MICHIGAN STATE UNIVERSITY

AFRE 802 Statistical Methods for Agricultural, Food, & Resource Economists

Multivariate probability distributions – Part 3 of 3
(WMS Ch. 5.5-5.8, 5.11, 5.12)

October 12, 2017

Nicole Mason
Michigan State University
Fall 2017

MICHIGAN STATE UNIVERSITY

GAME PLAN

- Review
- Graded in-class exercise

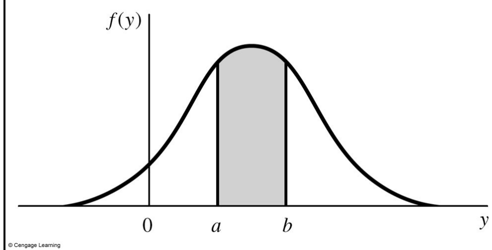
Multivariate probability distributions (Part 3 of 3)

1. Finish coverage of conditional probability distributions
2. Independent random variables
3. Expected values of (general) functions of RVs
4. Covariances and correlation coefficients
5. Expected values, variances, covariances, and correlations of linear functions of RVs
6. Conditional expectations

1

Graphical representation of univariate vs. bivariate PDFs

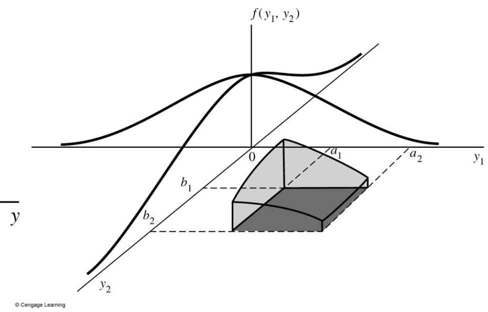
Univariate



Recall that in the univariate case, area under the PDF between a and b = P(a ≤ Y ≤ b)

$$\boxed{P(a \le Y \le b) = \int_a^b f(y)\,dy}$$

Bivariate



Bivariate PDFs are 2-dimensional surfaces, so the volume under the surface corresponds to a probability – e.g., above:

$$\boxed{P(a_1 \le Y_1 \le a_2, b_1 \le Y_2 \le b_2) = \int_{b_1}^{b_2}\int_{a_1}^{a_2} f(y_1,y_2)\,dy_1\,dy_2}$$

2

Review

Marginal probability distributions

Discrete RVs:

$$p_1(y_1) = \sum_{\text{all } y_2} p(y_1,y_2) \text{ and } p_2(y_2) = \sum_{\text{all } y_1} p(y_1,y_2)$$

Table 5.2 Joint probability function for Y_1 and Y_2, Example 5.5

	y_2	y_1 = 0	1	2	$p_2(y_2)$: Total
	0	0	3/15	3/15	6/15
	1	2/15	6/15	0	8/15
	2	1/15	0	0	1/15
$p_1(y_1)$:	*Total*	3/15	9/15	3/15	1

Continuous RVs:

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1,y_2)\,dy_2 \text{ and } f_2(y_2) = \int_{-\infty}^{\infty} f(y_1,y_2)\,dy_1$$

Review

MICHIGAN STATE UNIVERSITY

Conditional probability distributions

Recall that $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

Discrete RVs:

$$p(y_1 \mid y_2) = \frac{p(y_1, y_2)}{p_2(y_2)} \Leftrightarrow P(Y_1 = y_1 \mid Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)}$$

Continuous RVs:

$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)} \text{ and } f(y_2 \mid y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

Conditional CDF:

$F(y_1 \mid y_2)$ means $P(Y_1 \leq y_1 \mid Y_2 = y_2)$

Compute by integrating the conditional PDF over the relevant range.

Review

MICHIGAN STATE UNIVERSITY

EX) Conditional probability distributions for discrete RVs

Continuing with our previous example about a 2-person committee drawn from Democrats, Republicans, and independents (with bivariate probability distribution below), find the conditional probability distribution of Y_1 given that Y_2=1.

$$p(y_1 \mid y_2) = \frac{p(y_1, y_2)}{p_2(y_2)}$$

Table 5.2 Joint probability function for Y_1 and Y_2, Example 5.5

	y_1			$p_2(y_2)$:
y_2	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
$p_1(y_1)$: *Total*	3/15	9/15	3/15	1

Answer:

$$P(Y_1 = 0 \mid Y_2 = 1) = \frac{p(0,1)}{p_2(1)} = \frac{2/15}{8/15} = \frac{1}{4}$$

$$P(Y_1 = 1 \mid Y_2 = 1) = \frac{p(1,1)}{p_2(1)} = \frac{6/15}{8/15} = \frac{3}{4}$$

$$P(Y_1 = 2 \mid Y_2 = 1) = \frac{p(2,1)}{p_2(1)} = \frac{0/15}{8/15} = 0$$

Graded in-class exercise

EX) Conditional probability distributions for continuous RVs

Find the conditional probability, $P(Y_1 \leq 0.5 \mid Y_2 = 1.5)$, for the continuous bivariate PDF:

$$f(y_1, y_2) = \begin{cases} 0.5, & 0 \leq y_1 \leq y_2 \leq 2 \\ 0, & \text{elsewhere} \end{cases}$$

Conditional PDF :

$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

Independent random variables

*What is $P(A \cap B)$ if events A and B are **independent**?*

$$\boxed{P(A \cap B) = P(A)P(B)}$$

Can use similar approach to see if two RVs, Y_1 and Y_2, are independent:

CDF : $F(y_1, y_2) = F_1(y_1)F_2(y_2)$

Discrete probability distribution : $p(y_1, y_2) = p_1(y_1)p_2(y_2)$

Continuous probability density function : $f(y_1, y_2) = f_1(y_1)f_2(y_2)$

8

MICHIGAN STATE UNIVERSITY

EX) Checking for independence – discrete RVs

Continuing with our previous example about a 2-person committee drawn from Democrats, Republicans, and independents (with bivariate probability distribution below), is Y_1 independent of Y_2? Can check any point, but let's try (0,0).

Table 5.2 Joint probability function for Y_1 and Y_2, Example 5.5

	y_1			$p_2(y_2)$:
y_2	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
$p_1(y_1)$: *Total*	3/15	9/15	3/15	1

$\boxed{\textbf{Discrete probability distribution if independent: } p(y_1, y_2) = p_1(y_1)p_2(y_2)}$

9

EX) Checking for independence – continuous RVs

Suppose Y_1 and Y_2 have the continuous bivariate PDF below. Are these two RVs independent?

$$f(y_1, y_2) = \begin{cases} 6y_1 y_2^2, & 0 \le y_1 \le 1,\ 0 \le y_2 \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Continuous probability density function if indepedent : $f(y_1, y_2) = f_1(y_1) f_2(y_2)$

10

Another way to check for independence

Y_1 and Y_2 are independent RVs if:

$$f(y_1, y_2) = g(y_1)h(y_2)$$

where $g(y_1)$ is a nonnegative function of y_1 alone and $h(y_2)$ is a nonnegative function of y_2 alone ***AND only if
$f(y_1, y_2) > 0$ for $a \le y_1 \le b$ and $c \le y_2 \le d$
for constants a, b, c, d ***.

EX) Suppose Y_1 and Y_2 have the continuous bivariate PDF below. Are these two RVs independent?

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1,\ 0 \le y_2 \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

11

MICHIGAN STATE UNIVERSITY

Another example

$$f(y_1,y_2)=\begin{cases} 8y_1y_2, & 0\le y_1\le 1,\ 0\le y_2\le y_1 \\ 0, & \text{elsewhere} \end{cases}$$

We cannot use the alternative method here because the range over which y_2 has positive probability is a function of y_1.

If we found the marginal distributions for y_1 and y_2, we would see that:

$$f_1(y_1)=4y_1^3$$

and

$$f_2(y_2)=4y_2(1-y_2^2)$$

So $f(y_1,y_2)\neq f_1(y_1)f_2(y_2)$, thus Y_1 and Y_2 not independent

12

The expected value of a function of RVs

Recall that in the ***univariate*** case,

Discrete RVs: $E\left[g(Y)\right]=\sum_i g(y_i)p(y_i)$

Continuous RVs: $E[g(Y)]=\int_{-\infty}^{\infty} g(y)f(y)\,dy$

For the ***bivariate*** case,

Discrete RVs:

$$E\left[g(Y_1,Y_2)\right]=\sum_{\text{all } y_2}\sum_{\text{all } y_1} g(y_1,y_2)p(y_1,y_2)$$

Continuous RVs:

$$E\left[g(Y_1,Y_2)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1,y_2)f(y_1,y_2)\,dy_1\,dy_2$$

13

MICHIGAN STATE UNIVERSITY

Expected value of a function of RVs – example

Let Y_1 and Y_2 have joint density given by

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $E(Y_1 Y_2)$.

$$\boxed{E\left[g(Y_1,Y_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1,y_2)f(y_1,y_2)\,dy_1\,dy_2}$$

14

MICHIGAN STATE UNIVERSITY

Expected value of a function of RVs – **Rules**

1. $E\left[cg(Y_1,Y_2)\right] = cE\left[g(Y_1,Y_2)\right]$ for any constant c

2. $E\left[g_1(Y_1,Y_2)+g_2(Y_1,Y_2)+\ldots+g_k(Y_1,Y_2)\right]$
$= E\left[g_1(Y_1,Y_2)\right]+E\left[g_2(Y_1,Y_2)\right]+\ldots+E\left[g_k(Y_1,Y_2)\right]$

3. $E\left(c_1Y_1+c_2Y_2+\ldots+c_kY_k\right) = c_1E\left(Y_1\right)+c_2E\left(Y_2\right)+\ldots+c_kE\left(Y_k\right)$

4. If Y_1 and Y_2 are independent then
$E\left[g(Y_1)h(Y_2)\right] = E\left[g(Y_1)\right]E\left[h(Y_2)\right]$

15

Expected value of a function of RVs – **Rules** – examples

a. What is $E(Y_1 - Y_2)$ if $E(Y_1) = 0.75$ and $E(Y_2) = 0.375$

b. What is $E(3Y_1 - 7Y_2)$?

c. What is $E(Y_1 Y_2)$ if Y_1 and Y_2 are independent?

16

Covariance & correlation coefficient of 2 RVs

Recall the formula for the variance of an RV:

$$V(Y) = E\left[(Y-\mu)^2\right] = E\left[(Y-\mu)(Y-\mu)\right] = E(Y^2) - \left[E(Y)\right]^2 = E(Y^2) - \mu^2$$

The formula for the covariance of Y_1 and Y_2 is similar:

$$Cov(Y_1,Y_2) = E\left[(Y_1-\mu_1)(Y_2-\mu_2)\right] = E(Y_1Y_2) - E(Y_1)E(Y_2) = E(Y_1Y_2) - \mu_1\mu_2$$

where $\mu_1 = E(Y_1)$ and $\mu_2 = E(Y_2)$.

It is difficult to interpret the magnitude of the covariance, but we can standardize it to get the correlation coefficient, ρ. Note that $-1 \le \rho \le 1$.

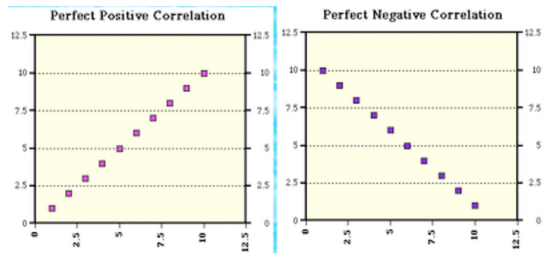
$$Corr(Y_1,Y_2) \equiv \rho = \frac{Cov(Y_1,Y_2)}{\sigma_1\sigma_2}$$

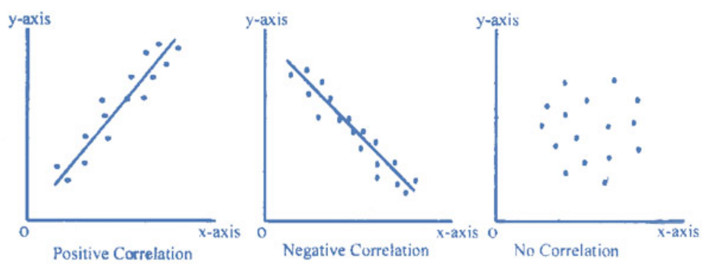
where σ_1 and σ_2 are the standard deviations of Y_1 and Y_2.

What is the relationship between the signs of Cov and Corr?

Covariance & correlation coefficient of 2 RVs

- *What does it mean to have a positive covariance (or correlation)? Negative? Zero?*
- *Correlation = +1? -1?*





Distribute handout

MICHIGAN STATE UNIVERSITY

Covariance, correlation, & independence

What is $E(Y_1 Y_2)$ if Y_1 and Y_2 are **independent**?

$$\boxed{E(Y_1 Y_2) = E(Y_1)E(Y_2) \text{ if } Y_1 \text{ and } Y_2 \text{ are } \textbf{independent}}$$

What is $Cov(Y_1 Y_2)$ if Y_1 and Y_2 are **independent**?

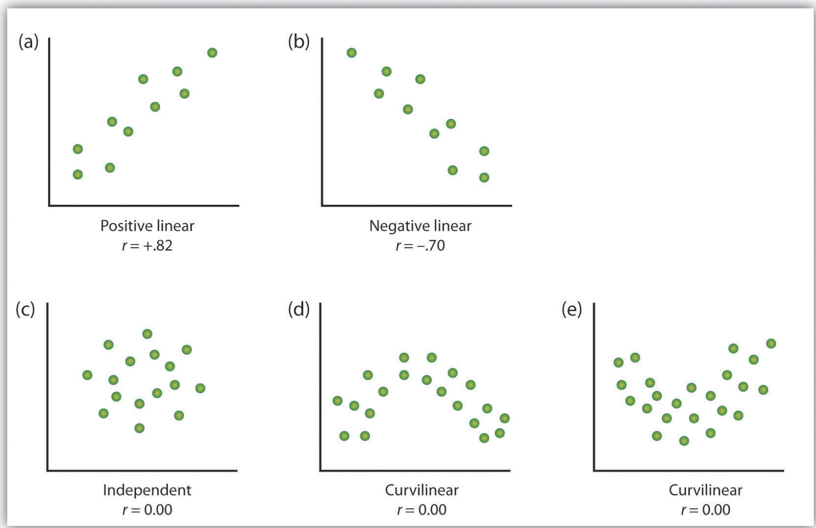
$$Cov(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = E(Y_1)E(Y_2) - E(Y_1)E(Y_2) = 0$$
if Y_1 and Y_2 are **independent**

What is $Corr(Y_1, Y_2)$ if Y_1 and Y_2 are **independent**?

$$\boxed{Corr(Y_1, Y_2) = \frac{Cov(Y_1, Y_2)}{\sigma_1 \sigma_2} = 0 \text{ if } Y_1 \text{ and } Y_2 \text{ are } \textbf{independent}}$$

***Note: **independence implies zero covariance (correlation)**
BUT zero covariance (correlation) does NOT imply independence.
Why?

Because **covariance** & **correlation** are about **linear** dependence and it is possible for two variables to have a **non-linear relationship but no linear relationship**



MICHIGAN STATE UNIVERSITY

Calculating the covariance – example #1

$$f(y_1, y_2) = \begin{cases} 2y_1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

For this joint PDF, $E(Y_1Y_2)$=1/3, $E(Y_1)$=2/3, and $E(Y_2)$=1/2.
Find Cov(Y_1, Y_2). Does the answer surprise you? Why or why not?

$$\boxed{Cov(Y_1,Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)}$$

21

Calculating the covariance – example #2

Show that Y_1 and Y_2 are dependent but have zero covariance.

	y_1		
y_2	-1	0	$+1$
-1	$1/16$	$3/16$	$1/16$
0	$3/16$	0	$3/16$
$+1$	$1/16$	$3/16$	$1/16$

$$\boxed{E\left[g(Y_1,Y_2)\right]=\sum_{\text{all } y_2}\sum_{\text{all } y_1} g(y_1,y_2)p(y_1,y_2)}$$

$$\boxed{Cov(Y_1,Y_2)=E(Y_1Y_2)-E(Y_1)E(Y_2)}$$

Rules for the expected value, variance, and covariance of linear functions of RVs: ***The bivariate case*** (see WMS pp. 271-273 for proof & multivariate case)

Random variables Y_1 and Y_2, and constants a_1, a_2, b_1 and b_2 :

$$\boxed{1.\ E(a_1Y_1+b_1+a_2Y_2+b_2)=a_1E(Y_1)+b_1+a_2E(Y_2)+b_2}$$

EX) $E(3Y_1-2-8Y_2+5)$

$$\boxed{2.\ V(a_1Y_1+b_1+a_2Y_2+b_2)=a_1^2V(Y_1)+a_2^2V(Y_2)+2a_1a_2Cov(Y_1,Y_2)}$$

EX) $V(Y_1+Y_2)$

EX) $V(Y_1-Y_2)$

EX) $V(3Y_1-2-8Y_2+5)$

$$\boxed{3.\ Cov(a_1Y_1+b_1, a_2Y_2+b_2)=a_1a_2Cov(Y_1,Y_2)}$$

EX) $Cov(Y_1,-Y_2)$

EX) $Cov(3Y_1-2,-8Y_2+5)$

MICHIGAN STATE UNIVERSITY

Conditional expectations

Motivation

- Covariance and correlation measure the linear relationship (linear dependence) between two RVs and treat them symmetrically
- In applied economics, we often want to explain one RV (Y) in terms of another RV (X)
- Call Y the "explained" variable, X the "explanatory" variable
- Recall conditional probability distributions and PDFs: $p(y|x)$ and $f(y|x)$
- We are often interested in the **conditional expectation** (a.k.a. the **conditional mean**): ***E(Y|X=x)*** or, for shorthand, ***E(Y|X)*** or sometimes ***E(Y|x)***

24

MICHIGAN STATE UNIVERSITY

Conditional expectations & variances - formulas

Conditional expectation of *Y* given *X*

Discrete RVs:

$$E(Y\,|\,X=x)=E(Y\,|\,X)=\sum_{\text{all } y} y\;p(y\,|\,x)$$

Continuous RVs:

$$E(Y\,|\,X=x)=E(Y\,|\,X)=\int_{-\infty}^{\infty} y\,f(y\,|\,x)\,dy$$

How would you use the E[g(Y)|X] formula to find the conditional variance, V(Y|X)?

$$V(Y\,|\,X)$$
$$=E(Y^2\,|\,X)-\left[E(Y\,|\,X)\right]^2$$

Conditional expectation of *g(Y)* given *X*

Discrete RVs:

$$E\left[g(Y)\,|\,X=x\right]=E\left[g(Y)\,|\,X\right]=\sum_{\text{all } y} g(y)\;p(y\,|\,x)$$

Continuous RVs:

$$E\left[g(Y)\,|\,X=x\right]=E\left[g(Y)\,|\,X\right]=\int_{-\infty}^{\infty} g(y)\,f(y\,|\,x)\,dy$$

25

Conditional expectations & variances - rules

<u>Gist</u>: treat the variable you are conditioning on as a constant

$\boxed{1.\ E\left[g(Y)\,|\,Y\right] = g(Y) \text{ for any function } g(.)}$

EX) $E(Y^2\,|\,Y)$

$\boxed{2.\ E\left[g(X)Y\,|\,X\right] = g(X)E(Y\,|\,X)}$

EX) $E(2X^2Y\,|\,X)$

$\boxed{\begin{array}{l} 3.\ \text{If } X \text{ and } Y \text{ are independent,} \\ \text{then } E(Y\,|\,X) = E(Y) \quad \text{and} \quad V(Y\,|\,X) = V(Y) \end{array}}$

$\boxed{4.\ \text{If } E(Y\,|\,X) = E(Y), \text{ then } Cov(X,Y) = 0}$

$\boxed{5.\ E\left[E(Y\,|\,X)\right] = E(Y) \quad \text{"the law of iterated expectations"}}$

EX) If $E(WAGE\,|\,EDUC) = 4 + 0.6\,EDUC$ and $E(EDUC) = 11.5$, find $E(WAGE)$.

26

MICHIGAN STATE UNIVERSITY

Homework:

- WMS Ch. 5 (part 3 of 3)
  - Expected value of a function of RVs & special theorems: 5.72, 5.74
  - Covariance: 5.89, 5.91, 5.92 (Hint: $E(Y_1)=0.25$ and $E(Y_2)=0.5$)
  - Expected values, variances, covariances, and correlations of linear functions of RVs: 5.102, 5.103 (consult Theorem 5.12), 5.110
  - Conditional expectations: none but review & internalize the rules (and include them on your cheat sheet!)

Next class:

- Finish Ch. 5 (if need be) and answer any questions you have about the material for the midterm (Ch. 1-5 in WMS and integration)

27

MICHIGAN STATE UNIVERSITY

In-class exercise #1: calculating the covariance

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

E(Y_1)=0.75, E(Y_2)=0.375. Find Cov(Y_1, Y_2).

$$\boxed{E\left[g(Y_1,Y_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1,y_2)f(y_1,y_2)dy_1dy_2}$$

$$\boxed{Cov(Y_1,Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)}$$

28

MICHIGAN STATE UNIVERSITY

In-class exercise #2: variance of a linear function of 3 random variables

Use Theorem 5.12 (copied on the next slide) and find the formula for:

$$V(a_1Y_1 + b_1 + a_2Y_2 + b_2 + a_3Y_3 + b_3)$$

29

MICHIGAN STATE UNIVERSITY

Rules for the expected value, variance, and covariance of linear functions of RVs: ***The general multivariate case***

Proof on pp. 272-3

Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random variables with $E(Y_i) = \mu_i$ and $E(X_j) = \xi_j$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then the following hold:

a $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$.

b $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{1 \le i < j \le n} a_i a_j \text{Cov}(Y_i, Y_j)$, where the double sum is over all pairs (i, j) with $i < j$.

c $\text{Cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i, X_j)$.