

Reference: Chapter 5 of Devore

Maghsoodloo

JOINT PROBABILITY DISTRIBUTION FUNCTIONS

Consider two production lines that manufacture a certain item. The production rates for both lines vary randomly from day to day. Line 1 has a capacity of 4 units per day while line II has a capacity of 3 units per day. Further, both lines produce at least one unit on any given day. Let X_1 = No. of units produced by line I, and X_2 = No. of units produced by line II per day. The joint probability (Pr)

distribution (JPD) of the bivariate vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is given below:

$X_1 \backslash X_2$	1	2	3	$p_1(x_1)$
1	0.01	0.05	0.04	0.10
2	0.05	0.10	0.10	0.25
3	0.10	0.15	0.10	0.35
4	0.04	0.15	0.11	0.30
$p_2(x_2)$	0.20	0.45	0.35	

The above table implies that the joint Pr $P(X_1 = 2, X_2 = 3) = p(2, 3) = 0.10$, and $p(4, 2) = 0.15$, etc. Further, $p_1(x_1)$ and $p_2(x_2)$ are referred to as the marginal Pr distributions

(mpds) of X_1 and X_2 , respectively. Note that $p_1(x_1) = \sum_{R_2} p(x_1, x_2)$ and $p_2(x_2) =$

$$\sum_{R_{x1}} p(x_1, x_2) = \sum_{R_1} p(x_1, x_2).$$

Further, $\mu_1 = E(X_1) = 0.10 + 0.50 + 1.05 + 1.20 = 2.85$ units/day, and

$$\mu_2 = E(X_2) = 0.20 + 0.90 + 1.05 = 2.15 \text{ units/day.}$$

Similarly,

$$E(X_1^2) = 9.05 \longrightarrow \sigma_1^2 = \sigma_{11} = 0.9275 \longrightarrow \sigma_1 = 0.9631$$

$$E(X_2^2) = 5.15 \longrightarrow \sigma_2^2 = \sigma_{22} = 0.5275 \longrightarrow \sigma_2 = 0.7263.$$

The covariance between 2 random variables (rvs) is defined as:

$$\sigma_{12} = \text{COV}(X_1, X_2) = E [(X_1 - \mu_1)(X_2 - \mu_2)] = E (X_1 X_2) - \mu_1 \mu_2.$$

For the above example,

$$E(X_1 X_2) = 0.01 + 2 \times 0.05 + 3 \times 0.04 + 2 \times 0.05 + 4 \times 0.10 + 6 \times 0.10 + \\ 3 \times 0.10 + 6 \times 0.15 + 9 \times 0.10 + 4 \times 0.04 + 8 \times 0.15 + 12 \times 0.11 = 6.11$$

$$\rightarrow \sigma_{12} = 6.11 - 2.85 (2.15) = -0.0175$$

The covariance matrix of the bivariate random vector X is given by:

$$\text{COV}(X) = \text{COV}\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 0.9275 & -0.0175 \\ -0.0175 & 0.5275 \end{bmatrix}$$

Note that the covariance matrix Σ is always symmetrical because $\sigma_{ij} = \sigma_{ji}$ for all $i \neq j$.

Further, covariance must be taken only between two rvs at a time (not 3).

The Pearson correlation coefficient between X_1 and X_2 is defined as:

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11} \sigma_{22}}} = \frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{-0.0175}{(0.9631)(0.7263)} = -0.02502.$$

It can be shown that $-1 \leq \rho \leq +1$, where $\rho = 0$ implies no correlation between X_1 and X_2 ($\rho = 0$ does not always imply that X_1 and X_2 are independent but shows that there is no linear relationship between X_1 and X_2). A value of $\rho = \pm 1$ implies perfect correlation between X_1 and X_2 . A positive $0 < \rho \leq 1$ implies that the relationship between x_1 and x_2 is linearly increasing and vice versa when $-1 \leq \rho < 0$. For example, there is a positive correlation between $X_1 =$ the amount of irrigation, and $X_2 =$ crop yield. While, there is a negative correlation between $X_1 =$ width of road, and $X_2 =$ accident rate. Just like covariance, the Pearson correlation coefficient is defined only between two rvs at-a-time.

CONDITIONAL PROBABILITY DISTRIBUTIONS

The conditional Pr distribution of X_2 given $X_1 = x_1$ is defined as:

$$p_2(x_2 | x_1) = \frac{p(x_1, x_2)}{p_1(x_1)}, \text{ and similarly, } p_1(x_1 | x_2) = \frac{p(x_1, x_2)}{p_2(x_2)}.$$

As an example, for the JPD on page 1, $p_2(x_2 | X_1 = 1) = \frac{p(1, x_2)}{0.10}$, i.e.,

$$p_2(x_2 | X_1 = 1) = \begin{cases} 0.10, & x_2 = 1 \\ 0.50, & x_2 = 2, \\ 0.40, & x_2 = 3 \end{cases} \text{ while } p_1(x_1 | X_2 = 3) = \begin{cases} 4/35, & x_1 = 1 \\ 10/35, & x_1 = 2, 3 \\ 11/35, & x_1 = 4. \end{cases}$$

Exercise 36.

(a) Obtain $p_2(x_2 | X_1 = i)$, $i = 2, 3$, or 4 .

(b) Obtain $p_1(x_1 | X_2 = i)$, $i = 1$ or 2 .

CONDITIONAL EXPECTATIONS

These are defined as follows: $E(X_2 | x_1) = \sum_{R_2} x_2 p_2(x_2 | x_1)$, and

$E(X_1 | x_2) = \sum_{R_1} x_1 p_1(x_1 | x_2)$, where $R_1 = R_{x_1}$ and $R_2 = R_{x_2}$. For example,

$E(X_2 | X_1 = 1) = 0.10 + 1 + 1.20 = 2.30$, and $E(X_1 | X_2 = 3) = 2.80$.

Exercise 36 (continued).

(c) Compute $E(X_2 | X_1 = i)$, $i = 2, 3$, or 4 and $E(X_1 | X_2 = i)$, $i = 1$ or 2 .

Note that for any bivariate random vector X , it is always true that

$p(x_1, x_2) = p_1(x_1) \times p(x_2 | x_1) = p_2(x_2) \times p(x_1 | x_2)$. For the JPD on page 67, $p(1, 3) = 0.04$, $p_1(1) \times p_2(3 | X_1 = 1) = 0.10 (4/10) = 0.04$, or $p_2(X_2 = 3) = 0.35$, $p_1(X_1 = 1 | X_2 = 3) = 4/35$, $p_2(X_2 = 3) \times p_1(X_1 = 1 | X_2 = 3) = 0.35 (4/35) = 0.04 = p(1, 3)$.

Exercise 36 (continued). (d) Verify that $p(3, 2) = p_1(3) \times p_2(X_2 = 2 | X_1 = 3) = p_2(2) \times p_1(X_1 = 3 | X_2 = 2)$. (e) Compute the $P(X_1 > 1 | X_2 > 2)$.

INDEPENDENCE OF TWO RANDOM VARIABLES

Two random variables, X_1 and X_2 , are independent iff (if and only if) $p(x_1, x_2) = p_1(x_1) \times p_2(x_2)$. If X_1 and X_2 are independent, then always $\sigma_{12} = 0$ and hence $\rho = 0$.

Note that the converse of this last claim is not necessarily true (see Exercise 38

below) unless the random vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ has a bivariate normal density

function. In short, two rvs are independent iff their JPDF factors out into the product of the individual mpds.

Exercise 37. A shop has 2 machines M_1 and M_2 . Let the rv X_i = Number of defective units produced per hour on M_i ($i = 1, 2$). The JPDF of random vector $X =$

$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is given below. (a) Obtain the mpdfs of X_1 and X_2 and the covariance matrix

$$\Sigma = \text{COV}\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right).$$

$X_1 \backslash X_2$	1	2	3	4	$p_1(x_1)$
0	0.02	0.08	0.08	0.02	
1	0.03	0.12	0.12	0.03	
2	0.03	0.12	0.12	0.03	
3	0.02	0.08	0.08	0.02	
$p_2(x_2)$					

(b) Compute $E(X_2 | X_1 = 3)$, $E(X_2 | X_1 = 2)$ and $E(X_2)$. (c) Determine if X_1 and X_2 are independent.

Exercise 38. Repeat all parts of Exercise 37 for the following JPDF.

$X_1 \backslash X_2$	0	1	2	3	$p_1(x_1)$
0	1/12	1/12	1/12	1/12	
1	1/12	0	0	1/12	
2	1/12	0	0	1/12	
3	1/12	1/12	1/12	1/12	
$p_2(x_2)$					

CONTINUOUS BIVARIATE RANDOM VARIABLES

Suppose X_1 represents surface tension and X_2 represents the acidity of the same sampling unit of a chemical product. The joint probability density function

(jpdf) of the random vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is given by

$$f(x_1, x_2) = C(6 - x_1 - x_2), \quad 0 \leq x_1 \leq 2, \quad 2 \leq x_2 \leq 4.$$

Example 32. (a) Determine the value of the above constant C such that $f(x_1, x_2)$ is a jpdf, i.e., find C such that the volume under $f(x_1, x_2)$ and rectangular region $R_X = [0 \leq x_1 \leq 2, \text{ and } 2 \leq x_2 \leq 4]$ is equal to 1. That is,

$$\int_{x_2=2}^4 \int_{x_1=0}^2 (6 - x_1 - x_2) dx_1 dx_2 = \int_2^4 C \left[6x_1 - \frac{x_1^2}{2} - x_2 x_1 \right]_0^2 dx_2 = 1$$

$$C \int_2^4 (12 - 2 - 2x_2) dx_2 = C \left[10x_2 - x_2^2 \right]_2^4 = C(24 - 16) = 8C = 1 \rightarrow$$

$C = 0.125$. Thus $f(x_1, x_2) = 0.125(6 - x_1 - x_2)$ is a jpdf over $R_X : 0 \leq x_1 \leq 2,$

$2 \leq x_2 \leq 4$ because the volume under $f(x_1, x_2)$ is equal to 100%.

(b) Compute the joint Pr that a randomly selected unit has a surface tension less than 1 and an acidity not exceeding 3.

$$P(X_1 \leq 1, X_2 \leq 3) = \frac{1}{8} \int_0^1 \int_2^3 (6 - x_1 - x_2) dx_2 dx_1 = 3/8.$$

(c) We next compute the $P(X_1 + X_2 \leq 4)$, or the $P(X_2 \leq 4 - X_1)$.

$$P(X_1 + X_2 \leq 4) = 0.125 \int_{x_1=0}^2 \int_{x_2=2}^{4-x_1} (6 - x_1 - x_2) dx_2 dx_1 = 2/3.$$

Exercise 39. (a) Re-compute the above $P(X_1 + X_2 \leq 4)$ by integrating with respect to (wrt) x_1 first followed by x_2 .

MARGINAL PROBABILITY DENSITY FUNCTIONS

Analogous to the discrete case, the mpdf of the continuous rv X_1 is defined as

$$f_1(x_1) = \int_{R_2} f(x_1, x_2) dx_2 = \int_{x_2=2}^4 0.125(6 - x_1 - x_2) dx_2 = \frac{3 - x_1}{4}, \quad 0 \leq x_1 \leq 2.$$

$$\text{Therefore, } E(X_1) = \int_0^2 x_1 f_1(x_1) dx_1 = \int_0^2 x_1 \frac{3 - x_1}{4} dx_1 = 5/6.$$

Exercise 39 (b). Obtain the mpdf of X_2 for the Example 32 and verify that both $f_1(x_1)$ and $f_2(x_2)$ are indeed probability density functions. (c) Compute the Pr $(X_2 \leq 3)$ and $E(X_2)$.

CONDITIONAL PROBABILITY DENSITY FUNCTIONS

The conditional pdf of X_2 given x_1 is defined as $f(x_2 | x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} =$
 $= \frac{6 - x_1 - x_2}{2(3 - x_1)}$, $2 \leq x_2 \leq 4$. Since the expression for $f(x_2 | x_1) = \frac{6 - x_1 - x_2}{2(3 - x_1)}$ is not

free of x_1 , then the rv X_2 is not independent of X_1 .

Exercise 39(d). Verify that $f(x_2 | x_1)$ is indeed a pdf over the range $R_2 = [2, 4]$. Then obtain $f(x_1 | x_2)$ and determine if X_1 is independent of X_2 . Verify your answer over the range $R_1 = [0, 2]$.

CONDITIONAL EXPECTATIONS

The conditional expectation of X_2 given the value of x_1 is defined as

$$E(X_2 | x_1) = \int_{R_2} x_2 f(x_2 | x_1) dx_2 = \int_2^4 x_2 \left(\frac{6 - x_1 - x_2}{6 - 2x_1} \right) dx_2 = \frac{26 - 9x_1}{3(3 - x_1)}.$$

Note that since X_2 is not independent of X_1 , then the $E(X_2 | x_1)$ is a function of x_1 over the range space $R_1 = [0, 2]$.

Exercise 39(e). Compute $E(X_2 | X_1 = 0.50)$ and obtain $E(X_1 | x_2)$ and use it to re-compute the unconditional expectation $E(X_1)$. Use $E(X_2 | x_1)$ and $f_1(x_1)$ to re-compute the unconditional $E(X_2)$. (f) Obtain the covariance matrix Σ . (ANS: $\sigma_{11} = 11/36$, $\rho = -1/11$). (g) Obtain the $V(X_2 | x_1)$.

Exercise 40. (a) Show that $-1 \leq \rho \leq 1$ for all bivariate random vectors. Hint: Expand $V(c_1 X_1 + c_2 X_2)$ and use the fact that $V(c_1 X_1 + c_2 X_2) \geq 0$ for all choices of real constants c_1 and c_2 . (b) Show that $\rho = +1$ if $X_2 = a + b X_1$, but $\rho = -1$ when $X_2 = a - b X_1$, where the constant $b > 0$.

Exercise 41. Consider the uniform jpdf

$$f(x_1, x_2) = \begin{cases} 1, & 0 \leq x_1 \leq 1, \quad -x_1 \leq x_2 \leq x_1 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Draw the triangular region $R_X = [0 \leq x_1 \leq 1, -x_1 \leq x_2 \leq x_1]$ and obtain the covariance matrix Σ . (b) Verify that $\rho = 0$ but yet X_1 and X_2 are not independent.

Note that a necessary (but not sufficient) condition for two rvs to be independent is that their range space R_X must be rectangular.

LINEAR COMBINATIONS (WHEN INDIVIDUAL COMPONENTS MAY BE CORRELATED)

Suppose X_1, X_2, \dots, X_n are random variables with known means $\mu_1, \mu_2, \dots, \mu_n$ and known variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, respectively, and covariances σ_{ij}

($i \neq j$). Then the rv $Y = \sum_{i=1}^n c_i X_i$, where c_i 's are known constants, is called a linear

combination (LC). In other words, we have complete information about the 1st two moments of the n inputs X_i 's, and the objective is to use them to compute $E(Y)$ and $V(Y)$, i.e., the 1st two moments of the linear output Y , as shown below.

$$\mu_Y = E(Y) = E\left[\sum_{i=1}^n c_i X_i\right] = \sum_{i=1}^n c_i E(X_i) = \sum_{i=1}^n c_i \mu_i \quad (31a)$$

Note that the $E(Y)$ is the same LC of μ_i 's as Y is of X_i 's! We next compute the σ_Y^2 by applying the nonlinear variance operator V .

$$\sigma_Y^2 = V(Y) = E(Y - \mu_Y)^2 = E\left[\left(\sum_{i=1}^n c_i X_i - \sum_{i=1}^n c_i \mu_i\right)^2\right] = E\left[\sum_{i=1}^n c_i (X_i - \mu_i)\right]^2$$

$$\begin{aligned}
&= E \left[\sum_{i=1}^n c_i^2 (X_i - \mu_i)^2 + \sum_{j \neq i} \sum_{i=1}^{n-1} c_i c_j (X_i - \mu_i)(X_j - \mu_j) \right] \\
&= \sum_{i=1}^n c_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n c_i c_j E[(X_i - \mu_i)(X_j - \mu_j)] = \\
&= \sum_{i=1}^n c_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n c_i c_j \sigma_{ij} = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \sigma_{ij} \quad (31b)
\end{aligned}$$

If the rvs X_1, X_2, \dots, X_n are independent, then $\sigma_{ij} = 0$ for all $i \neq j$ in equation (31b) and as a result the $V(Y)$ reduces to $\sum_{i=1}^n c_i^2 \sigma_i^2$ as before. Further, if X_i 's are also normally

distributed (besides being jointly independent), then $Y \sim N(\sum_{i=1}^n c_i \mu_i, \sum_{i=1}^n c_i^2 \sigma_i^2)$. For

example, the sample mean \bar{X} from a normal universe is a LC whose $c_i = 1/n$ for all $i = 1, 2, \dots, n$ so that $\bar{X} \sim N(\mu, \sum_{i=1}^n (1/n)^2 \sigma_x^2)$, or $\bar{X} \sim N(\mu, \sigma_x^2 / n)$. However, if X_i 's are

correlated (i.e., $\sigma_{ij} \neq 0$) and normally distributed, then the linear combination $Y =$

$$\sum_{i=1}^n c_i X_i \text{ is also Gaussian with } E(Y) = \sum_{i=1}^n c_i \mu_i \text{ and } V(Y) = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \sigma_{ij} .$$

Simple RANDOM SAMPLING

Suppose X is a continuous random variable with pdf $f(x; \mu, \sigma^2)$ and let a random sample of size n be drawn from this population. Denote the n sample values by x_1, x_2, \dots, x_n ; then X_1, X_2, \dots, X_n are random variables with pdfs $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$. The method of sampling, which possesses the following two properties, is called random sampling:

(1) X_1, X_2, \dots, X_n are mutually independent. (2) $f(x_i) = f(x)$ for all i .

Therefore, if X_1, X_2, \dots, X_n are elements of a random sample, then $E(X_i) = \mu$ and $V(X_i) = \sigma^2$ for all i because all X_i 's are identically distributed like the parent pdf $f(x; \mu, \sigma^2)$.

Exercise 42. Let \bar{x} be the mean of a random sample of size n from a population with mean μ and variance σ^2 . Show that $E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma^2/n$. (b) Further, if the population is normal, then \bar{X} is also $N(\mu, \sigma^2/n)$. (c) Now consider the LC: $Y = 2X_1 - 3X_2 - 4X_3 + 5X_4$, where $\mu_1 = 50$, $\mu_2 = \mu_3 = 25$, $\mu_4 = 35$, $\sigma_1^2 = \sigma_4^2 = 1.25$, $\sigma_2^2 = \sigma_3^2 = 1.95$, $\sigma_{12} = 1.40$, $\sigma_{34} = 1.50$ and all other covariances are 0. Assuming that Y is normally distributed, compute the $\Pr(Y > 110)$.

Exercise 43. Work Exercises 1, 3, 15, 37, 39, 42, 46, 47, 50, 53, 56, 58, 59, 60, 65, 73, 76, 77, and 78 on pages 215-250 of Devore's 6th Edition.

Exercise 44. The smog content of air in a certain area is monitored daily. The acceptable content of a particular constituent is at 7.7%. If the actual content, X , of this constituent is $N(7.6, 0.0016)$, and the measuring instrument has an error ε which is $N(0, 0.0009)$, compute: (a) The Pr that a single measurement will exceed 7.7%, (b) The Pr that the mean of 5 measurements is less than 7.55. ANS: (a) 0.02275, (b) 0.012674.

Exercise 45. Suppose X , Y and Z are NID (normally and independently distributed) with means 100, 48, 48 and variances 10, 13 and 13, respectively. (a) Compute the $\Pr(X > Y + Z)$. ANS: 0.74751. (b) Work Exercise 89 on page 251 of your text.

TEST 3