Multivariate Distributions – Expected Value **Notes 06**

Associated Reading: Wackerly 7, Chapter 5, Sections 5-8 and 11

Y_1 $\mathbb{E}[Y_1] = \int y_1 f(y_1)\, dy_1$ (Domain of Y_1)

In Chapters 3 and 4, you were introduced to the expected value operator, which takes the weighted average of a random variable (or a function of that random variable) with respect to an assumed probability distribution. Now we will lay out definitions related to the expected value operator and how it acts upon multivariate distributions.

First, we combine Theorems 5.6-5.8 into one statement:

linearity of expectation.

$$\mathbb{E}[a_1 g_1(Y_1,Y_2) + a_2 g_2(Y_1,Y_2) + \cdots + g_n(Y_1,Y_2) + b] = \sum_{i=1}^{n} a_i\, \mathbb{E}[g_i(Y_1,Y_2)] + b.$$

$Y_1 \; Y_2 \; \cdots \; Y_n$

$$\mathbb{E}[g(Y_1,Y_2)] = \begin{cases} \iint_D g(y_1,y_2)\cdot f(y_1,y_2)\, dy_1\, dy_2 \\ \sum_{(y_1,y_2)} g(y_1,y_2)\cdot p(y_1,y_2) \end{cases}$$

$g: \mathbb{R}^2 \to \mathbb{R}$.

joint pdf. joint pmf.

Second, we note Theorem 5.9, which indicates to us that in general, $E[Y_1Y_2] \neq E[Y_1]E[Y_2]$:

$$\mathbb{E}[g(Y_1)\cdot h(Y_2)] = \iint_{D_2 D_1} g(y_1)\, h(y_1)\cdot f(y_1,y_2)\, dy_1\, dy_2.$$

① Suppose $\boxed{Y_1 \perp Y_2}$ (independent) $f(y_1,y_2) = f_1(y_1)\cdot f_2(y_2)$

$$\boxed{\mathbb{E}[g(Y_1)h(Y_2)]} = \iint_{D_2 D_1} g(y_1)\,h(y_2)\cdot f_1(y_1)\cdot f_2(y_2)\, dy_1\, dy_2.$$

$$= \int_{D_1} g(y_1) f_1(y_1)\, dy_1 \cdot \int_{D_2} h(y_2) f_2(y_2)\, dy_2.$$

$$\Leftarrow \boxed{\mathbb{E}[g(Y_1)] \cdot \mathbb{E}[h(Y_2)]} \quad (*)$$

suppose we take $g(Y_1) = Y_1 - \mu_1$, $h(Y_2) = Y_2 - \mu_2$. if $Y_1 \perp Y_2$,

$$\underbrace{\mathbb{E}[(Y_1-\mu_1)\cdot(Y_2-\mu_2)]}_{Cov(Y_1,Y_2)} = \mathbb{E}[Y_1-\mu_1]\cdot\mathbb{E}[Y_2-\mu_2] = 0$$

$\Downarrow$ $Y_1 \perp Y_2 \Downarrow Cov(Y_1,Y_2) = 0$

② "reverse direction" if $(*)$ holds true for every funtion pair (g,h) $\Rightarrow Y_1 \perp Y_2$.

$Cov(Y_1,Y_2) = 0 \not\Downarrow Y_1 \perp Y_2$.

These definitions, given in Sections 5.5 and 5.6, will be applied to two specific cases in Sections 5.7 and 5.8:

general def of $\mathbb{E}$. → covariance / correlation ⇒ section 5.7

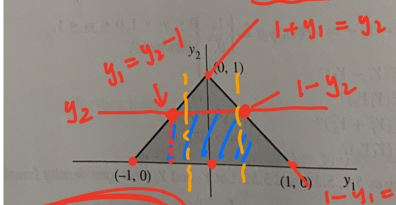
general def of $\mathbb{E}$. → linear funtions ⇒ section 5.8

$\forall g$ $\mathbb{E}[g]$

(Y_1, Y_2)

→ **EXAMPLE.** Wackerly 7, Exercise 5.79. Note that if we are working with a bivariate uniform distribution, then $f(y_1, y_2)$ within the region of integration is the reciprocal of the area of the region of integration.

5.79 Suppose that, as in Exercise 5.11, Y_1 and Y_2 are uniformly distributed over the triangle shaded in the accompanying diagram. Find $E(Y_1Y_2)$.



$$f(y_1,y_2) = \begin{cases} \dfrac{1}{\text{area(support)}} & (y_1,y_2)\in \text{support} \\ 0 & \text{o.w.} \end{cases}$$

$$f(y_1,y_2) = \frac{1}{2\times 1\times \frac12} = 1$$

$$\int_{D_1}\int_{D_2} y_1y_2 \, dy_2\, dy_1 = 0$$

offline

$$E[Y_1Y_2] = \int_{D_2}\int_{D_1} y_1y_2 \cdot \boxed{f(y_1,y_2)}\, dy_1\, dy_2.$$

$$= \int_{D_2}\int_{D_1} y_1y_2 \, dy_1\, dy_2.$$

$$= \int_0^1 \int_{y_2-1}^{1-y_2} y_1y_2\, dy_1\, dy_2 = \int_0^1 y_2\left(\frac{y_1^2}{2}\Big|_{y_2-1}^{1-y_2}\right) dy_2 = \int_0^1 y_2\cdot 0 \, dy_2 = \boxed{0}$$

→ **EXAMPLE.** Wackerly 7, Exercise 5.81. Using the test of Notes Set 5, we can convince ourselves that Y_1 and Y_2 are independent random variables.

5.81 In Exercise 5.18, Y_1 and Y_2 denoted the lengths of life, in hundreds of hours, for components of types I and II, respectively, in an electronic system. The joint density of Y_1 and Y_2 is

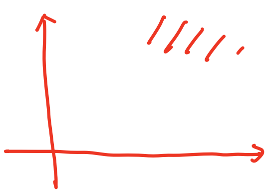
$$f(y_1,y_2) = \begin{cases} (1/8)y_1 e^{-(y_1+y_2)/2}, & y_1 > 0,\ y_2 > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

One way to measure the relative efficiency of the two components is to compute the ratio Y_2/Y_1. Find $E(Y_2/Y_1)$. [*Hint:* In Exercise 5.61, we proved that Y_1 and Y_2 are independent.]

option 2

$$\int_0^\infty\int_0^\infty \frac{y_2}{y_1} f(y_1,y_2)\, dy_1\, dy_2 = ?$$

exercise!



① support of $f(y_1,y_2)$ ⇒ rectangular.

② $f(y_1,y_2) = \frac18 y_1 \cdot e^{-\frac{(y_1+y_2)}{2}}$

$$= \underbrace{\frac18 y_1 \cdot e^{-\frac{y_1}{2}}}_{f_1(y_1)} \cdot \underbrace{e^{-\frac{y_2}{2}}}_{f_2(y_2)}$$

⇒ $Y_1 \perp Y_2$

$$E\left[\frac{Y_2}{Y_1}\right] = E\left[\frac{1}{Y_1}\cdot Y_2\right]$$

by independence → $\boxed{= E\left[\frac{1}{Y_1}\right]\cdot E[Y_2]}$

$E[Y_2] = \int_{D_2} y_2 \cdot f_2(y_2)\, dy_2 = 2$ ← Exp(2)

$$\boxed{f_2(y_2)} = \int_{D_1} f(y_1,y_2)\, dy_1$$

$$= \int_0^\infty \frac18 y_1 e^{-\frac{y_1}{2}} \cdot e^{-\frac{y_2}{2}}\, dy_1$$

$$= \frac18 e^{-\frac{y_2}{2}} \int_0^\infty \underbrace{y_1 e^{-\frac{y_1}{2}}}\, dy_1 = e^{-\frac{y_2}{2}}\cdot\frac18\cdot 2^2\cdot\Gamma(2) \underbrace{\int_0^\infty \frac{1}{2^2\cdot\Gamma(2)} y_1 e^{-\frac{y_1}{2}}\, dy_1}_{=1}$$

~ Gamma($\alpha = 2$, $\beta = 2$). exactly pdf Gamma(2,2)

pdf Exp(2) ⟸ $\boxed{= \frac12 e^{-\frac{y_2}{2}}}$

$E\left[\frac{1}{Y_1}\right]$ first need to $f_1(y_1)$

$$f_1(y_1) = \int_0^\infty f(y_1,y_2)\, dy_2$$

$$= \int_0^\infty \frac18 y_1 e^{-\frac{y_1}{2}} e^{-\frac{y_2}{2}}\, dy_2$$

Check this! $\boxed{=}$ $\frac14 y_1 e^{-\frac{y_1}{2}}$ ~ Gamma(2,2)

$$E\left[\frac{1}{Y_1}\right] = \int_0^\infty \cancel{\frac{1}{y_1}}\cdot\frac14 \cancel{y_1} e^{-\frac{y_1}{2}}\, dy_1 = \frac12$$

in Summary: $E\left[\frac{Y_2}{Y_1}\right] = E\left[\frac{1}{Y_1}\right]\cdot E[Y_2] = 1$

The covariance between two random variables Y_1 and Y_2, a measure of dependence between Y_1 and Y_2, is defined as

Covariance or $\boxed{Cov(Y_1, Y_2)} := \mathbb{E}[\underbrace{(Y_1-\mu_1)\cdot(Y_2-\mu_2)}_{g(Y_1,Y_2).}]$

which can also be written in shortcut form as

$$\boxed{Cov(Y_1,Y_2)} = \mathbb{E}[Y_1Y_2 - Y_2\mu_1 - Y_1\mu_2 + \mu_1\mu_2]$$

here $\mu_1 = \mathbb{E}[Y_1]$, $\mu_2 = \mathbb{E}[Y_2]$.

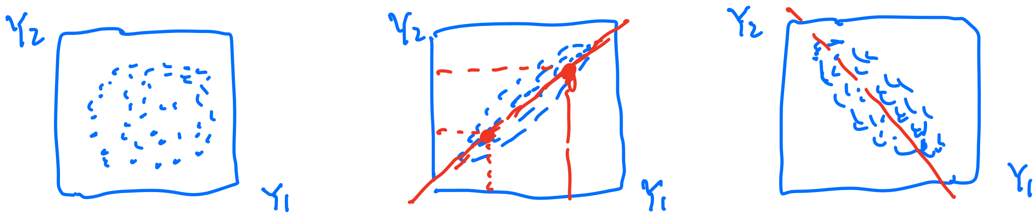
$$= \mathbb{E}[Y_1Y_2] - \underbrace{\mathbb{E}[Y_2]}_{\mu_2}\cdot\mu_1 - \underbrace{\mathbb{E}[Y_1]}_{\mu_1}\cdot\mu_2 + \mu_1\mu_2$$

$$= \boxed{\mathbb{E}[Y_1Y_2] - \mu_1\mu_2.}$$

remark: $Y_1 \perp Y_2 \Rightarrow \mathbb{E}[Y_1Y_2] = \mathbb{E}[Y_1]\cdot\mathbb{E}[Y_2]$

$\Rightarrow Cov(Y_1,Y_2) = 0$

The following contour plots will provide intuition about covariance and the values it can take on:



$Cov(Y_1,Y_2) \approx 0$ $\quad$ $Cov(Y_1,Y_2) > 0$ $\quad$ $Cov(Y_1,Y_2) < 0$

However when $Cov(Y_1,Y_2) = 0$ $\not\Rightarrow Y_1 \perp Y_2$

Covariance not be an optimal measure of dependence, as its value *is not readily interpretable*. An alternative way of specifying the linkage between two r.v.'s is given by the correlation coefficient, ρ:[a]

$$\boxed{\rho_{Y_1Y_2} = \frac{Cov(Y_1,Y_2)}{\sigma_{Y_1}\cdot\sigma_{Y_2}}} = \frac{\mathbb{E}[(Y_1-\mu_1)(Y_2-\mu_2)]}{\sqrt{\mathbb{E}[(Y_1-\mu_1)^2]\cdot\mathbb{E}[(Y_2-\mu_2)^2]}}$$

Variance (Y_1) · Variance (Y_2)

$Cov(Y_1,Y_2) \in [-\sigma_{Y_1}\sigma_{Y_2}, \sigma_{Y_1}\sigma_{Y_2}]$.

$\boxed{\text{Next time:}}$

$\boxed{-1 \le \rho_{Y_1Y_2} \le 1}$ ✓

$\rho = 1$: Y_1, Y_2 perfectly positive correlated

$\rho = -1$: Y_1, Y_2 ... negative correlated.

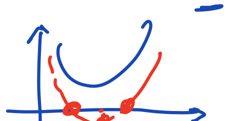
for t. $(tY_1 - Y_2)^2 \geq 0$

$\mathbb{E}(tY_1 - Y_2)^2 \geq 0$ $\Rightarrow$ $\underbrace{(2\mathbb{E}[Y_1Y_2])^2 - 4\cdot\mathbb{E}Y_1^2\cdot\mathbb{E}Y_2^2}_{\le 0}$

$\forall t$ $\quad t^2\underbrace{\mathbb{E}[Y_1^2]}_{a} - \underbrace{2t\,\mathbb{E}[Y_1Y_2]}_{b} + \underbrace{\mathbb{E}[Y_2^2]}_{c} \geq 0$ ✓

$$\frac{(\mathbb{E}[Y_1Y_2])^2}{\mathbb{E}[Y_1^2]\cdot\mathbb{E}[Y_2^2]} \le 1$$

recall $ax^2 + bx + c \geq 0$ holds for all x.



$\boxed{b^2 - 4ac \le 0}$

if $b^2 - 4ac > 0$

$\Rightarrow$ there are two Solutions to $ax^2 + bx + c = 0$

suppose $\mu_1 = \mathbb{E}[Y_1] = 0$, $\mu_2 = 0$

$\rho^2_{Y_1Y_2} \le 1$

[a] As with previous derivations of this ilk, this need not be reproduced on a test.

review for solving quadratic inequality of one variable.

$-1 \le \rho_{Y_1 Y_2} \le 1$

→ **EXAMPLE.** Wackerly 7, Exercise 5.89

5.89 In Exercise 5.1, we determined that the joint distribution of Y_1, the number of contracts awarded to firm A, and Y_2, the number of contracts awarded to firm B, is given by the entries in the following table.

(Y_1, Y_2)

y_2 \ y_1	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

Question: compute $\mathrm{cov}(Y_1, Y_2)$.

$\mathrm{cov}[Y_1 Y_2] = \mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_1] \cdot \mathbb{E}[Y_2]$

$\mathbb{E}[Y_1 Y_2] = \sum_{y_1} \sum_{y_2} y_1 y_2\, p(y_1, y_2)$

$= 1 \cdot \frac{2}{9} + 0 = \frac{2}{9}.$

$\mathbb{E}[Y_1] = \sum_{y_1} y_1 P_1(y_1) = \frac{2}{3}.$

$P(Y_1 = 0) = \sum_{y_2} p(0, y_2)$

$= \frac{4}{9}$

$P(Y_1 = 1) = \frac{4}{9}$

$P(Y_1 = 2) = \frac{1}{9}.$

$\mathbb{E}[Y_2] = \sum_{y_2} y_2 \left(\sum_{y_1} p(y_1, y_2) \right)$

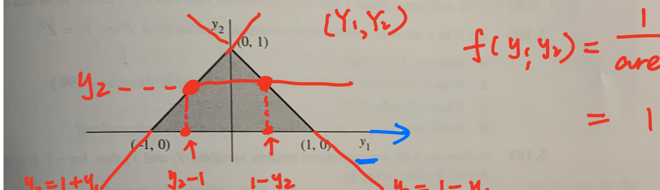
$= \frac{2}{3}.$

$\mathrm{cov}(Y_1 Y_2) = \frac{2}{9} - \frac{2}{3} \cdot \frac{2}{3} = -\frac{2}{9}.$

large value of Y_1 associated with smaller value of Y_2, vice-versa

→ **EXAMPLE.** Wackerly 7, Exercise 5.93

5.93 Suppose that, as in Exercises 5.11 and 5.79, Y_1 and Y_2 are uniformly distributed over the triangle shaded in the accompanying diagram.



$f(y_1, y_2) = \frac{1}{\text{area(support)}}$

$= 1$

a Find $\mathrm{Cov}(Y_1, Y_2)$.

b Are Y_1 and Y_2 independent? (See Exercise 5.55.)

c Find the coefficient of correlation for Y_1 and Y_2.

d Does your answer to part (b) lead you to doubt your answer to part (a)? Why or why not?

a) $\mathrm{cov}(Y_1 Y_2) = \underbrace{\mathbb{E}[Y_1 Y_2]}_{=0} - \mathbb{E}[Y_1] \cdot \mathbb{E}[Y_2] = 0$

$\mathbb{E}[Y_1] = \int y_1 f_1(y_1)\, dy_1$ $\quad f_1(y_1)$

$= \int y_1 \cdot \boxed{\int f(y_1, y_2)\, dy_2}\, dy_1 = \iint y_1 f(y_1, y_2)\, dy_1\, dy_2$

$= \int_0^1 \underbrace{\boxed{\int_{y_2 - 1}^{1 - y_2} y_1 f(y_1, y_2)\, dy_1}}_{=0} dy_2 = 0$

$\int_{y_2-1}^{1-y_2} y_1\, dy_1$

$= \frac{y_1^2}{2} \Big|_{y_2 - 1}^{1 - y_2} = 0$

b). $Y_1 \perp Y_2$?

$\Rightarrow$ not independent because shape of support is not rectangular.

$\Rightarrow$ notes 5.

c) $\rho_{Y_1 Y_2} = \frac{\mathrm{cov}(Y_1, Y_2)^{=0}}{\sigma_{Y_1} \cdot \sigma_{Y_2}} = 0$

d) not. $\mathrm{cov}(Y_1 Y_2) = 0$

$\not\Rightarrow Y_1 \perp Y_2.$

→ **EXAMPLE.** Wackerly 7, Exercise 5.97

5.97 The random variables Y_1 and Y_2 are such that $E(Y_1) = 4$, $E(Y_2) = -1$, $V(Y_1) = 2$ and $V(Y_2) = 8$.

a What is $\mathrm{Cov}(Y_1, Y_1)$?

b Assuming that the means and variances are correct, as given, is it possible that $\mathrm{Cov}(Y_1, Y_2) = 7$? [*Hint:* If $\mathrm{Cov}(Y_1, Y_2) = 7$, what is the value of ρ, the coefficient of correlation?]

c Assuming that the means and variances are correct, what is the largest possible value for $\mathrm{Cov}(Y_1, Y_2)$? If $\mathrm{Cov}(Y_1, Y_2)$ achieves this largest value, what does that imply about the relationship between Y_1 and Y_2?

a) $\mathrm{Cov}(Y_1, Y_1) = \mathbb{E}[Y_1 \cdot Y_1] - \mathbb{E}[Y_1] \cdot \mathbb{E}[Y_1] = \mathbb{E}[Y_1^2] - (\mathbb{E} Y_1)^2 = V[Y_1] = 2$

b) $\mathrm{cov}(Y_1 Y_2) = 7$?

$-1 \le \rho_{Y_1 Y_2} = \frac{\mathrm{cov}(Y_1 Y_2)}{\sigma_{Y_1} \sigma_{Y_2}} \le 1 \Rightarrow \mathrm{cov}(Y_1 Y_2) \in [\underbrace{-\sigma_{Y_1} \sigma_{Y_2}}_{-4}, \underbrace{\sigma_{Y_1} \sigma_{Y_2}}_{4}]$

$\sqrt{2}\ \sqrt{8}$

c) $\mathrm{cov}(Y_1 Y_2)_{max} = 4. \Rightarrow \rho_{Y_1 Y_2} = 1$

Y_1 & Y_2 $\Rightarrow Y_1 Y_2$ are perfectly positive correlated.

Let $U_1 = \sum_{i=1}^n a_i Y_i$ and $U_2 = \sum_{j=1}^m b_j X_j$ be linear functions of the n r.v.'s Y_i and the m r.v.'s X_j. The following statement is *always* true, regardless of whether the individual r.v.'s are independent:

$$\mathbb{E}[U_1] = \sum_{i=1}^n a_i \, \mathbb{E}[Y_i] \quad (\text{linearity of expectation operator}). \quad \mathbb{E}[U_2] = \sum_{j=1}^m b_j \mathbb{E}X_j$$

This next statement is also always true, with the second term being zero if the r.v.'s are independent:

$$V[U_1] = V\left[\sum_{i=1}^n a_i Y_i\right] \underset{\text{claim.}}{=} \sum_{i=1}^n a_i^2 V(Y_i) + 2\sum\sum_{1 \le i < j \le n} a_i a_j \operatorname{cov}(Y_i, Y_j)$$

Suppose $\mathbb{E}(Y_i) = 0$ $\forall i$

$$\mathbb{E}\left[(\underbrace{\textstyle\sum a_i Y_i}_{U_1}) \cdot (\underbrace{\textstyle\sum a_i Y_i}_{U_1})\right] = \mathbb{E}\left[\sum_i \sum_j a_i a_j \operatorname{cov}(Y_i, Y_j)\right]$$

This formula is the usual one that is presented in beginning mathematical statistics textbooks, since they generally assume that the reader has no knowledge of matrices. However, practicing statisticians by-and-large only use the matrix version of the formula. Hence I will present it below, right after I provide...

→ A Very Short Introduction to Matrices.[a]

matrix: a rectangular array of numbers.

$$X = \begin{array}{cc} \text{column 1} & \\ \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} & \text{row 1} \end{array} \in \mathbb{R}^{2\times 2}$$

single element: $X_{(\text{row}),(\text{column})}$

e.g. $X_{21} = 2 \quad X_{12} = 3.$

What we need to know for now.

① the vector of coefficients $\{a_1, a_2, \dots, a_n\}$ is an $n \times 1$ matrix. e.g. $a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^{n\times 1}$

② a matrix transpose reverses the rows & columns of the original matrix. e.g. $a^T = (a_1 \; a_2 \dots a_n) \in \mathbb{R}^{1\times n}$

e.g. $X^T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

③ matrix product.

Let $Y = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix}$

$$XY = \begin{pmatrix} \boxed{1 \quad 3} \\ \boxed{2 \quad 4} \end{pmatrix} \begin{pmatrix} \boxed{\begin{matrix}2\\1\end{matrix}} & \boxed{\begin{matrix}4\\3\end{matrix}} \end{pmatrix} = \begin{pmatrix} \overset{1\times 2 + 3\times 1}{\overset{\shortparallel}{5}} & 13 \\ 8 & 20 \end{pmatrix} \quad 20 = 2\times 4 + 4\times 3.$$

[a]The following is written assuming no knowledge of matrices at all, which is appropriate since linear algebra is not a pre- or co-requisite for this course. If you are already familiar with matrices, you can safely ignore this material. If you are not familiar with matrices, note that in addition to what follows here I have also written up an introductory document on matrices, which I have put on Canvas, under "Math." Note that in that document I provide examples of how to define and work with matrices in R, which may be useful for anyone in the class.

$Cov(Y_i Y_j) = \rho_{ij} \cdot \sigma_i \cdot \sigma_j$

The covariance between two random variables Y_i and Y_j is $\text{Cov}(Y_i, Y_j) = E[Y_iY_j] - E[Y_i]E[Y_j] = \rho_{ij}\sigma_i\sigma_j$. If one has k random variables, one can populate a matrix $\mathbf{\Sigma}$ with all pairwise covariances:

covariance matrix $(Y_1, Y_2, \cdots Y_n)$.

$$\Sigma = \begin{pmatrix} cov(Y_1 Y_1) & \cdots & & cov(Y_1 Y_n) \\ \vdots & & cov(Y_i Y_j) & \\ & & & \ddots \\ Cov(Y_n, Y_1) & & & cov(Y_n, Y_n) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \rho_{1n}\sigma_1\sigma_n & \cdots & & \sigma_n^2 \end{pmatrix}$$

$\in \mathbb{R}^{n \times n}$

Note that this matrix is *symmetric*; e.g., $\Sigma_{i,j} = \Sigma_{j,i}$, where the first and second subscripts denote the row and column of the matrix, respectively.

Now, let's look at a linear function of two random variables, $U = a_1Y_1 + a_2Y_2$. Now define two matrices with the coefficients a_1 and a_2:

$$a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad a^T = (a_1 \ a_2)$$

Then the variance of U is

Claim: $V[U] = \underset{1\times 2}{a^T} \underset{2\times 2}{\Sigma} \underset{2\times 1}{a} \in \mathbb{R}^1$

Does this match the result you'd get using the book's formulation? (Yes.)

$$V[U] = V[a_1Y_1 + a_2Y_2] = \sum_{i=1}^{2} a_i^2 V[Y_i] + 2\sum_{1\le i<j\le 2} a_i a_j cov(Y_i, Y_j)$$

(from before).

$$= a_1^2 V[Y_1] + a_2^2 V[Y_2] + 2 a_1 a_2 cov[Y_1 Y_2]$$

matrix notation: $a^T \Sigma a = (a_1 \ a_2) \begin{bmatrix} V[Y_1] & cov(Y_1 Y_2) \\ cov[Y_1 Y_2] & V[Y_2] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$

Last, this statement for $\text{Cov}(U_1, U_2)$ is always true, but it is equal to zero if the $Y_i's$ and $X_j's$ are all independent of each other:

$$U_1 = \sum_{i=1}^{n} a_i Y_i = (a_1 \ a_2 \cdots a_n) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$U_2 = \sum_{j=1}^{m} b_j X_j = (b_1 \cdots b_m) \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}.$$

$$V[U_1] = a^T \Sigma_Y a \qquad \Sigma_Y = \begin{pmatrix} Cov(Y_1 Y_1) & \cdots & cov(Y_1 Y_n) \\ \vdots & \ddots & \\ cov(Y_n Y_1) & & cov(Y_n, Y_n) \end{pmatrix}$$

$$V[U_2] = b^T \Sigma_X b \qquad \Sigma_X = \begin{pmatrix} cov(X_1 X_1) & \cdots & cov(X_1 X_m) \\ \vdots & \ddots & \\ Cov(X_m X_1) & & Cov(X_m X_m) \end{pmatrix}$$

① $Cov[U_1 + c, U_2 + d]$

c, d are two constants

$= \mathbb{E}[(U_1 + \cancel{c} - (\mathbb{E}U_1 + \cancel{c}))(U_2 + \cancel{d} - (\mathbb{E}U_2 + \cancel{d}))]$.

$= cov[U_1, U_2]$

② $cov[aU_1, bU_2]$

suppose $\mathbb{E}U_1 = \mathbb{E}U_2 = 0$

$= \mathbb{E}(aU_1 \cdot bU_2)$

$= ab \cdot Cov[U_1, U_2]$

$\boxed{Cov[U_1 U_2]} = \boxed{a^T \Sigma_{YX} b}$ $\qquad \Sigma_{YX} = \begin{pmatrix} cov(Y_1 X_1) & \cdots & Cov(Y_1 X_m) \\ \vdots & \ddots & \end{pmatrix} \in \mathbb{R}^{m \times n}$

$\Sigma_{ij} = \text{cov}(Y_i, X_j)$ ... $\text{cov}(Y_n, X_1)$ ... $\text{cov}(Y_n, X_m)$

→ **EXAMPLE.** Wackerly 7, Exercise 5.103

5.103 Assume that Y_1, Y_2, and Y_3 are random variables, with

$$E(Y_1) = 2,\quad V(Y_1) = 4,\quad \text{Cov}(Y_1, Y_2) = 1,$$
$$E(Y_2) = -1,\quad V(Y_2) = 6,\quad \text{Cov}(Y_1, Y_3) = -1,$$
$$E(Y_3) = 4,\quad V(Y_3) = 8,\quad \text{Cov}(Y_2, Y_3) = 0.$$

Find $E(3Y_1 + 4Y_2 - 6Y_3)$ and $V(3Y_1 + 4Y_2 - 6Y_3)$.

$$V[U_1] = \boxed{a^T \Sigma a}$$

$$\Sigma = \begin{pmatrix} 4 & 1 & -1 \\ 1 & 6 & 0 \\ -1 & 0 & 8 \end{pmatrix}$$

$\text{Cov}(Y_1 Y_2)$ $\quad$ $\text{cov}(Y_2 Y_3)$

① $\mathbb{E}[3Y_1 + 4Y_2 - 6Y_3] = 3\mathbb{E}Y_1 + 4\mathbb{E}Y_2 - 6\mathbb{E}Y_3 = -22$

② $V[3Y_1 + 4Y_2 - 6Y_3]$ $\quad$ let $\boxed{a^T = (3, 4, -6)}$

$$U_1 = a^T \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = 3Y_1 + 4Y_2 - 6Y_3.$$

$= 480.$

↑ check this!

→ **EXAMPLE.** Wackerly 7, Exercise 5.111(a)

5.111 Suppose that Y_1 and Y_2 have correlation coefficient ρ_{Y_1,Y_2} and for constants a, b, c and d let $W_1 = a + bY_1$ and $W_2 = c + dY_2$.

a Show that the correlation coefficient between W_1 and W_2, ρ_{W_1,W_2}, is such that $|\rho_{Y_1,Y_2}| = |\rho_{W_1,W_2}|$.

... results that you obtained in Exercise 5.110?

$$\rho_{W_1 W_2} = \frac{\text{cov}(W_1 W_2)}{\sqrt{V(W_1)\cdot V(W_2)}}$$

$\boxed{V(W_1)} = V(a + bY_1)$
$= b^2 \cdot V(Y_1)$

$\boxed{\text{cov}(W_1, W_2)} = \text{cov}(\cancel{a} + bY_1, \cancel{c} + dY_2).$
$= bd \cdot \text{cov}(Y_1, Y_2)$

$\boxed{V(W_2)} = V(c + dY_2) = d^2 V(Y_2).$

$$\rho_{W_1 W_2} = \frac{bd\, \text{cov}(Y_1 Y_2)}{|b| \sqrt{V(Y_1)} \cdot |d| \sqrt{V(Y_2)}} = \frac{bd}{|bd|} \cdot \rho_{Y_1 Y_2}$$

$\Rightarrow |\rho_{W_1 W_2}| = |\rho_{Y_1 Y_2}|$

→ **EXAMPLE.** Wackerly 7, Exercise 5.115(b,c)

5.115 Refer to Exercise 5.88. If Y denotes the number of tosses of the die until you observe each of the six faces, $Y = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6$ where Y_1 is the trial on which the first face is tossed, $Y_1 = 1$, Y_2 is the number of additional tosses required to get a face different than the first, Y_3 is the number of additional tosses required to get a face different than the first two distinct faces, ..., Y_6 is the number of additional tosses to get the last remaining face after all other faces have been observed.

a Show that $\text{Cov}(Y_i, Y_j) = 0, i, j = 1, 2, \ldots, 6, i \neq j$.

b Use Theorem 5.12 to find $V(Y)$.

c Give an interval that will contain Y with probability at least 3/4.

a) $Y_1 \cdots Y_6$ are independent $\Rightarrow \text{cov}(Y_i, Y_j) = 0$

b) $V(Y) = V\left(\boxed{\sum_{i=1}^{6} Y_i}\right) = a^T \Sigma a = \boxed{\sum_{i=1}^{6} V[Y_i]}$

let $U_1 = \sum_{i=1}^{6} a_i Y_i$

Set $a_i = 1$

$$\Sigma = \begin{pmatrix} V[Y_1] & & 0 \\ & V[Y_2] & \\ & & \ddots \\ 0 & & V[Y_6] \end{pmatrix}$$

$Y_1 = 1$

$Y_2 \sim \text{Geom}\left(\frac{5}{6}\right)$

$Y_3 \sim \text{Geom}\left(\frac{4}{6}\right)$

$\vdots$

$Y_i \sim \text{Geom}\left(P_i = \frac{7-i}{6}\right)$

$$\sum_{i=1}^{6} V[Y_i] = \sum_{i=1}^{6} \cdot \frac{(1-P_i)}{P_i^2} = \sum_{i=1}^{6} \left(\frac{6}{7-i}\right)^2 \cdot \left(\frac{i-1}{6}\right) \approx 39.$$

$$\mathbb{E}[Y] = \sum_{i=1}^{6} \cdot \frac{1}{P_i} = \sum_{i=1}^{6} \cdot \frac{6}{7-i} = 14.7.$$

$[14.7 - 2\sqrt{39},\ 14.7 + 2\sqrt{39}]$

c). Goal: find interval (a, b) st: $P(Y \in \boxed{(a, b)}) \geq \frac{3}{4}.$

Recall Chebesheff inequality.

$P(-k\sigma \le Y - EY \le k\sigma) \ge 1 - \frac{1}{k^2}$

$k = 2$

$P(-2\sqrt{39} \le Y - 14.7 \le 2\sqrt{39}) \ge \frac{3}{4}$

Recall that the conditional pmf and pdf are written as $p(y_1|y_2)$ and $f(y_1|y_2)$. We can now extend the concept of the conditional by combining it with the Law of ~~the Unconscious~~ Statistician:

$$p(y_1|y_2) = \frac{p(y_1,y_2)}{p_2(y_2)} = \frac{p(y_1,y_2)}{\sum_{y_1} p(y_1,y_2)} \quad \text{joint pmf}$$

$$f(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)} = \frac{f(y_1,y_2)}{\int f(y_1,y_2)\,dy_1}$$

$$E[Y_1 | Y_2 = y_2] = \sum_{y_1} y_1 \boxed{p(y_1|y_2)}$$

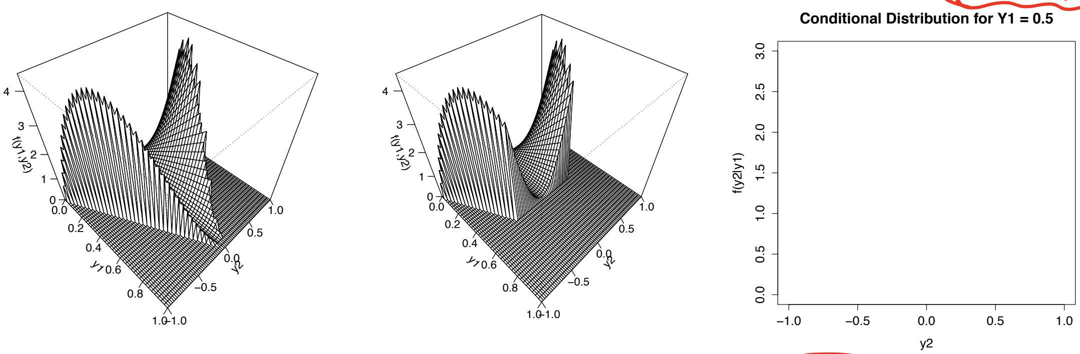
or

$$E[g(Y_1) | Y_2 = y_2] = \int y_1 f(y_1|y_2)\,dy_1$$

$$= \sum_{y_1} g(y_1)\, p(y_1|y_2).$$

The following will help illustrate the concept of conditional expectation:

$E[Y_1 | Y_2 = y_2] = h(y_2)$



Perhaps you are given a situation where you are given (or can derive) $E[Y_1|Y_2]$. Then, assuming Y_1 and Y_2 are continuous r.v.'s, we can write the unconditional expectation as:

$$E[Y_1] = \int_{D_1} y_1 \boxed{f_1(y_1)}\,dy_1$$

$$= \int_{D_1} y_1 \boxed{\int_{D_2} f(y_1,y_2)\,dy_2}\,dy_1$$

$$= \int_{D_1} y_1 \int_{D_2} f(y_1|y_2) f_2(y_2)\,dy_2\,dy_1$$

$$= \int_{D_2} f_2(y_2) \underbrace{\boxed{\int_{D_1} y_1 f(y_1|y_2)\,dy_1}}_{E[Y_1 | Y_2 = y_2]}\,dy_2.$$

$$= \int_{D_2} f_2(y_2) \cdot E[Y_1 | Y_2 = y_2]\,dy_2.$$

with respect to Y_1

$$= E\big(\underbrace{E[Y_1|Y_2]}_{h(Y_2)}\big)$$

with respect to Y_2 $\qquad h(Y_2)$.

→ **EXAMPLE.** Wackerly 7, Exercise 5.133

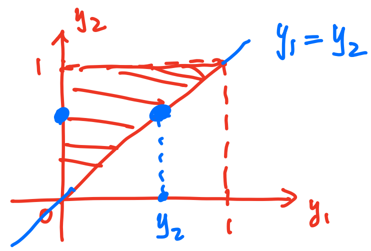
5.133 In Exercise 5.9, we determined that

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

is a valid joint probability density function.

a Find $E(Y_1 | Y_2 = y_2)$.

b Use the answer derived in part (a) to find $E(Y_1)$. (Compare this with the answer found in Exercise 5.77.)



a) $\mathbb{E}[Y_1 | Y_2 = y_2] = \int y_1 f(y_1 | y_2)\, dy_1$

$= \int y_1 \cdot \frac{f(y_1, y_2)}{f_2(y_2)}\, dy_1.$

$f_2(y_2) = \int_{D_1} f(y_1, y_2)\, dy_1 = \int_0^{y_2} 6(1-y_2)\, dy_1 = 6(1-y_2)\cdot y_2.$

$\mathbb{E}[Y_1 | Y_2 = y_2] = \int_0^{y_2} y_1 \frac{\cancel{6(1-y_2)}}{\cancel{6(1-y_2)}\cdot y_2}\, dy_1 = \frac{y_2}{2}$ (for $0 \le y_2 \le 1$).

b). $\mathbb{E}[Y_1] = \mathbb{E}[\underbrace{\mathbb{E}[Y_1 | Y_2]}]$.

marginal of y_2.

$= \mathbb{E}[\frac{Y_2}{2}] = \int_0^1 f_2(y_2)\cdot \frac{y_2}{2}\, dy_2 = 3\int_0^1 \underbrace{(1-y_2)\cdot y_2^2}\, dy_2 = 3 \cdot B(3, 2).$

Beta$(\alpha = 3, \beta = 2) = 3 \cdot \frac{\Gamma(3)\cdot\Gamma(2)}{\Gamma(5)} = \frac{1}{4}.$

use $\Gamma(n) = (n-1)!$

→ **EXAMPLE.** Wackerly 7, Exercise 5.139

5.139 Suppose that a company has determined that the the number of jobs per week, N, varies from week to week and has a Poisson distribution with mean λ. The number of hours to complete each job, Y_i, is gamma distributed with parameters α and β. The total time to complete all jobs in a week is $T = \sum_{i=1}^{N} Y_i$. Note that T is the sum of a random number of random variables. What is

a $E(T \mid N = n)$?

b $E(T)$, the expected total time to complete all jobs?

$N \sim \text{Poiss}(\lambda)$.

$Y_i \sim \text{Gamma}(\alpha, \beta)$.

$T = \sum_{i=1}^{N} Y_i$

a). $\mathbb{E}(T | N = n) = \mathbb{E}[\sum_{i=1}^{n} Y_i | N = n] = \sum_{i=1}^{n} \underbrace{\mathbb{E}[Y_i | N = n]}_{\alpha\beta}$

Gamma(α, β)

$= n \cdot \alpha\beta$

b) $\mathbb{E}[T] = \mathbb{E}[\underbrace{\mathbb{E}[T | N]}_{N\cdot\alpha\cdot\beta}] = \alpha\beta\, \mathbb{E}[N] = \lambda \cdot \alpha \cdot \beta.$

Poiss(λ)

$P(\underbrace{N, Y_i}, T) \Rightarrow P(T) \Rightarrow \mathbb{E}[T]$

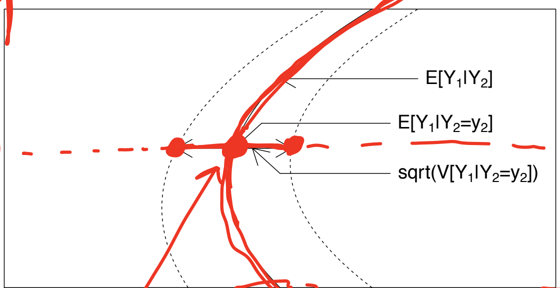
$\int y_1^2\, p(Y_1|Y_2)\, dy_1$

The last topic we will touch upon in these notes is the concept of conditional variance. (We follow the book by not generalizing this to the conditional variance of $g(Y_1)$...and for notational simplicity we change "$Y_2 = y_2$" to "Y_2.")

$$V[Y_1|Y_2] = \underbrace{E[Y_1^2|Y_2]}_{h(Y_2)} - (\underbrace{E[Y_1|Y_2]}_{g(Y_2)})^2$$

Conditional probabilities are probabilities (short-cut formula for variance still works)

Given $V[Y_1|Y_2]$, we can state the unconditional variance $V[Y_1]$.[a]



$$\boxed{V[Y_1] = E(V[Y_1|Y_2]) + V(E[Y_1|Y_2]).}$$

Pf: $E[V[Y_1|Y_2]]$

$= E[E[Y_1^2|Y_2] - (E[Y_1|Y_2])^2]$

$= E[Y_1^2] - E[E[Y_1|Y_2]^2]$

$V[E[Y_1|Y_2]]$

$= E(E[Y_1|Y_2])^2 - (\underbrace{E[E[Y_1|Y_2]]}_{})^2$

$= E[E[Y_1|Y_2]^2] \qquad (E[Y_1])^2$

Sum up. $= \boxed{E[Y_1^2] - [EY_1]^2} = V(Y_1)$

$E[Y_1 | Y_2 = y_2]$

$Y_1 | Y_2 = y_2$

the average variation $V[Y_1|Y_2]$. Variation in $E[Y_1|Y_2]$

$\Rightarrow E[V[Y_1|Y_2]] \quad \Rightarrow V[E[Y_1|Y_2]]$.

→ **EXAMPLE.** Wackerly 7, Exercise 5.135

5.135 In Exercise 5.41, we considered a quality control plan that calls for randomly selecting three items from the daily production (assumed large) of a certain machine and observing the number of defectives. The proportion p of defectives produced by the machine varies from day to day and has a uniform distribution on the interval (0, 1). Find the

a expected number of defectives observed among the three sampled items.

b variance of the number of defectives among the three sampled.

$\boxed{p \sim Unif(0,1).}$

Y: # of defective items.

$Y|p \sim Bin(n=3, p)$.

$E[p] = \frac{1}{2} \quad V[p] = \frac{1}{12}$.

a) $E[Y]$.

$= E[E[Y|p]]$

$= E[3 \times p] = 3E[p] = \frac{3}{2}$

b). $V[Y]$.

$= E[V[Y|p]] + V[E[Y|p]]$.

$= E[3p(1-p)] + V[3 \times p]$.

$= 3E[p - p^2] + 9 \cdot V[p]$.

$= 3 \cdot (\frac{1}{2} - E[p^2]) + 9 \cdot V(p) = 5$

[a]See Wackerly 7, p. 287 for the derivation.

$= 3\cdot(2 - E[P]) + \ldots V(P) \ldots \, 4.$

$(E(P))^2 + V(P)$ $\quad \frac{1}{12}.$

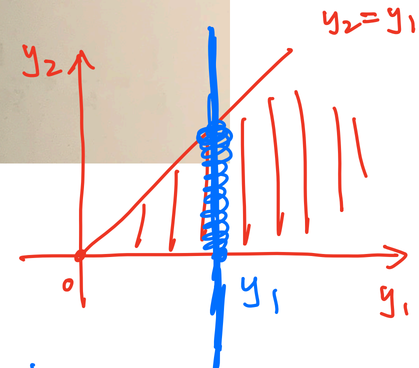
→ **EXAMPLE.** Wackerly 7, Exercise 5.141

5.141 Let Y_1 have an exponential distribution with mean λ and the conditional density of Y_2 given $Y_1 = y_1$ be

pdf

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

Find $E(Y_2)$ and $V(Y_2)$, the unconditional mean and variance of Y_2.



$Y_1 \sim \text{Exp}(\lambda)$

$Y_2 \mid Y_1 \sim \text{Unif}(0, Y_1)$

a) $\mathbb{E}[Y_2] = \mathbb{E}[\underbrace{\mathbb{E}[Y_2 \mid Y_1]}_{\frac{Y_1}{2}}] = \frac{1}{2}\mathbb{E}[\underbrace{Y_1}_{\text{Exp}(\lambda)}] = \frac{1}{2}\lambda$

b) $V[Y_2] = \mathbb{E}[V[Y_2 \mid Y_1]] + V[\mathbb{E}[Y_2 \mid Y_1]].$

$\mathbb{E}[\underbrace{V[Y_2 \mid Y_1]}_{Y_2 \mid Y_1 \sim \text{Unif}[0,\, Y_1]}] = \mathbb{E}[\frac{1}{12} Y_1^2] = \frac{1}{12}\underbrace{\mathbb{E}[Y_1^2]}_{\parallel} = \frac{\lambda^2}{6}$

$(\underbrace{E[Y_1]}_{\Downarrow \, \lambda^2})^2 + \underbrace{Var(Y_1)}_{\lambda^2}.$

$V[\mathbb{E}[Y_2 \mid Y_1]] = V[\frac{1}{2}Y_1] = \frac{1}{4}\cdot V[Y_1] = \frac{\lambda^2}{4}$

$\Rightarrow \quad V[Y_2] = \frac{\lambda^2}{4} + \frac{\lambda^2}{6} = \frac{5}{12}\lambda^2.$

---

an equivalent way:

$f_1(y_1) = \frac{1}{\lambda}\exp(-\frac{y_1}{\lambda}) \quad (y_1 \ge 0).$

$f_2(y_2 \mid y_1) = \frac{1}{y_1} \quad (y_2 \le y_1)$

$f(y_1, y_2) = f_1(y_1) f_2(y_2 \mid y_1) = \frac{1}{\lambda y_1}\exp(-\frac{y_1}{\lambda}).$ $\quad (y_2 \le y_1)$

$\mathbb{E}[Y_2] = \int_{y_2} y_2 \underbrace{\boxed{\int_{y_1 \ge y_2} f(y_1, y_2)\, dy_1}}_{f_2(y_2)} dy_2$

$\Rightarrow$ double integration

$V(Y_2) = \ldots$