Linear Combinations, Covariance and Correlation

Rob McCulloch

1. Mean and Variance of a Linear Function
2. Covariance and Correlation for RVs
3. Mean and Variance of a Linear Combination

1. Mean and Variance of a Linear Function

Previously we looked at

$$Y = c_0 + c_1 X$$

where Y and X were discrete.

Our formulas for the mean and variance work the same way for discrete and continuous random variables.

Let **Y** and **X** be random variables such that

$$Y = c_0 + c_1 X$$

Then,

$$E(Y) = c_0 + c_1 E(X)$$

$$Var(Y) = c_1^2 Var(X)$$

$$\sigma_Y = |c_1| \sigma_X$$

Example:

Suppose you plan to play the game with winnings W with $E(W) = 0$ and $Var(W) = 100$.

There is a \$3 cover charge to get in to play so your total winnings T is represented by:

$$T = -3 + W$$

Then,,

$E(T) = -3 + E(W) = -3.$

$Var(T) = Var(W) = 100.$

Suppose sneak in, don't pay the cover and "double your money":

$$T = 2W$$

$E(T) = 2E(W) = 2(0) = 0.$

$Var(T) = 4Var(W) = 400.$

We don't have to know the distribution of W.
We don't have to know if W is discrete or continuous.

Example:

Suppose you think returns on the market next period can be represented by

$$M \sim N(6, 225),$$

where M represents the uncertain market return (in percent).

You also could invest in a "risk-free" asset which pays r_f for sure.

Suppose you put fraction w_1 into the risk free asset and w_2 into the market. ($w_1 + w_2 = 1$, $w_i \geq 0$).

Then the return on your portfolio is:

$$P = w_1 \, r_f + w_2 \, M$$

Suppose $r_f = 2\%$ and $w_2 = .6$.

Then,

$$P = .4(2) + .6\,M$$
$$= .8 + .6\,M$$

What are the mean and standard deviation of P?

$$E(P) = .8 + .6\,E(M) = .8 + .6(6) = 4.4.$$

$$\sigma_P = .6\,\sigma_M = .6\,(15) = 9.$$

2. Covariance and Correlation for RVs

Suppose we have a pair of random variables (X, Y).

We would like to have a measure of how dependent the random variables are.

The covariance between bivariate discrete random variables X and Y is

$$cov(X, Y) = \sigma_{XY} = \sum p(x, y)(x - \mu_x)(y - \mu_y)$$

where the sum is over all possible (x, y) pairs.

For continuous random variables there is a formula using calculus.

Example:

$\mu_X = .1, \quad \mu_Y = .1.$

$\sigma_X = .05, \; \sigma_Y = .05.$

		X	
		.05	.15
Y	.05	.4	.1
	.15	.1	.4

x	y	prob	x-E(X)	y-E(Y)	prod
0.05	0.05	0.4	-0.05	-0.05	0.0025
0.15	0.05	0.1	0.05	-0.05	-0.0025
0.05	0.15	0.1	-0.05	0.05	-0.0025
0.15	0.15	0.4	0.05	0.05	0.0025

$\text{Cov}(X, Y) = \sigma_{XY}$

$= .4 * .0025 + .1 * (-.0025) + .1 * (-.0025) + .4 * .0025 = .0015.$

Intuition:

There is an 80% chance X and Y move *in the same direction*.

Example:

$\mu_X = .1, \quad \mu_Y = .1.$

$\sigma_X = .05, \;\; \sigma_Y = .05.$

		X	
		.05	.15
Y	.05	.1	.4
	.15	.4	.1

x	y	prob	x-E(X)	y-E(Y)	prod
0.05	0.05	0.1	-0.05	-0.05	0.0025
0.15	0.05	0.4	0.05	-0.05	-0.0025
0.05	0.15	0.4	-0.05	0.05	-0.0025
0.15	0.15	0.1	0.05	0.05	0.0025

$\text{Cov}(X, Y) = \sigma_{XY}$

$= .1 * .0025 + .4 * (-.0025) + .4 * (-.0025) + .1 * .0025 = -.0015.$

Intuition:

There is an 80% chance X and Y move *in opposite directions*.

The **correlation** between random variables (discrete or continuous) is

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

ρ: the basic facts

$$-1 \leq \rho \leq 1$$

If ρ is close to 1, then it means there is a line, with positive slope, such that (X,Y) is likely to fall close to it.

If ρ is close to -1, same thing, but the line has a negative slope.

Example:

		X	
		.05	.15
Y	.05	.4	.1
	.15	.1	.4

The correlation is:

$$\rho_{XY} = \frac{.0015}{.05 * .05} = 0.6.$$

Example:

		X	
		.05	.15
Y	.05	.1	.4
	.15	.4	.1

The correlation is:

$$\rho_{XY} = \frac{-.0015}{.05 * .05} = -0.6.$$

Example

		X	
		0	1
Y	0	.25	.25
	1	.25	.25

Let us compute the covariance:
.25(-.5)(-.5) + .25(-.5)(.5) + .25(.5)(-.5) +.25(.5)(.5)=0

The covariance is 0 and so is the correlation: not surprising, right?

Independence and correlation

Suppose two RV's are independent.

That means they have **nothing to do with each other**.

That means they have **nothing to do with each other linearly**

That means the correlation is 0.

If X and Y are independent, then

$$\rho_{XY} = 0 = \mathrm{cov}(X, Y)$$

Note: the converse is not necessarily true. Cov=0 **does not** necessarily mean they are independent.

3. Mean and Variance of a Linear Combination

We looked at the returns on a portfolio which mixed a riskless asset and a risky asset.

How about mixing two risky assets?

Let X be the return on one risky asset and Y be the return on another.

Suppose you put fraction w_1 of your wealth into X and w_2 into Y.

Let P denote the return on your portfolio.

Then

$$P = w_1 X + w_2 Y$$

The random variable P is a linear combination of the random variables X and Y.

Example:

Suppose X and Y have the joint distribution given below and $w_1 = w_2 = .5$.

$$P = .5\,X + .5\,Y$$

$\mu_X = .1, \quad \mu_Y = .1.$

$\sigma_X = .05, \;\; \sigma_Y = .05.$

		X	
		.05	.15
Y	.05	.1	.4
	.15	.4	.1

What is the distribution of P?

If X turns out to be .15 and Y turns out to be .05 the P turns out to be $.5 * .15 + .5 * .05 = .1$

```
     x    y prob    p
1 0.05 0.05  0.1 0.05
2 0.15 0.05  0.4 0.10
3 0.05 0.15  0.4 0.10
4 0.15 0.15  0.1 0.15
```

$P = .5\,X + .5\,Y$

P:

	prob	p
1	0.1	0.05
2	0.4	0.10
3	0.4	0.10
4	0.1	0.15

or P:

	prob	p
1	0.1	0.05
2	0.8	0.10
3	0.1	0.15

What are the mean and variance of P?

$\mu_P = \mathrm{E}(P) = .1 * .05 + .4 * .1 + .4 * .1 + .1 * .15 = 0.1$

$\sigma_P^2 = \mathrm{Var}(P) = .1 * (.05 - .1)^2 + .8 * (.1 - .1)^2 + .1 * (.15 - .1)^2$
$= .0005$

$\sigma_P = \sqrt{.0005} = .0224.$

The mean makes sense

but why is the standard deviation so low ??!!

$P = .5\,X + .5\,Y$

	x	y	prob	p
1	0.05	0.05	0.1	0.05
2	0.15	0.05	0.4	0.10
3	0.05	0.15	0.4	0.10
4	0.15	0.15	0.1	0.15

P has a way lower variance because 80% of the time if one of X and Y are above average, the other is below and the two cancel out in P which is just the average.

80% of the time, P is equal to $\mu_P = .1$.

Apparently, covariance matters when we linearly combine RVs !!

Mean and Variance of a Combination of two RVs

Let Y, X_1, and X_2 be random variables such that

$$Y = c_0 + c_1 X_1 + c_2 X_2.$$

then,

$$\mu_y = c_0 + c_1 \mu_{X_1} + c_2 \mu_{X_2}.$$

$$\sigma_Y^2 = c_1^2 \sigma_{X_1}^2 + c_2^2 \sigma_{X_2}^2 + 2 c_1 c_2 \sigma_{X_1 X_2}.$$

Example:

In the example we just did we have

$\mu_X = .1, \quad \mu_Y = .1.$

$\sigma_X = .05, \;\; \sigma_Y = .05.$

$\sigma_{XY} = -.0015.$

$$P = .5\,X + .5\,Y$$

$\mu_P = .5 * .1 + .5 * .1 = .1$

$\sigma_P^2 = .5^2 * .05^2 + .5^2 * .05^2 + 2 * .5 * .5 * (-.0015) = .0005.$

$\sqrt{.0005} = .0224.$

Example:

Let's use the formulas on our positive covariance example.

$\mu_X = .1, \quad \mu_Y = .1. \ \sigma_X = .05, \ \sigma_Y = .05.$
$\sigma_{XY} = .0015.$

$$P = .5\,X + .5\,Y$$

$\mu_P = .5 * .1 + .5 * .1 = .1$

$\sigma_P^2 = .5^2 * .05^2 + .5^2 * .05^2 + 2 * .5 * .5 * (.0015) = .002$

$\sqrt{.002} = .0447.$

With a positive covariance, X and Y tend to go up and down together, so you don't get much variance reduction by averaging.

In general, with a combination of k random variables we have:

Let Y, X_1, X_2, X_3, ...X_k (have k X's) be random variables such that

$$Y = c_0 + c_1 X_1 + c_2 X_2 + \ldots + c_k X_k$$

then,

$$\mu_y = c_0 + c_1 \mu_{X_1} + c_2 \mu_{X_2} + \ldots + c_k \mu_{X_k}.$$

$$\begin{aligned} \sigma_Y^2 &= c_1^2 \sigma_{X_1}^2 + c_2^2 \sigma_{X_2}^2 + \ldots + c_k^2 \sigma_{X_k}^2 \\ &+ 2 \sum_{\text{all pairs}} c_i c_j \sigma_{X_i X_j} \end{aligned}$$

Example:

With $k = 3$ we have:

$$Y = c_0 + c_1 X_1 + c_2 X_2 + c_3 X_3$$

$$\mu_y = c_0 + c_1 \mu_{X_1} + c_2 \mu_{X_2} + c_3 \mu_{X_3}.$$

$$\begin{aligned} \sigma_Y^2 &= c_1^2 \sigma_{X_1}^2 + c_2^2 \sigma_{X_2}^2 + c_3^2 \sigma_{X_3}^2 \\ &+ 2 \left(c_1 c_2 \sigma_{X_1 X_2} + c_2 c_3 \sigma_{X_2 X_3} + c_1 c_3 \sigma_{X_1 X_3} \right) \end{aligned}$$

Note:

The return on a portfolio in general.

Suppose we put wealth fraction w_i into asset with return r_i, $i = 1, 2, \ldots, k$.

Then, the return on the portfolio is

$$P = w_1 \, r_1 + w_2 \, r_2 + \ldots + w_k \, r_k = \sum w_i r_i.$$

Example

$$Y = .2X_1 + .5X_2 + .3X_3$$

$\mu_{X_1} = .05$

$\mu_{X_2} = .1$

$\mu_{X_3} = .15$

$\sigma^2_{X_1} = .01$

$\sigma^2_{X_2} = .009$

$\sigma^2_{X_3} = .008$

$\rho_{X_1X_2} = .3$

$\rho_{X_1X_3} = -.2$

$\rho_{X_2X_3} = .2$

$\sigma_{X_1X_2} = .002846$

$\sigma_{X_1X_3} = -.001789$

$\sigma_{X_2X_3} = .001697$

μ_y = .2*(.05) + .5*(.1) + .3*(.15)= 0.105000

$\sigma^2_Y =$ .2*.2*(.01) + .5*.5*(.009) + .3*.3*(.008) + 2*.2*.5*.002846 + 2*.2*.3*(-.001789) + 2*.5*.3*.001697 = 0.00423362

Think in terms of financial assets and portfolios.

Special Cases

Here are some important special cases of our formulae
(remember, these work for any RV's, continuous or discrete)

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2\,cov(X_1, X_2)$$

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2\,cov(X_1, X_2)$$

If the correlation is 0, then:

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2)$$

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2)$$

For example:

$$Y = X_1 + X_2$$
$$= 0 + (1)\,(X_1) + (1)\,(X_2)$$

$$E(Y) = 0 + (1)\,E(X_1) + (1)\,E(X_2) =$$
$$E(X_1) + E(X_2).$$

$$Var(Y) = (1^2)\,Var(X_1) + (1^2)\,Var(X_2) + 2(1)(1)\,cov(X_1, X_2) =$$
$$Var(X_1) + Var(X_2) + 2\,cov(X_1, X_2).$$

For example:

$$Y = X_1 - X_2$$
$$= 0 + (1)\,(X_1) + (-1)\,(X_2)$$

$$E(Y) = 0 + (1)\,E(X_1) + (-1)\,E(X_2) =$$
$$E(X_1) - E(X_2).$$

$$Var(Y) = (1^2)\,Var(X_1) + (-1)^2\,Var(X_2) + 2(1)(-1)\,cov(X_1, X_2) =$$
$$Var(X_1) + Var(X_2) - 2\,cov(X_1, X_2).$$

Example:

Suppose the wining on a game are represented by the random variable W with

$$E(W) = -1, \qquad Var(W) = 100$$

You pay a dollar to play and after that you are just as likely to lose as win.

Suppose you double the bet so that

$$T_1 = 2W.$$

$E(T_1) =$

$Var(T_1) =$

Suppose you double the bet so that your total winnings are

$$T_1 = 2W.$$

$E(T_1) = 2E(W) = 2(-1) = -2.$

$Var(T_1) = 2^2 Var(W) = 4 Var(W) = 400.$

Suppose you play twice so that your total winnings are

$$T_2 = W_1 + W_2$$

$E(T_2) =$

$Var(T_2) =$

Suppose you play twice so that your total winnings are

$$T_2 = W_1 + W_2$$

$$E(T_2) = E(W_1) + E(W_2) = 2E(W) = 2(-1) = -2.$$

$$Var(T_2) = Var(W_1) + Var(W_2) = 2Var(W) = 200.$$

Suppose you play the game twice and are interested in the *average* winnings

$$A = \frac{W_1 + W_2}{2} = \frac{1}{2}W_1 + \frac{1}{2}W_2$$

$E(A) =$

$Var(A) =$

Suppose you play the game twice and are interested in the *average* winnings

$$A = \frac{W_1 + W_2}{2} = \frac{1}{2}W_1 + \frac{1}{2}W_2$$

$E(A) = \frac{1}{2}E(W_1) + \frac{1}{2}E(W_2) = \frac{2}{2}E(W) = E(W) = -1.$

$Var(A) = \frac{1}{2^2}Var(W_1) + \frac{1}{2^2}Var(W_2) = \frac{2}{4}Var(W) = \frac{Var(W)}{2} = 50.$

Double:

- mean doubles.
- variance goes up by 4.

Sum two:

- mean doubles.
- variance goes up by 2.

Average two

- mean is the same.
- variance goes down by $\frac{1}{2}$.