[Chapter 5. Multivariate Probability Distributions]

5.1 Introduction

Suppose that $Y_1, Y_2, \ldots, Y_n$ denote the outcomes of n successive trials of an experiment. A specific set of outcomes, or sample measurements, may be expressed in terms of the intersection of n events

$$(Y_1 = y_1), (Y_2 = y_2), \ldots, (Y_n = y_n)$$

which we will denote as

$$(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

or more compactly, as

$$(y_1, y_2, \ldots, y_n).$$

Calculation of the probability of this intersection is essential in making inferences about the population from which the sample was drawn and is a major reason for studying multivariate probability distributions.

5.2 Bivariate and Multivariate probability distributions

Many random variables can be defined over the same sample space.

(Example) Tossing a pair of dice.
The sample space contains 36 sample points. Let Y_1 be the number of dots appearing on die 1, and Y_2 be the sum of the number of dots on the dice. We would like to obtain the probability of $(Y_1 = y_1, Y_2 = y_2)$ for all the possible values of y_1 and y_2. That is the joint distribution of Y_1 and Y_2.

(Def 5.2) For any r.v. Y_1 and Y_2 the joint (bivariate) distribution function $F(y_1, y_2)$ is given by

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

for $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$.

(Theorem 5.2) If Y_1 and Y_2 are r.v. 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 $a_1^\star \geq a_1$ and $b_2^\star \geq b_2$, then

$$F(a_1^\star, b_2^\star) - F(a_1^\star, b_2) - F(a_1, b_2^\star) + F(a_1, b_2) \\ = P(a_1 < Y_1 \leq a_1^\star, b_2 < Y_2 \leq b_2^\star) \geq 0.$$

(1) Discrete variables:

(Def 5.1) Let Y_1 and Y_2 be discrete r.v. The *joint probability distribution* for Y_1 and Y_2 is given by

$$p(y_1, y_2) = p(Y_1 = y_1, Y_2 = y_2)$$

for $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$. The function $p(y_1, y_2)$ will be referred to as the joint probability function.

Note that if Y_1 and Y_2 are discrete r.v. with joint probability function $p(y_1, y_2)$, its CDF is

$$\begin{aligned} F(y_1, y_2) &= P(Y_1 \le y_1, Y_2 \le y_2) \\ &= \sum_{t_1 \le y_1} \sum_{t_2 \le y_2} p(t_1, t_2) \end{aligned}$$

(Theorem 5.1) If Y_1 and Y_2 are discrete r.v. with joint probability function $p(y_1, y_2)$, then

1. $p(y_1, y_2) \ge 0$ for all y_1, y_2.
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$, where the sum is over all values (y_1, y_2) that are assigned nonzero probabilities.
3. $P\left[(y_1, y_2) \in A\right] = \sum_{(y_1, y_2) \in A} p(y_1, y_2)$ for $A \subseteq S$. So,

$$P(a_1 \le Y_1 \le a_2, b_1 \le Y_2 \le b_2) = \sum_{t_1 = a_1}^{a_2} \sum_{t_2 = b_1}^{b_2} p(t_1, t_2)$$

(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 distribution of Y_1 and Y_2.

(Example 5.2) Consider the Y_1 and Y_2 in (Example 5.1). Find $F(-1, 2)$, $F(1.5, 2)$ and $F(5, 7)$.

(2) Continuous variables:

Two random variables are said to be jointly continuous if their joint distribution function $F(y_1, y_2)$ is continuous in both arguments.

(Def 5.3) Let Y_1 and Y_2 be continuous r.v. with joint distribution function $F(y_1, y_2)$. If there exists a nonnegative function $f(y_1, y_2)$ such that

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

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

(Theorem 5.3) If Y_1 and Y_2 are jointly continuous random variables with a joint density function $f(y_1, y_2)$, then

1. $f(y_1, y_2) \geq 0$ for all y_1, y_2.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$.
3. $p\left[(y_1, y_2) \in A\right] = \int \int_A f(y_1, y_2) dy_2 dy_1$. So,

$$P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2) = \int_{b_1}^{b_2} \int_{a_1}^{a_2} f(y_1, y_2) dy_1 dy_2$$

(Example 5.3)

(Example 5.4)

(Exercise 5.5)

(Exercise 5.9)

(Question) How about the case of the intersection of n events

$$(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)?$$

For discrete r.v.,
the probability function is given by

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

and its joint distribution function is given by

$$\begin{aligned} F(y_1, y_2, \ldots, y_n) &= P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) \\ &= \sum_{t_1 \le y_1} \sum_{t_2 \le y_2} \sum_{t_n \le y_n} p(y_1, y_2, \ldots, y_n). \end{aligned}$$

For continuous r.v.,
the joint distribution function is given by

$$P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) = F(y_1, \ldots, y_n)$$
$$= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n) dt_1 \ldots dt_n$$

for every set of real numbers $(y_1, y_2, \ldots, y_n)$ and its joint density is given by

$$f(y_1, y_2, \ldots, y_n).$$

5.3 Marginal and Conditional probability distributions

Given Theorem 5.1 and Definition 5.2,

[Discrete random variables]

(Def 5.4)
a. Let Y_1 and Y_2 be jointly discrete r.v. with probability function $p(y_1, y_2)$. Then the *marginal probability functions* of Y_1 and Y_2 are given by

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2), \quad p_2(y_2) = \sum_{y_1} p(y_1, y_2).$$

(Def 5.5)
If Y_1 and Y_2 are jointly discrete r.v. 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

$$\begin{aligned} p(y_1 \mid y_2) &= P(Y_1 = y_1 \mid Y_2 = y_2) \\ &= \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)} \end{aligned}$$

provided that $p_2(y_2) > 0$.

Note that

1. Multiplicative law(Theorem 2.5(p.55)): $P(A \cap B) = P(A)P(B \mid A)$.

2. 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). Then, the bivariate probability for (y_1, y_2) is

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

3. $p(y_1 \mid y_2)$: the probability that the r.v. Y_1 equals y_1, given that that Y_2 takes on the value y_2.

4. $p(y_1 \mid y_2)$ is undefined if $p_2(y_2) = 0$.

(Example 5.5, 5.7)

(Example) Contracts for two construction jobs are randomly assigned to one or more of three firms A, B and C. Let Y_1 and Y_2 be the number of contracts assigned to firm A and B, respectively. Recall that each firm can receives 0, 1, or 2 contracts.

a. Find the joint probability distribution for Y_1 and Y_2.

b. Calculate $F(1, 0)$, $F(3, 4)$ and $F(1.5, 1.6)$

c. Find the marginal probability distribution of Y_1 and Y_2.

d. Find the conditional probability function for Y_2 given $Y_1 = 1$.

e. Find the conditional probability function for Y_2 given $Y_1 = 0$.

Given Definition 5.3 and Theorem 5.3,

[Continuous random variables]

(Def 5.4)
b. Let Y_1 and Y_2 be jointly continuous r.v. with probability function $f(y_1, y_2)$. Then the *marginal density functions* of Y_1 and Y_2 are given by

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

For continuous Y_1 and Y_2, $P(Y_1 = y_1 \mid Y_2 = y_2)$ can not be defined as in the discrete case, because both $(Y_1 = y_1)$ and $(Y_2 = y_2)$ are events with zero probability.

(Def 5.6)
If Y_1 and Y_2 are jointly continuous r.v. with joint density function $f(y_1, y_2)$, then the conditional distribution function of Y_1 given $Y_2 = y_2$ is

$$F(y_1 \mid y_2) = P(Y_1 \leq y_1 \mid Y_2 = y_2) = \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_2(y_2)} dt_1$$

Note that one can derive conditional density function of Y_1 given $Y_2 = y_2$, $f(y_1 \mid y_2)$ from the calculation of $F(y_1)$:

(Def 5.7)
If Y_1 and Y_2 are jointly continuous r.v. with joint density function $f(y_1, y_2)$ 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

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

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

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

Note that i) $f(y_1 \mid y_2)$ is undefined for all y_2 such that $f_2(y_2) = 0$, ii)$f(y_2 \mid y_1)$ is undefined for all y_1 such that $f_1(y_1) = 0$.

(Example 5.8)

(Example) Let Y_1 and Y_2 have joint probability density function(pdf) given by

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

a. Find the value of k such that this is a pdf.

b. Calculate $P(Y_1 \le 3/4, Y_2 \ge 1/2)$

c. Find the marginal density function of Y_1 and Y_2.

d. Calculate $P(Y_1 \le 1/2 \mid Y_2 \le 3/4)$

e. Find the conditional density function of Y_1 given Y_2.

f. Find the conditional density function of Y_2 given Y_1.

g. Calculate $P(Y_2 \ge 3/4 \mid Y_1 = 1/2)$

5.4 Independent random variables

- Independent random variables :
  Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
  Suppose we are concerned with events of the type $(a \le Y_1 \le b) \cap (c \le Y_2 \le d)$. If Y_1 and Y_2 are independent, does the following equation hold?

$$P(a \le Y_1 \le b, c \le Y_2 \le d) = P(a \le Y_1 \le b)P(c \le Y_2 \le d)$$

(Def 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*.

Extension of (Def 5.8) to n dimensions :

Suppose we have n random variables, $Y_1, \ldots, Y_n$, where Y_i has distribution function $F_i(y_i)$, for $i = 1, 2, \ldots, n$; and where $Y_1, \ldots, Y_n$ have joint distribution $F(y_1, y_2, \ldots, y_n)$.
Then $Y_1, \ldots, Y_n$ are independent if and only if

$$F(y_1, y_2, \ldots, y_n) = F_1(y_1) \cdots F_n(y_n)$$

for all real numbers $y_1, y_2, \ldots, y_n$.

(Theorem 5.4)

- Discrete r.v. : If Y_1 and Y_2 are discrete r.v. 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 pair of real numbers (y_1, y_2).

(Theorem 5.4)

• Continuous r.v. : If Y_1 and Y_2 are continuous r.v. 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 pair of real numbers (y_1, y_2).

(Example 5.10)

(Example 5.12)

The key benefit of the following theorem 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.

(Theorem 5.5) If 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 r.v. 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.

(Example 5.13)

(Exercise 5.61)

5.5 Expected value of a function of r.v.

(Def 5.9)

• Discrete r.v. : Let $g(Y_1, Y_2, \dots, Y_k)$ a function of the discrete r.v., $Y_1, Y_2, \dots, Y_k$, which have probability function $p(y_1, y_2, \dots, y_k)$. Then the *expected value* of $g(Y_1, Y_2, \dots, Y_k)$ is

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

• Continuous r.v. : If $Y_1, \dots, Y_k$ are continuous r.v. with joint density function $f(y_1, \dots, y_k)$, then

$$E[g(Y_1, \dots, Y_k)] \\ = \int_{y_k} \cdots \int_{y_1} g(y_1, \dots, y_k) f(y_1, \dots, y_k) dy_1 \dots dy_k.$$

• Derivation of $E(Y_1)$ from (Def 5.9)(and Def 4.4).

(Example 5.15)

(Example 5.16)

5.6 Special theorems

(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 r.v. 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,Y_2 be r.v. 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

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

(Example 5.20)

(Theorem 5.9) Let Y_1 and Y_2 be independent r.v. 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 expectation exist.

(Proof)

(Example 5.21)

(Exercise 5.77)

5.7 Covariance of two random variables

The covariance between Y_1 and Y_2 is a measure of the linear dependence between them.

(Def 5.10) If Y_1 and Y_2 are r.v. with means $\mu_1 = E(Y_1)$ and $\mu_2 = E(Y_2)$, respectively, the *covariance* of Y_1 and Y_2 is given by

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

- $-\infty < Cov(Y_1, Y_2) < \infty$
- The larger the absolute value of the covariance of Y_1 and Y_2, the greater the *linear dependence* between Y_1 and Y_2.
- Positive values indicate that $Y_1(Y_2)$ increases as $Y_2(Y_1)$ increases.
- Negative values indicate that $Y_1(Y_2)$ decreases as $Y_2(Y_1)$ increases.
- A zero value indicates no linear dependence between Y_1 and Y_2.
- If $Y_1 = Y_2$, $Cov(Y_1, Y_2) =$

It is difficult to employ the covariance as an absolute measure of dependence because its value depends upon the scale of measurement. So, it is difficult to determine whether a particular covariance is large or small.

The *correlation coefficient*, ρ is defined as

$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_1 \sigma_2}$$

where $\sigma_1 = \sqrt{Var(Y_1)}$ and $\sigma_2 = \sqrt{Var(Y_2)}$ are the standard deviations of Y_1 and Y_2.

- ρ also measures the linear dependence between Y_1 and Y_2 and its sign is the same as the sign of the covariance. But, $-1 < \rho < 1$.

- $\rho > 0$: $Y_1(Y_2)$ increases as $Y_2(Y_1)$ increases.

- $\rho < 0$: $Y_1(Y_2)$ decreases as $Y_2(Y_1)$ increases.

- $\rho = 0$: no correlation between Y_1 and Y_2.

- $\rho = +1(-1)$: perfect linear correlation between Y_1 and Y_2, with all points on a straight line with positive(negative) slope.

(Theorem 5.10) Let Y_1 and Y_2 are r.v. with means μ_1 and μ_2, respectively, then

$$Cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] = E(Y_1 Y_2) - \mu_1 \mu_2$$

(Proof)

(Example 5.23)

(Theorem 5.11) Let Y_1 and Y_2 are independent r.v., then $Cov(Y_1, Y_2) = 0$.

(Proof)

[Note] *Zero covariance DOES NOT IMPLY the independence between two random variables.* BUT, one exception : *If Y_1 and Y_2 are NORMAL random variables and they are uncorrelated(i.e., they have zero covariance), they are also independent(See Chapter 5.10).*

(Example 5.24)

(Exercise)

5.8 Expected value and variance of linear combinations of r.v.'s

Let U_1 be a linear function of the r.v.'s $Y_1, Y_2, \ldots, Y_n$. Then, we have

$$U_1 = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n = \sum_{i=1}^{n} a_i Y_i$$

for constants $a_1, a_2, \ldots, a_n$.

We may be interested in the moments(expected values, covariance,...) of such linear combinations.

(Theorem 5.12) Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be r.v. with $E(Y_i) = \mu_i$ and $E(X_j) = \xi_j$. 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 $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$.

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

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

(Example 5.25)

(Example 5.27)

5.9 Multinomial probability distribution

From Chapter 3, we know that a binomial r.v. results from an experiment consisting of n trials with two possible outcomes per trial. If one encounter the situations in which the number of possible outcomes per trial is more than two, one needs another concept.

A multinomial experiment is a generalization of the binomial experiment.

(Def 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, $i = 1, 2, \ldots, k$

and remains the same from trial to trial. Notice that $p_1 + p_2 + \cdots + p_k = 1$.

4. The trials are independent.

5. The r.v.'s 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 $Y_1 + Y_2 + \ldots + Y_k = n$.

(Def 5.12) Assume that $p_1, p_2, \ldots, p_k$ are such that $\sum_{i=1}^{k} p_i = 1$, and $p_i > 0$ for $i = 1, 2, \ldots, k$. The r.v.'s $Y_1, Y_2, \ldots, Y_k$ 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 $i, y_i = 0, 1, \ldots, n$ and $\sum_{i=1}^{k} y_i = n$.

Note that $k = 2$ provides the binomial experiment/distribution.

(Example 5.30)

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

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

(Example) Suppose that a fair die is rolled 9 times. Let Y_i be the number of trials for which number i appears.

(a) What is the probability that 1 appears three times, 2 and 3 twice each, 4 and 5 once each, and 6 not at all?

(b) Find $E(Y_i)$, $V(Y_i)$ and $Cov(Y_s, Y_t)$ where $i, s, t = 1, \ldots, 6$ and $s \neq t$.

5.10 Bivariate normal distribution

The bivariate density function for two normal r.v.'s Y_1 and Y_2 is

$$f(y_1, y_2) = \frac{e^{-Q/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

for $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$ where

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

- ρ is the correlation between Y_1 and Y_2.
- Marginal distributions of Y_1 and Y_2.
- Suppose Y_1 and Y_2 are independent($\rho = 0$), and they have marginal normal distributions. Then (Y_1, Y_2) is a bivariate normal with $\rho = 0$.
  We know that if two variables are independent, then they are (linearly) uncorrelated. But, in general the following is NOT true: *if they are (linearly) uncorrelated, they are independent.*
  However, for normal Y_1 and Y_2, they are also independent if they are uncorrelated.

5.11 Conditional expectations

Conditional expectations are related to conditional probability/density functions discussed in Section 5.3.

(Def 5.13) If Y_1 and Y_2 are any two r.v.'s, the *conditional expectation* of $g(Y_1)$ given that $Y_2 = y_2$ is defined to be

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

if Y_1 and Y_2 are jointly continuous and

$$E(g(Y_1) \mid Y_2 = y_2) = \sum_{y_1} g(y_1) p(y_1 \mid y_2)$$

if Y_1 and Y_2 are jointly discrete.

(Example 5.31)

In general, the conditional expectation of Y_1 given $Y_2 = y_2$ is a function of y_2. If we let Y_2 range over all of its possible values, we can think of the conditional expectation $E(Y_1 \mid Y_2)$ as a function of the r.v. Y_2.

Because $E(Y_1 \mid Y_2)$ is a function of the r.v. Y_2, it is itself a random variable; and as such, it has a mean and a variance.

(Theorem 5.14) Let Y_1 and Y_2 denote random variables. Then

$$E(Y_1) = E[E(Y_1 \mid Y_2)]$$

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)

Note that

- The conditional variance of Y_1 given $Y_2 = y_2$ is defined by

$$V(Y_1 \mid Y_2 = y_2) \\ = E(Y_1^2 \mid Y_2 = y_2) - [E(Y_1 \mid Y_2 = y_2)]^2.$$

- The conditional variance is a function of y_2.

- Letting Y_2 range over all of its possible values, $V(Y_1 \mid Y_2)$ is a random variable that is a function of Y_2.

(Theorem 5.15)
Let Y_1 and Y_2 denote random variables. Then

$$V(Y_1) = E[V(Y_1 \mid Y_2)] + V[E(Y_1 \mid Y_2)]$$

(Proof)

(Example 5.32) A quality control plan for an assembly line involves sampling $n = 10$ finished items per day and counting Y, the number of defectives. If p denotes the probability of observing a defective, then Y has a binomial distribution, assuming that a large number of items are produced by the line. But p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to 1/4. Find expected value and the variance of Y.