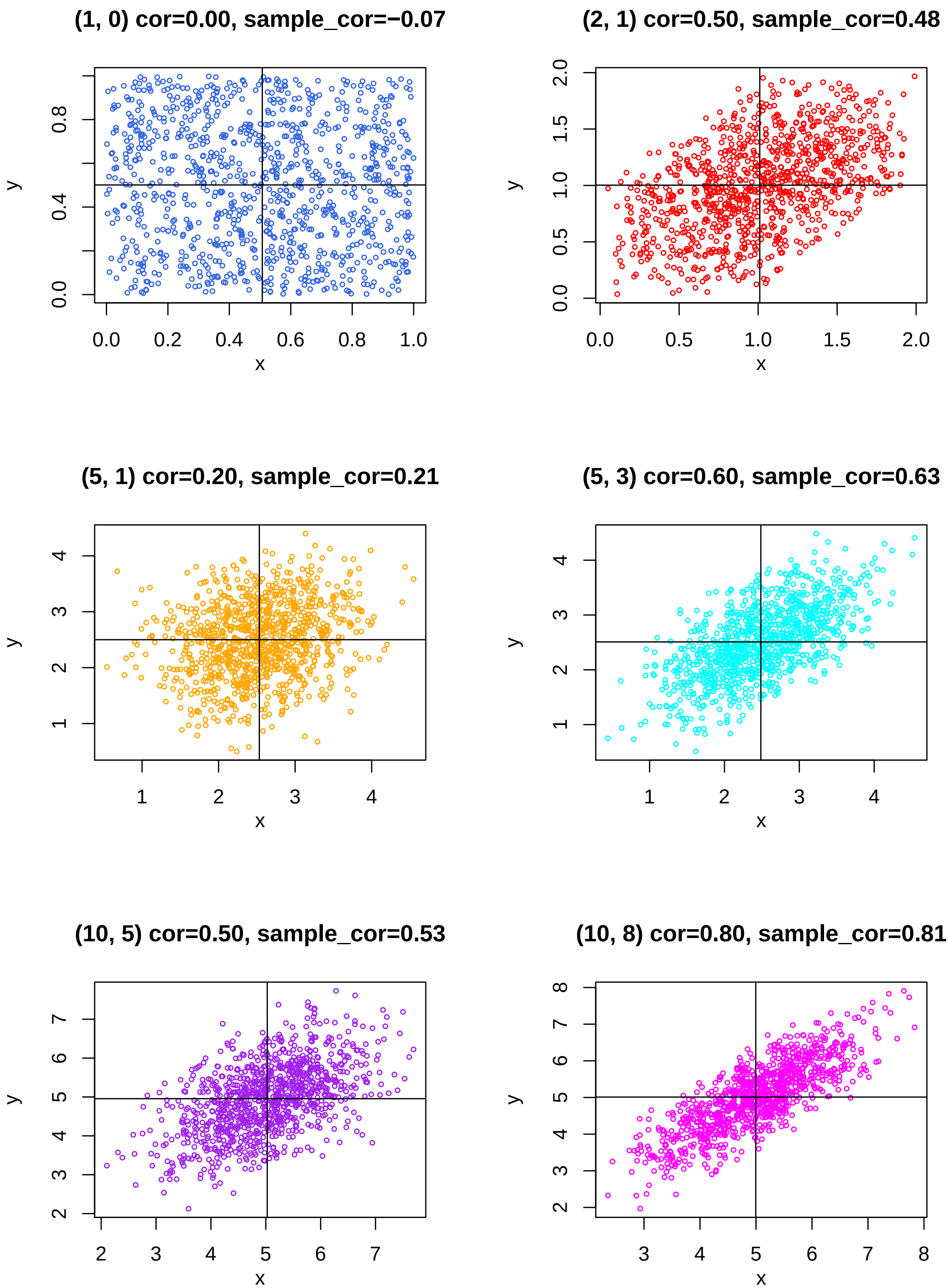




3.4 Overlapping uniform distributions

We ran simulations in R of the following scenario. $X_1, X_2, \ldots, X_{20}$ are i.i.d and follow a $U(0,1)$ distribution. X and Y are both sums of the same number of X_i. We call the number of X_i common to both X and Y the overlap. The notation in the figures below indicates the number of X_i being summed and the number which overlap.

For example, 5,3 indicates that X and Y were each the sum of 5 of the X_i and that 3 of the X_i were common to both sums. (The data was generated using `rand(1,1000);`)


(1, 0) cor=0.00, sample_cor=−0.07
(2, 1) cor=0.50, sample_cor=0.48
(5, 1) cor=0.20, sample_cor=0.21
(5, 3) cor=0.60, sample_cor=0.63
(10, 5) cor=0.50, sample_cor=0.53
(10, 8) cor=0.80, sample_cor=0.81
x
y