CHAPTER 5
Multivariate Probability Distributions

Motivation

- The intersection of two or more events is frequently of interest to an experimenter.
  - A gambler playing blackjack is interested in the event of drawing both an ace and a face card from a 52-card deck
  - Observing both the height and the weight of an individual

Motivation

- Suppose that $Y_1, Y_2, \dots, Y_n$ denote the outcomes of n successive trials of an experiment.
  For example,
  - the weights of n people
  - the measurements of *n* physical characteristics for a single person.

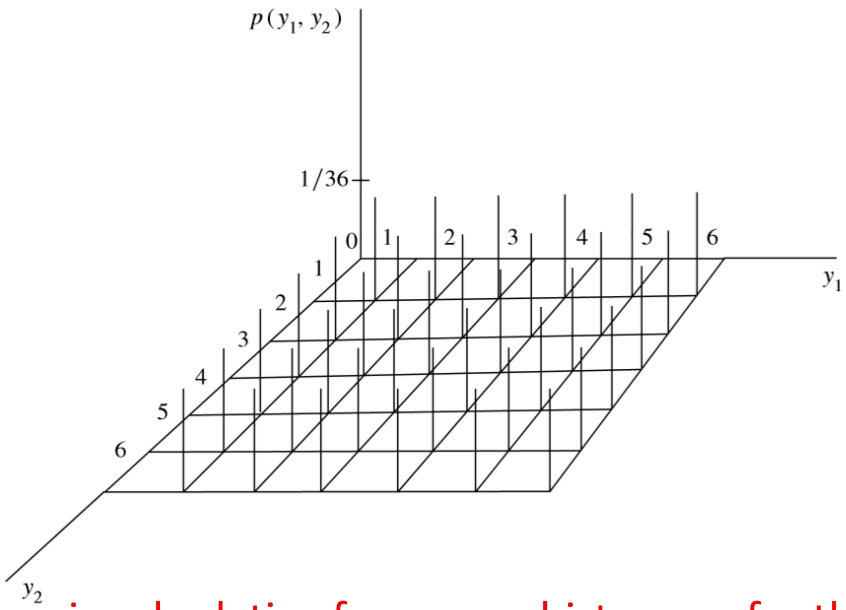
- A specific set of outcomes, or sample measurements, may be expressed in terms of the intersection of the n events
  - $(Y_1 = y_1), (Y_2 = y_2), \dots, (Y_n = y_n)$
  - $(Y1 = y_1, Y_2 = \ y_2, \dots, Y_n = y_n)$,
  - $(y_1, y_2, \dots, y_n)$.

- Calculation of the probability of this intersection is essential in making inferences about the population from which the sample has been drawn.
  - ➢ Main reason for studying multivariate probability distribution

Joint Events in Sample Space

- Many random variables can be defined over the same sample space. For example, tossing a pair of dice
  - The sample space contains 36 sample points, corresponding to the $mn = (6)(6) = 36$ ways
  - The 36 sample points associated with the experiment are equiprobable and correspond to the 36 numerical events (y_1, y_2).
  - For this simple example, the intersection (y_1, y_2) contains at most one sample point. Hence, the bivariate probability function is $p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) = \frac{1}{36}$,

$$y_1 = 1, 2, \ldots, 6, y_2 = 1, 2, \ldots, 6.$$



<Theoretical, three-dimensional relative frequency histogram for the pairs of observations (y_1, y_2)>

The Joint Probability Function (Discrete r.v.)

DEFINITION 5.1

Let Y_1 and Y_2 be discrete random variables. The *joint* (or bivariate) *probability function* for Y_1 and Y_2 is given by

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

- the bivariate case the joint probability function $p(y1, y2)$
  - Assigns nonzero probabilities to only a finite or countable number of pairs of values $(y1, y2)$.
  - The nonzero probabilities must sum to 1.

Properties of The Joint Probability Function

THEOREM 5.1

If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$, then

1. $p(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\Sigma_{y_1, y_2} p(y1, y2) = 1$, where the sum is over all values (y_1, y_2) that are assigned nonzero probabilities.

- The joint probability function for discrete random variables is sometimes called the **joint probability mass function** because it specifies the probability (mass) associated with each of the possible pairs of values for the random variables.

- Once the joint probability function has been determined for discrete random variables Y_1 and Y_2, calculating joint probabilities involving Y_1 and Y_2 is straightforward.
  - For the die-tossing experiment, $P(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2)$ is

$$P(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2) = p(2,1) + p(2,2) + p(3,1) + p(3,2) = 4/36 = 1/9.$$

The Joint Probability Function : Example

EXAMPLE 5.1

A local supermarket has three checkout counters. Two customers arrive at the counters at different times when the counters are serving no other customers. Each customer chooses a counter at random, independently of the other. Let Y_1 denote the number of customers who choose counter 1 and Y_2, the number who select counter 2. Find the joint probability function of Y_1 and Y_2.

The Joint Distribution Function

DEFINITION 5.2

For any random variables Y_1 and Y_2, the joint (bivariate) distribution function $F(y_1, y_2)$ is

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty.$$

The Joint Distribution Function : Example

EXAMPLE 5.2

Consider the random variables Y_1 and Y_2 of Example 5.1. Find $F(-1, 2)$, $F(1.5, 2)$, and $F(5, 7)$.

Joint Probability Density Function

DEFINITION 5.3

Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. If there exists a nonnegative function $f(y_1, y_2)$, such that

$$F(y_1, y_2) = \int_{-\infty}^{y1} \int_{-\infty}^{y2} f\,(t_1, t_2) dt_2\, dt_1,$$

for all $-\infty < y_1 < \infty, -\infty < y_2 < \infty$, then Y_1 and Y_2 are said to be *jointly continuous random variables*. The function $f(y_1, y_2)$ is called the *joint probability density function*.

- Note that the characteristics of $F(y_1, y_2)$ specifies whether Y_1 and Y_2 are continuous or not

Properties of Joint Distribution Function

THEOREM 5.2

If Y_1 and Y_2 are random variables with joint distribution function $F(y_1, y_2)$, then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$.
2. $F(\infty, \infty) = 1$.
3. If $y_1^* \geq y_1$ and $y_2^* \geq y_2$, then

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) \geq 0.$$

- Notice that $F(\infty, \infty) \equiv \lim_{y_1 \to \infty} \lim_{y_2 \to \infty} F(y_1, y_2) = 1$ implies that the joint density function $f(y_1, y_2)$ must be such that the integral of $f(y_1, y_2)$ over all values of (y_1, y_2) is 1.

- Part 3 follows because

$$F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) + F(y_1, y_2) = P(y_1 \leq Y_1 \leq y_1^*, y_2 \leq Y_2 \leq y_2^*) \geq 0.$$

Properties of Joint Density Function

THEOREM 5.2

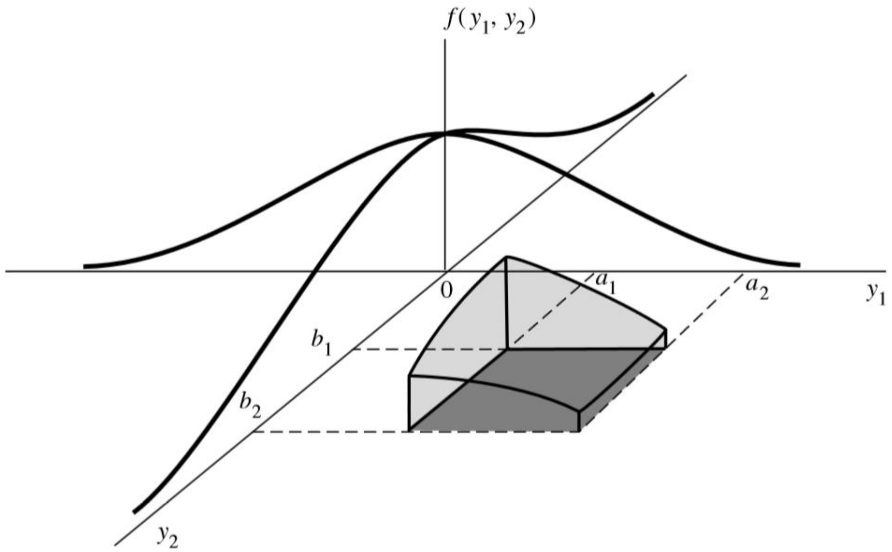
If Y_1 and Y_2 are jointly continuous random variables with a joint density function given by $f(y_1, y_2)$, then

1. $f(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1\, dy_2 = 1$.

- The bivariate probability density function $f(y_1, y_2)$ traces a probability density surface over the (y_1, y_2) plane

- Volumes under this surface correspond to probabilities. Thus, $P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2)$ is the shaded volume shown in Figure is

$$\int_{b_1}^{b_2} \int_{a_1}^{a_2} f(y_1, y_2) dy_1 dy_2$$



Properties of Joint Density Function : Example

EXAMPLE 5.3

Suppose that a radioactive particle is randomly located in a square with sides of unit length. That is, if two regions within the unit square and of equal area are considered, the particle is equally likely to be in either region. Let *Y*1 and *Y*2 denote the coordinates of the particle's location. A reasonable model for the relative frequency histogram for *Y*1 and *Y*2 is the bivariate analogue of the univariate uniform density function:

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & e\ lsewhere \end{cases}$$

a Sketch the probability density surface.
b Find $F(0.2, 0.4)$
c Find $P(0.1 \le Y_1 \le 0.3, 0 \le Y_2 \le 0.5)$

Properties of Joint Density Function : Example

SOLUTION 5.3

Properties of Joint Density Function : Example

EXAMPLE 5.4

Gasoline is to be stocked in a bulk tank once at the beginning of each week and then sold to individual customers. Let Y_1 denote the proportion of the capacity of the bulk tank that is available after the tank is stocked at the beginning of the week. Because of the limited supplies, Y_1 varies from week to week. Let Y_2 denote the proportion of the capacity of the bulk tank that is sold during the week. Because Y_1 and Y_2 are both proportions, both variables take on values between 0 and 1. Further, the amount sold, y_2, cannot exceed the amount available, y_1. Suppose that the joint density function for Y_1 and Y_2 is given by

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

Find the probability that less than one-half of the tank will be stocked and more than one-quarter of the tank will be sold.

Properties of Joint Density Function : Example

SOLUTION 5.4

Motivation

- Recall that the distinct values assumed by a discrete random variable represent mutually exclusive events.

- Similarly, for all distinct pairs of values y_1, y_2, the bivariate events $(Y_1 = y_1, Y_2 = y_2)$, represented by (y_1, y_2), are mutually exclusive events

- It follows that the univariate event $(Y_1 = y_1)$ is the union of bivariate events of the type $(Y_1 = y_1, Y_2 = y_2)$, with the union being taken over all possible values for y_2.

Marginal Probability Distribution: Motivation

Consider the die-tossing experiment

Y_1 = number of dots on the upper face of die 1,
Y_2 = number of dots on the upper face of die 2.

Then,

$$P(Y_1 = 1) = p(1,1) + p(1,2) + p(1,3) + \cdots + p(1,6) = 1/36 + 1/36 + 1/36 + \cdots + 1/36 = 6/36 = 1/6$$

$$P(Y_1 = 2) = p(2,1) + p(2,2) + p(2,3) + \cdots + p(2,6) = 1/6$$

$$\vdots$$

$$P(Y_1 = 6) = p(6,1) + p(6,2) + p(6,3) + \cdots + p(6,6) = 1/6.$$

Expressed in summation notation, probabilities about the variable $Y1$ alone are

$$P(Y_1 = y_1) = p_1(y_1) = \sum_{y_2=1}^{6} p(y_1, y_2)$$

$$P(Y_2 = y_2) = p_2(y_2) = \sum_{y_1=1}^{6} p(y_1, y_2)$$

Marginal Probability Distribution

DEFINITION 5.4

a Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1, y_2)$. Then the ***marginal probability functions*** of Y_1 and Y_2, respectively, are given by

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

b Let Y_1 and Y_2 be jointly continuous random variables with joint density function $f(y_1, y_2)$. Then the ***marginal density functions*** of Y_1 and Y_2, respectively, are given by

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

Marginal Probability Distribution : Example

EXAMPLE 5.5

From a group of three Republicans, two Democrats, and one independent, a committee of two people is to be randomly selected. Let Y_1 denote the number of Republicans and Y_2 denote the number of Democrats on the committee. Find the joint probability function of Y_1 and Y_2 and then find the marginal probability function of Y_1.

Marginal Probability Distribution : Example

SOLUTION 5.5

Marginal Probability Distribution : Example

EXAMPLE 5.6

Let

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

Sketch $f(y_1, y_2)$ and find the marginal density functions for Y_1 and Y_2.

Marginal Probability Distribution : Example

SOLUTION 5.6

Motivation

- The multiplicative law gives the probability of the intersection $A \cap B$ as $P(A \cap B) = P(A)P(B|A)$,
- Now consider the intersection of the two numerical events, $(Y_1 = y_1)$ and $(Y_2 = y_2)$, represented by the bivariate event (y_1, y_2).
  - It follows directly from the multiplicative law of probability that the bivariate probability for the intersection (y_1, y_2) is $p(y_1, y_2) = p_1(y_1)p(y_2|y_1) = p_2(y_2)p(y_1|y_2)$.

Conditional Probability Distribution: Motivation

- The probability of the intersection $A \cap B$

$$P(A \cap B) = P(A)P(B|A),$$

- The probability of the intersection of the two numerical events, $(Y_1 = y_1)$ and $(Y_2 = y_2)$, represented by the bivariate event (y_1, y_2):

$$\begin{aligned} p(y_1, y_2) &= p_1(y_1)p(y_2|y_1) \\ &= p_2(y_2)p(y_1|y_2). \end{aligned}$$

  - ✓ The probabilities $p_1(y_1)$ and $p_2(y_2)$ are associated with the univariate probability distributions for Y_1 and Y_2 individually
  - ✓ $p(y1|y2)$ is the probability that the random variable Y_1 equals y_1, given that Y_2 takes on the value y_2.

Conditional Probability Distribution : Discrete Case

DEFINITION 5.5

If Y_1 and Y_2 are jointly discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then the *conditional discrete probability function* of Y_1 given Y_2 is

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

provided that $p_2(y_2) > 0$.

- Note that $p(y_1|y_2)$ is undefined if $p_2(y_2) = 0$.

Conditional Probability Distribution : Continuous r.v.

- If Y1 and Y_2 are continuous, $P(Y_1 = y_1 | Y_2 = y_2)$ **cannot be defined** as in the discrete case
  - ✓ because both $(Y_1 = y_1)$ and $(Y_2 = y_2)$ are events with zero probability.

- Assuming that Y_1 and Y_2 are jointly continuous with density function $f(y_1, y_2)$, we might be interested in a probability of the form

$$P(Y_1 \leq y_1 | Y_2 = y_2) = F(y_1 | y_2),$$

which, as a function of y_1 for a fixed y_2, is called the conditional distribution function of Y_1, given $Y_2 = y_2$.

Conditional Probability Distribution Function : Continuous r.v.

DEFINITION 5.6

If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$, then the *conditional distribution function* of Y_1 given $Y_2 = y_2$ is

$$F(y_1|y_2) = P(Y_1 \leq y_1 | Y_2 = y_2).$$

- Notice that $F(y_1|y_2)$ is a function of y_1 for a fixed value of y_2.

$$F(y_1) = \int_{-\infty}^{\infty} F(y_1|y_2) f_2(y_2) dy_2 \quad \left(\because \text{the law of total probability } \sum_{y_2} P(y_1|y_2)P(y_2) = P(y_1)\right)$$

$$F(y_1) = \int_{-\infty}^{y_1} f_1(t_1) dt_1 = \int_{-\infty}^{y_1} \left[\int_{-\infty}^{\infty} f(t_1, y_2) dy_2\right] dt_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1 dy_2 \quad \left(\because f_1(t_1) = \int_{-\infty}^{\infty} f(t_1, y_2) d_2\right)$$

$$\int_{-\infty}^{\infty} F(y_1|y_2) f_2(y_2) dy_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1 dy_2 \Rightarrow F(y_1|y_2) = \frac{\int_{-\infty}^{y_1} f(t_1, y_2)}{f_2(y_2)}$$

Conditional Probability Density : Continuous r.v.

DEFINITION 5.7

Let Y_1 and Y_2 be jointly continuous random variables with joint density $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. For any y_2 such that $f_2(y_2) > 0$, the conditional density of Y_1 given $Y_2 = y_2$ is given by

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

and, for any y_1 such that $f_1(y_1) > 0$, the conditional density of Y_2 given $Y_1 = y_1$ is given by

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}.$$

- Note that the conditional density $f(y_1|y_2)$ is undefined for all y_2 such that $f_2(y_2)$. Similarly, $f(y_2|y_1)$ is undefined if y_1 is such that $f_1(y_1) = 0$.

Conditional Probability Density : Example

EXAMPLE 5.8

A soft-drink machine has a random amount Y_2 in supply at the beginning of a given day and dispenses a random amount Y_1 during the day (with measurements in gallons). It is not resupplied during the day, and hence $Y_1 \leq Y_2$. It has been observed that Y_1 and Y_2 have a joint density given by

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

That is, the points (y_1, y_2) are uniformly distributed over the triangle with the given boundaries. Find the conditional density of Y_1 given $Y_2 = y_2$. Evaluate the probability that less than $1/2$ gallon will be sold, given that the machine contains 1.5 gallons at the start of the day.

Conditional Probability Density : Example

SOLUTION 5.8

Motivation

- Probabilities associated with Y_1 were the same, regardless of the observed value of Y_2.
- We now present a formal definition of independence of random variables.
- The two events A and B are independent if

$$P(A \cap B) = P(A) \times P(B).$$

- The two random variables Y_1 and Y_2 are independent, if

$$P(a < Y_1 \leq b, c < Y_2 \leq d) \; = \; P(a < Y1 \leq b) \times P(c < Y2 \leq d)$$

  - That is, if Y_1 and Y_2 are independent, the joint probability can be written as the product of the marginal probabilities

Independence

DEFINITION 5.8

Let Y_1 have distribution function $F_1(y_1)$, Y_2 have distribution function $F_2(y_2)$, and Y_1 and Y_2 have joint distribution function $F(y_1, y_2)$. Then Y_1 and Y_2 are said to be *independent* if and only if

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

for every pair of real numbers (y_1, y_2).

If Y_1 and Y_2 are not independent, they are said to be *dependent*.

Generalization:

- Suppose that we have n random variables, $Y_1, \dots, Y_n$,
  - Y_i has distribution function $F_i(y_i)$, for $i = 1,2, \dots, n$
  - $Y_1, \dots, Y_n$ have joint distribution function $F(Y_1, \dots, Y_n)$

- $Y_1, \dots, Y_n$ are independent if and only if

$$F(Y_1, \dots, Y_n) = F(Y_1) \cdots F(Y_n)$$

for all real numbers $y_1, y_2, \dots, y_n$, with the obvious equivalent forms for the discrete and continuous cases.

Independence

THEOREM 5.4

If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$$p(y_1, y_2) = p_1(y_1)p_2(y_2)$$

for all pairs of real numbers (y_1, y_2).

If Y_1 and Y_2 are continuous random variables with joint density function $f(y_1, y_2)$ and marginal density functions $f_1(y_1)$ and $f_2(y_2)$, respectively, then Y_1 and Y_2 are independent if and only if

$$f(y_1, y_2) = f_1(y_1)f_2(y_2)$$

for all pairs of real numbers (y_1, y_2).

Proof: omitted

Independence : Example

EXAMPLE 5.11

Let

$$f(y_1, y_2) = \begin{cases} 6y_1y_2^2, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & elsewhere \end{cases}$$

Show that Y_1 and Y_2 are independent.

Independence : Example

EXAMPLE 5.12

Let

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

Show that Y_1 and Y_2 are dependent.

Checking for Independence

THEOREM 5.5

Let Y_1 and Y_2 have a joint density $f(y_1, y_2)$ that is positive if and only if $a \le y_1 \le b$ and $c \le y_2 \le d$, for constants a, b, c, and d; and $f(y_1, y_2) = 0$ otherwise. Then Y_1 and Y_2 are independent random variables if and only 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.

Proof: omitted

- The key benefit of the Theorem 5.5 is that we do not actually need to derive the marginal densities.
  - ✓ Indeed, the functions $g(y_1)$ and $h(y_2)$ need not, themselves, be density functions

Checking for Independence : Example

EXAMPLE 5.13

Let Y_1 and Y_2 have a joint density given by

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

Are Y_1 and Y_2 independent variables?

Motivation

$g(Y_1, Y_2, \ldots, Y_k)$: a function of **joint** random variables $Y_1, Y_2, \ldots, Y_k$

$$\Downarrow$$

$E[g(Y_1, Y_2, \ldots, Y_k)]$

Expected Value

DEFINITION 5.9

Let $g(Y_1, Y_2, \ldots, Y_k)$ be a function of the discrete random variables, $Y_1, Y_2, \ldots, Y_k$, which have probability function $p(y_1, y_2, \ldots, y_k)$. Then the *expected value* of $g(Y_1, Y_2, \ldots, Y_k)$ is

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \sum_{\text{all } y_k} \cdots \sum_{\text{all } y_2} \sum_{\text{all } y_1} g(y_1, y_2, \ldots, y_k) p(y_1, y_2, \ldots, y_k)$$

If $Y_1, Y_2, \ldots, Y_k$ are continuous random variables with joint density function $f(y_1, y_2, \ldots, y_k)$, then

$$E[g(Y_1, Y_2, \ldots, Y_k)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_k) \times f(y_1, y_2, \ldots, y_k) dy_1 dy_2 \ldots dy_k.$$

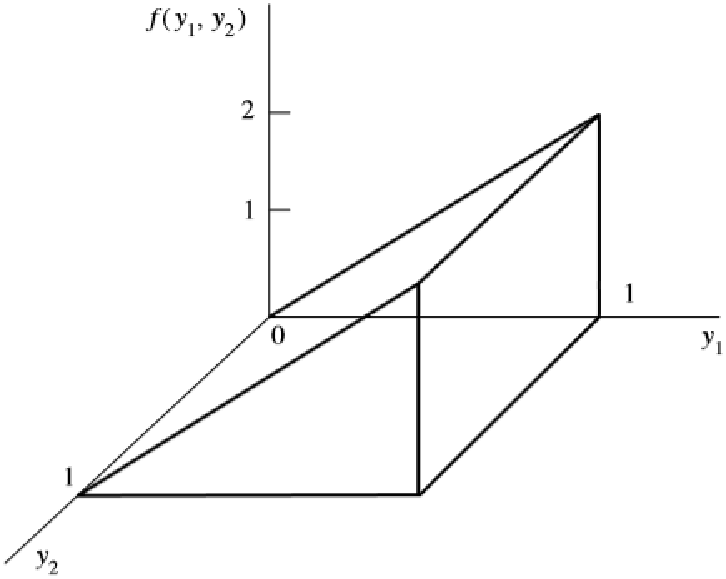
Expected Value : Example

EXAMPLE 5.15

Let Y_1 and Y_2 have a joint density given by

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

Find the expected value of $Y_1 Y_2$.



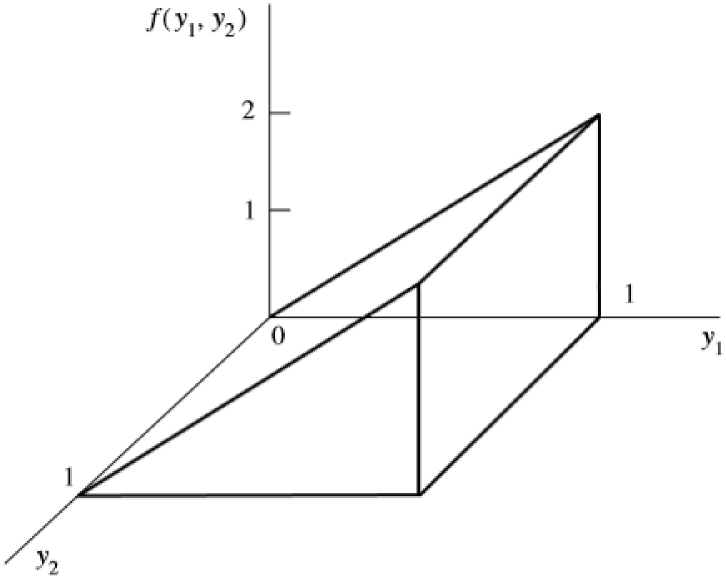
Expected Value : Example

EXAMPLE 5.16

Let Y_1 and Y_2 have a 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 the expected value of Y_1.



Expected Value : Example

EXAMPLE 5.19

A process for producing an industrial chemical yields a product containing two types of impurities. For a specified sample from this process, let Y_1 denote the proportion of impurities in the sample and let Y_2 denote the proportion of type I impurities among all impurities found. Suppose that the joint distribution of Y_1 and Y_2 can be modeled by the following probability density function:

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

Find the expected value of the proportion of type I impurities in the sample.

Theorems for Expectation

THEOREM 5.6

Let c be a constant. Then

$$E(c) = c.$$

THEOREM 5.7

Let $g(Y_1, Y_2)$ be a function of the random variables Y_1 and Y_2 and let c be a constant. Then

$$E[cg(Y_1, Y_2)] = cE[g(Y_1, Y_2)].$$

THEOREM 5.8

Let Y_1 and Y_2 be random variables and $g_1(Y_1, Y_2), g_2(Y_1, Y_2), \ldots, g_k(Y_1, Y_2)$ be functions of Y_1 and Y_2. Then

$$E[g_1(Y_1, Y_2) + g_2(Y_1, Y_2) + \cdots + g_k(Y_1, Y_2)] = E[g_1(Y_1, Y_2)] + E[g_2(Y_1, Y_2)] + \cdots + E[g_k(Y_1, Y_2)].$$

Theorems for Expectation

THEOREM 5.9

Let Y_1 and Y_2 be independent random variables and $g(Y_1)$ and $h(Y_2)$ be functions of only Y_1 and Y_2, respectively. Then

$$E[g(Y_1)h(Y_2)] = E[g(Y_1)]E[h(Y_2)],$$

provided that the expectations exist.

Proof:

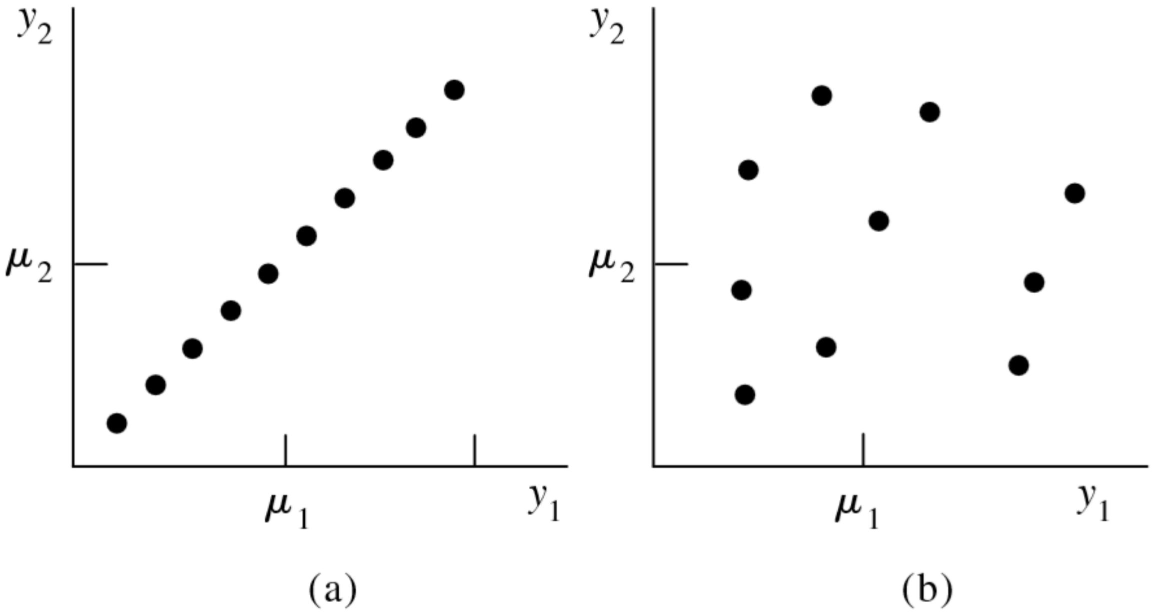
Theorems for Expectation : Example

EXAMPLE 5.21

Refer to Example 5.19. In that example we found $E(Y_1Y_2)$ directly. By investigating the form of the joint density function given there, we can see that Y_1 and Y_2 are independent. Find $E(Y_1Y_2)$ by using the result that $E(Y_1Y_2) = E(Y_1)E(Y_2)$ if Y_1 and Y_2 are independent.

Motivation

- We think of the dependence of two random variables Y_1 andY_2 as implying that one variable—say, Y_1—either increases or decreases as Y_2 changes



- The average value of $(Y_1 - \mu_1)(Y_2 - \mu_2)$ provides a measure of the linear dependence between Y_1 and Y_2.
  - ➢ The quantity, $E[(Y1\ - \mu1)(Y2\ - \mu2)]$, is called the covariance of Y_1 and Y_2.

Covariance

DEFINITION 5.10

If Y_1 and Y_2 are random variables with means μ_1 and μ_2, respectively, the *covariance* of Y_1 and Y_2 is

$$\mathrm{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] .$$

- The larger the absolute value of the covariance of $Y1$ and $Y2$, the greater the linear dependence between $Y1$ and $Y2$.
  - ✓ Positive values indicate that $Y1$ increases as $Y2$ increases;
  - ✓ negative values indicate that $Y1$ decreases as $Y2$ increases.
  - ✓ A zero value of the covariance indicates that the variables are uncorrelated and that there is no linear dependence between $Y1$ and $Y2$.

Correlation Coefficient

- it is difficult to employ the covariance as an absolute measure of dependence because its value depends upon the scale of measurement
  - *correlation coefficient*, ρ, a quantity related to the covariance and defined as

$$\rho = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2} \qquad (-1 \leq \rho \leq 1)$$

✓ $\rho = +1$: Perfect positive correlation
✓ $\rho = -1$: Perfect negative correlation
✓ $\rho = 0$: zerp correlation

Covariance

THEOREM 5.10

If Y_1 and Y_2 are random variables with means μ_1 and μ_2, respectively, then

$$\mathrm{Cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] = E(Y_1 Y_2) - E(Y_1)E(Y_2).$$

Proof:

Independence

THEOREM 5.11

If Y_1 and Y_2 are independent random variables, then

$$\mathrm{Cov}(Y_1, Y_2) = 0.$$

Thus, independent random variables must be uncorrelated.

Proof:

Covariance : Example

EXAMPLE 5.23

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 the covariance of Y_1 and Y_2.

Motivation

Linear function $U_1 = \sum_{i=1}^{n} a_i Y_i$ of random Variables $Y_1, Y_2, \ldots, Y_n$

Linear function $U_2 = \sum_{j=1}^{m} b_i X_j$ of random Variables $X_1, X_2, \ldots, X_m$

- What is the expectation $E(U_1)$ of U_1 (a linear function of random variables)?
- What is the variance $\mathrm{Var}(U_1)$ of U_1 (a linear function of random variables)?
- What is the covariance $\mathrm{Cov}(U_1, U_2)$?

Expected Value of Linear functions of r.v.

THEOREM 5.12

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 $E\left(X_j\right) = \xi_j$. 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 $E(U_1) = \Sigma_{i=1}^{n} a_i \mu_i$.

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

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

Proof:

Expected Value of Linear functions of r.v. : Example

EXAMPLE 5.27

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$.
(These variables may denote the outcomes of n independent trials of an experiment.)
Define

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

and show that $E(\bar{Y}) = \mu$ and $V(\bar{Y}) = \sigma^2/n$.

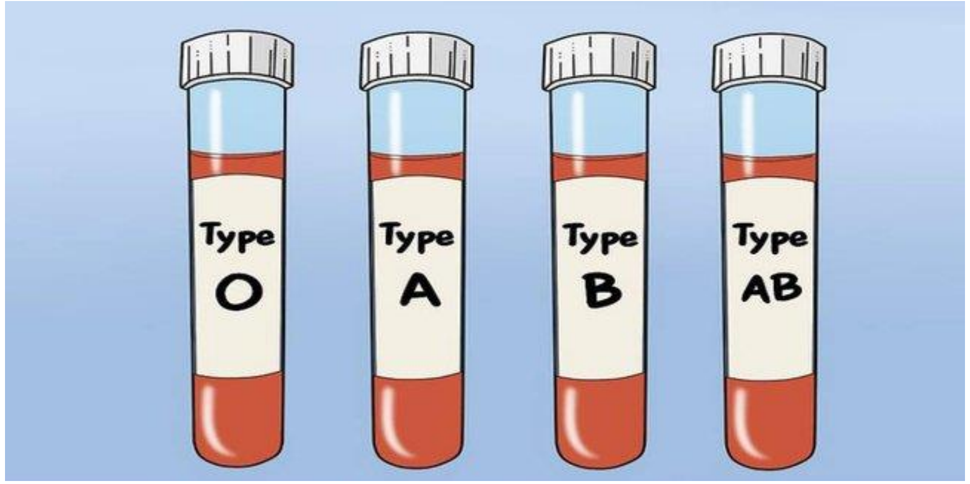
Expected Value of Linear functions of r.v. : Example

EXAMPLE 5.28

The number of defectives Y in a sample of $n = 10$ items selected from a manufacturing process follows a binomial probability distribution. An estimator of the fraction defective in the lot is the random variable $\hat{p} = Y/n$. Find the expected value and variance of $\hat{p}$.

Motivation

- Recall from Chapter 3 that a binomial random variable results from an experiment consisting of n trials with two possible outcomes per trial.
- Frequently we encounter similar situations in which the number of possible outcomes per trial is more than two. For example,
  - ✓ Experiments that involve blood typing typically have at least four possible outcomes per trial.
  - ✓ Experiments that involve sampling for defectives may categorize the type of defects observed into more than two classes.



Multinomial Experiment

DEFINITION 5.11

A *multinomial experiment* possesses the following properties:

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k classes or cells.
3. The probability that the outcome of a single trial falls into cell i, is p_i, $\quad i = 1, 2, \ldots, k$ and remains the same from trial to trial. Notice that

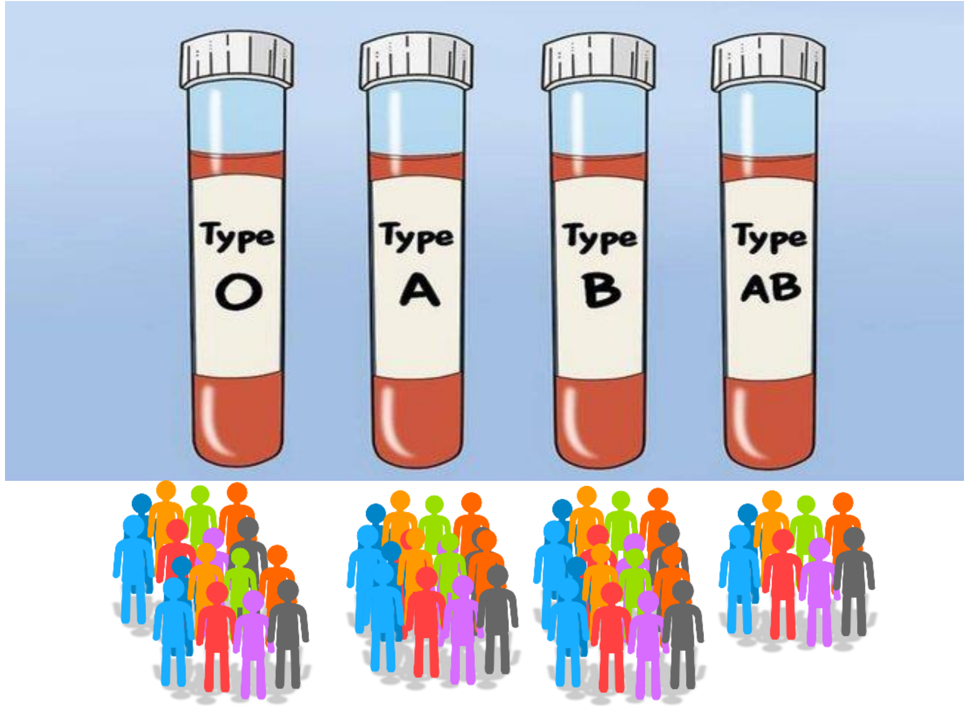
$$p_1 + p_2 + p_3 + \cdots + p_k = 1.$$

4. The trials are independent.
5. The random variables of interest are $Y_1, Y_2, \ldots, Y_k$, where Y_i equals the number of trials for which the outcome falls into cell i. Notice that

$$Y_1 + Y_2 + Y_3 + \cdots + Y_k = n.$$

Multinomial Experiment

p_1 p_2 p_3 p_4



Y_1 Y_2 Y_3 Y_4

Multinomial Distribution

DEFINITION 5.12

Assume that $p_1, p_2, \ldots, p_k$ are such that $\Sigma_{i=1}^{k} p_i = 1$, and $p_i > 0$ for $i = 1, 2, \ldots, k$. The random variables $Y_1, Y_2, \ldots, Y_k$, are said to have a *multinomial distribution* with parameters n and $p_1, p_2, \ldots, p_k$ if the joint probability function of $Y_1, Y_2, \ldots, Y_k$ is given by

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1!\, y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k},$$

where, for each i, $y_i = 0, 1, 2, \ldots, n$ and $\Sigma_{i=1}^{k} y_i = n$.

Multinomial Distribution : Example

EXAMPLE 5.30

According to recent census figures, the proportions of adults (persons over 18 years of age) in the United States associated with five age categories are as given in the following table.

Age	Proportion
18–24	.18
25–34	.23
35–44	.16
45–64	.27
65↑	.16

If these figures are accurate and five adults are randomly sampled, find the probability that the sample contains one person between the ages of 18 and 24, two between the ages of 25 and 34, and two between the ages of 45 and 64.

Multinomial Distribution : Example

SOLUTION 5.30

Mean and Variance of Multinomial Distribution

THEOREM 5.13

If $Y_1, Y_2, \ldots, Y_k$ have a multinomial distribution with parameters n and $p_1, p_2, \ldots, p_k$, then

1. $E(Y_i) = np_i, V(Y_i) = np_i q_i$.

2. $Cov(Y_s, Y_t) = -np_s p_t$, if $s \neq t$.

Proof:

Motivation

DEFINITION 4.8 (Univariate Normal Distribution)

A random variable Y is said to have a *normal probability distribution* if and only if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of Y is

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(y-\mu)^2}{(2\sigma^2)}\right], \qquad -\infty < y < \infty.$$

- For $k = 2$

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{1}{1-\rho^2}\left\{\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)(y_1-\mu_1)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right\}\right]$$

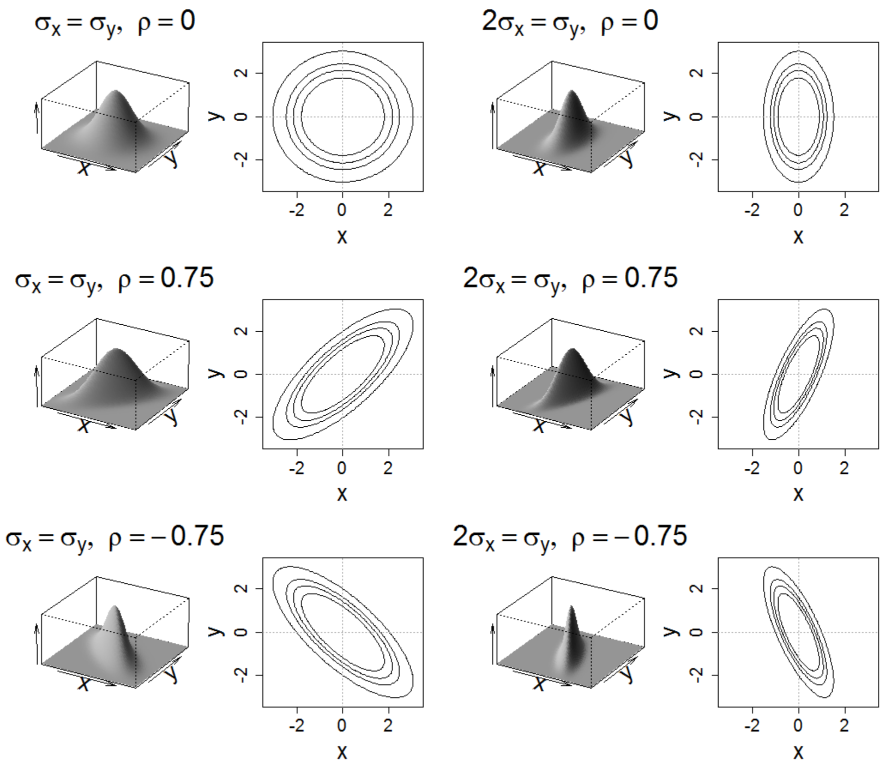
$$-\infty < y_1 < \infty, \qquad -\infty < y_2 < \infty$$

Bivariate Normal Distribution

- For $k = 2$

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{1}{1-\rho^2}\left\{\frac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(y_1-\mu_1)(y_1-\mu_1)}{\sigma_1\sigma_2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2}\right\}\right]$$

$$-\infty < y_1 < \infty, \qquad -\infty < y_2 < \infty$$



- If $\mathrm{Cov}(Y_1, Y_2) = 0$ or, equivalently, if $\rho = 0$
  - then $f(y1, y2) = g(y_1)h(y_2)$,
  - $g(y_1)$ is a nonnegative function of y_1 alone and $h(y_2)$ is a nonnegative function of y_2 alone.
    - ➢ Therefore, if $\rho = 0$, Theorem 5.5 implies that Y_1 and Y_2 are independent.

Multivariate Normal Distribution (Matrix Notation)

- **Univariate Gaussian**

$$N(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Mean: $\mu=E[X]$

Variance: $\sigma^2=\mathrm{var}(\mathrm{X})=\mathrm{E}[(X-E[X])^2]$

- **Multivariate Gaussian**

$$N(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Mean vector

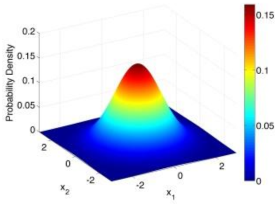
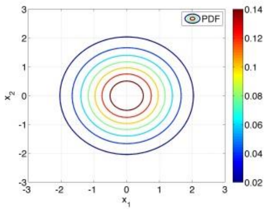
$$\boldsymbol{\mu}=\begin{bmatrix}E[X_1]\\ \vdots\\ E[X_n]\end{bmatrix}$$

Covariance matrix

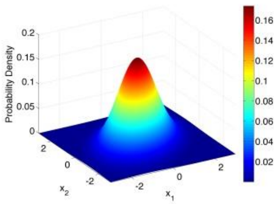
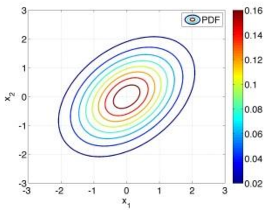
$$\boldsymbol{\Sigma}=\begin{bmatrix}\mathrm{Cov}[X_1,X_1] & \cdots & \mathrm{Cov}[X_1,X_n]\\ \vdots & \ddots & \vdots\\ \mathrm{Cov}[X_n,X_1] & \cdots & \mathrm{Cov}[X_n,X_n]\end{bmatrix}$$

$$\mathrm{Cov}[X,Z]=E[(X-E[X])(Z-E[Z])]$$

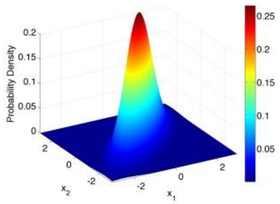
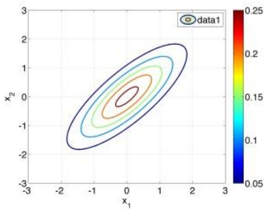
$\boldsymbol{\mu}=\begin{bmatrix}0\\0\end{bmatrix}$ $\boldsymbol{\Sigma}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$

$\boldsymbol{\mu}=\begin{bmatrix}0\\0\end{bmatrix}$ $\boldsymbol{\Sigma}=\begin{bmatrix}1&0.4\\0.4&1\end{bmatrix}$

$\boldsymbol{\mu}=\begin{bmatrix}0\\0\end{bmatrix}$ $\boldsymbol{\Sigma}=\begin{bmatrix}1&0.8\\0.8&1\end{bmatrix}$

Motivation

- **Unconditional Expectation:**

$$E(g(Y_1)) = \sum_{\text{all } y_1} g(y_1)p(y_1)$$

$$E(g(Y_1)) = \int_{-\infty}^{\infty} g(y_1)f(y_1)dy_1$$

- **Conditional Expectation:**

$$E(g(Y_1)|Y_2 = y_2) = \sum_{\text{all } y_1} g(y_1)p(y_1|y_2)$$

$$E(g(Y_1)|Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1)f(y_1|y_2)dy_1$$

Conditional Expectations

DEFINITION 5.13

if Y_1 and Y_2 are jointly **continuous random variables**, the *conditional expectation* of $g(Y_1)$, given that $Y_2 = y_2$, is defined to be

$$E(g(Y_1)|Y_2 = y_2) = \int_{-\infty}^{\infty} g(y_1) f(y_1|y_2) dy_1$$

if Y_1 and Y_2 are jointly **discrete random variables**, the *conditional expectation* of $g(Y_1)$, given that $Y_2 = y_2$, is defined to be

$$E(g(Y_1)|Y_2 = y_2) = \sum_{\text{all } y_1} g(y_1) p(y_1|y_2)$$

Conditional Expectation

THEOREM 5.14

Let Y_1 and Y_2 denote random variables. Then

$$E(Y_1) = E[E(Y_1|Y_2)],$$

where on the right-hand side the inside expectation is with respect to the conditional distribution of Y_1 given Y_2 and the outside expectation is with respect to the distribution of Y_2.

Proof:

Conditional Variance

THEOREM 5.15

Let Y_1 and Y_2 denote random variables. Then

$$V(Y_1) = E[V(Y_1|Y_2)] + V[E(Y_1|Y_2)].$$

Proof:

Conditional Variance : Example

EXAMPLE 5.33

Refer to Example 5.32. Find the variance of Y.