Section 5.1-5.2

Tuesday, November 1, 2016 1:04 PM

Section
5.1-5.2

5 Multivariate Distributions

Complementary reading from WMS: Chapter 5.

5.1 Introduction

REMARK: Up until now, we have discussed **univariate** random variables (and their associated probability distributions, moment generating functions, means, variances, etc.). In practice, however, one is often interested in multiple random variables. Consider the following examples:

- In an educational assessment program, we want to predict a student's posttest score (Y_2) from her pretest score (Y_1).
- In a clinical trial, physicians want to characterize the concentration of a drug (Y) in one's body as a function of the time (X) from injection.
- An insurance company wants to estimate the amount of loss related to collisions Y_1 and liability Y_2 (both measured in 1000s of dollars).
- Agronomists want to understand the relationship between yield (Y, measured in bushels/acre) and the nitrogen content of the soil (X).
- In a marketing study, the goal is to forecast next month's sales, say Y_n, based on sales figures from the previous $n-1$ periods, say $Y_1, Y_2, ..., Y_{n-1}$.

2 r.v.s,

NOTE: In each of these examples, it is natural to posit a relationship between (or among) the random variables that are involved. This relationship can be described mathematically through a probabilistic model. This model, in turn, allows us to make probability statements involving the random variables (just as univariate models allow us to do this with a single random variable).

TERMINOLOGY: If Y_1 and Y_2 are random variables, then

$$\boldsymbol{Y} = (Y_1, Y_2)$$

is called a **bivariate random vector**. If $Y_1, Y_2, ..., Y_n$ denote n random variables, then

$$\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$$

is called an **n-variate random vector**.

5.2 Discrete random vectors

TERMINOLOGY: Let Y_1 and Y_2 be discrete random variables. Then, (Y_1, Y_2) is called a **discrete random vector**, and the **joint probability mass function (pmf)** of Y_1 and Y_2 is given by

$$p_{Y_1,Y_2}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2),$$

Univariat

$p_Y(y) = P(Y=y)$

for all $(y_1, y_2) \in R$. The set $R \subseteq \mathcal{R}^2$ is the two dimensional support of (Y_1, Y_2). The function $p_{Y_1,Y_2}(y_1, y_2)$ has the following properties:

(1) $p_{Y_1,Y_2}(y_1, y_2) > 0$, for all $(y_1, y_2) \in R$

(2) $\sum_R p_{Y_1,Y_2}(y_1, y_2) = 1$.

RESULT: Suppose that (Y_1, Y_2) is a discrete random vector with pmf $p_{Y_1,Y_2}(y_1, y_2)$. Then,

$$P[(Y_1, Y_2) \in B] = \sum_B p_{Y_1,Y_2}(y_1, y_2),$$

for any set $B \subset \mathcal{R}^2$. That is, the probability of the event $\{(Y_1, Y_2) \in B\}$ is obtained by adding up the probability (mass) associated with each support point in B. If $B = (-\infty, y_1] \times (-\infty, y_2]$, then

$$P[(Y_1, Y_2) \in B] = P(Y_1 \leq y_1, Y_2 \leq y_2) \equiv F_{Y_1,Y_2}(y_1, y_2) = \sum_{t_1 \leq y_1} \sum_{t_2 \leq y_2} p_{Y_1,Y_2}(t_1, t_2)$$

is called the **joint cumulative distribution function (cdf)** of (Y_1, Y_2).

Example 5.1. Tornados are natural disasters that cause millions of dollars in damage each year. An actuary determines that the annual numbers of tornadoes in two Iowa counties (Lee and Van Buren) are jointly distributed as indicated in the table below. Let Y_1 and Y_2 denote the number of tornados seen each year in Lee and Van Buren counties, respectively.

$P(Y_1 = 0) = 0.25$

$p_{Y_1,Y_2}(y_1, y_2)$	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$	$y_2 = 3$
$y_1 = 0$	0.12	0.06	0.05	0.02
$y_1 = 1$	0.13	0.15	0.12	0.03
$y_1 = 2$	0.05	0.15	0.10	0.02

(a) What is the probability that there is no more than one tornado seen in the two counties combined?

SOLUTION. We want to compute $P(Y_1 + Y_2 \boxed{\leq 1})$. Note that the support points which correspond to the event $\{Y_1 + Y_2 \leq 1\}$ are $(0, 0), (0, 1)$ and $(1, 0)$. Thus,

$$\begin{aligned} P(Y_1 + Y_2 \leq 1) &= p_{Y_1,Y_2}(0, 0) + p_{Y_1,Y_2}(1, 0) + p_{Y_1,Y_2}(0, 1) \\ &= 0.12 + 0.13 + 0.06 = 0.31. \end{aligned}$$

(b) What is the probability that there are two tornadoes in Lee County?

SOLUTION. We want to compute $P(Y_1 = 2)$. Note that the support points which correspond to the event $\{Y_1 = 2\}$ are $(2, 0), (2, 1), (2, 2)$ and $(2, 3)$. Thus,

$$\begin{aligned} P(Y_1 = 2) &= p_{Y_1,Y_2}(2, 0) + p_{Y_1,Y_2}(2, 1) + p_{Y_1,Y_2}(2, 2) + p_{Y_1,Y_2}(2, 3) \\ &= 0.05 + 0.15 + 0.10 + 0.02 = 0.32. \ \square \end{aligned}$$

5.3 Continuous random vectors

TERMINOLOGY: Let Y_1 and Y_2 be continuous random variables. Then, (Y_1, Y_2) is called a **continuous random vector**, and the **joint probability density function (pdf)** of Y_1 and Y_2 is denoted by $f_{Y_1,Y_2}(y_1, y_2)$. The joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ is a three-dimensional function whose domain is R, the two-dimensional support of (Y_1, Y_2).