Chapter 5: Multivariate Distributions

- In previous chapters, we have considered observing a single r.v. per individual.
- However, in many applications, we measure several r.v.'s per individual.

Example 1: For a set of households, we measure Y_1 = income and Y_2 = number of children.

Example 2: For a set of people, we measure Y_1 = height, Y_2 = weight, and Y_3 = cholesterol level.

- The set of r.v.'s measured on each individual is called a random vector, usually denoted by

5.2 Bivariate Probability Distributions

- If Y_1 and Y_2 are r.v.'s, we may consider the bivariate random vector

$$\underset{\sim}{Y} = (Y_1, Y_2).$$

Discrete Bivariate Distributions

Defn. The joint probability function (joint pmf) of two discrete r.v.'s Y_1 and Y_2 is

- The joint (bivariate) pmf gives the probability that the <u>pair</u> (Y_1, Y_2) takes a specific pair of values (y_1, y_2).

<u>Example 3</u>: Researchers at South Carolina, Georgia, and Florida are applying for two separate grants. Suppose all the proposals are equally good, and so which universities get the contracts can be seen as a random selection. Let $Y_1 =$ number of contracts awarded to S. Carolina and $Y_2 =$ number of contracts awarded to Georgia.

Sample space:

- Note all sample points are equally likely here.
- The (bivariate) joint pmf of Y_1 and Y_2 can be represented with a two-way table.

Theorem: (Properties of a discrete joint pmf):
If Y_1 and Y_2 are discrete r.v.'s with
joint pmf $p(y_1, y_2)$, then:

(1)

(2)

where we sum over all possible (y_1, y_2) pairs.
- Property (2) implies that there are a
finite or <u>countably infinite</u> number of
(y_1, y_2) pairs having positive probability.
<u>Example 3 again</u>: Find the probability that USC
obtains at least as many contracts as Georgia.

Defn. (joint cdf): The (bivariate) joint cdf of the random vector (Y_1, Y_2) is

- If Y_1, Y_2 are jointly discrete,

Example 3 again: Find $F(1,1)$.

Continuous Bivariate Distributions

A random vector (Y_1, Y_2) is <u>continuous</u> (i.e., Y_1 and Y_2 are <u>jointly continuous</u>) if their joint cdf $F(y_1, y_2)$ is continuous in both arguments.

Defn. (joint pdf): Let Y_1 and Y_2 be jointly continuous r.v.'s with joint cdf $F(y_1, y_2)$. If there is a nonnegative function $f(y_1, y_2)$ such that

$$F(y_1, y_2) = $$

then $f(y_1, y_2)$ is called the <u>joint probability density function</u> (joint pdf) of (Y_1, Y_2).

Properties of joint cdf

(1)

(2)

(3)

(4)

Properties of joint pdf

(1)

(2)

- Note that property (2) for joint pdf's follows directly from property (2) of joint cdf's.
- A (bivariate) joint pdf may be plotted as a 3-d function on the (y_1, y_2) plane:

- Volumes under this surface correspond to probabilities (like areas under curve in univariate case).

Example 4: Suppose the random vector (Y_1, Y_2) has joint pdf

$$f(y_1, y_2) = \begin{cases} y_1 + y_2 & \text{for } 0 < y_1 < 1, \ 0 < y_2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the joint cdf.

- Consider the diagram:

Finding Probabilities with Joint pdf's

- With bivariate distributions, we should <u>always</u> sketch the region of support in the (y_1, y_2) plane.

- Finding a probability amounts to integrating the joint pdf over a particular region of that support.

Brief Review of Double Integrals over a Region

- To integrate $f(y_1, y_2)$ over some region in the (y_1, y_2) plane, we first determine the type of region:

<u>Rectangular region</u> :

<u>Vertically Simple region:</u>

Horizontally simple region:

Complex region:

-In some cases (e.g., a triangular region), the region may be treated as __either__ vertically simple __or__ horizontally simple.

Example 4 again: Find $P(Y_1 + Y_2 < 1)$

Region of support of (Y_1, Y_2):

Note: $Y_1 + Y_2 < 1 \Rightarrow$

Example 5: Suppose the time (in hours) to complete task 1 and task 2 for a random employee has the joint pdf:

$$f(y_1, y_2) = \begin{cases} e^{-(y_1 + y_2)} & \text{for } y_1 > 0, \ y_2 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

- Find the probability that a random employee takes less than 2 hours on task 1 and between 1 and 3 hours on task 2.

Example 5(a): What is the probability the employee takes longer on task 2 than on task 1?