36225
Probability

ZhenTing Liu

Contents

Chapter 1

Multivariate Distribution

1.1 Multivariate distribution

Definition 1.1.1

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

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

In the single-variable case , the probability function for a discrete random variable Y assigns nonzero probabilities to a finite or countable number of distinct values of Y in such a way that the sum of the probabilities is equal to 1. Similarly, in the bivariate case the joint probability function $p(y_1, y_2)$ assigns nonzero probabilities to only a finite or countable number of pairs of values (y_1, y_2). Further, the nonzero probabilities must sum to 1.

Theorem 1.1.1 Discrete random variables

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

1. $p(y_1, y_2) \geqslant 0 \quad \forall y_1, y_2$
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$

As in the case of univariate random variables, the distinction between jointly discrete and jointly continuous random variables may be characterized in terms of their (joint) distribution functions.

Definition 1.1.2

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

$$F(y_1, y_2) = P(Y_1 \leqslant y_1, Y_2 \leqslant y_2) \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

Note:-

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. $F(\infty, \infty) = 1$

Theorem 1.1.2 Continous random variables

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

1.2 Marginal and Conditional distribution

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.

Definition 1.2.1: Marginal Distribution

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

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

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

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

The term **marginal**, as applied to the univariate probability functions of Y_1 and Y_2 , has intuitive meaning. To find $p_1(y_1)$, we sum $p(y_1, y_2)$ over all values of y_2, and hence accumulate the probabilities on the y_1 axis (or margin).

Note:-

To find marginal pdf

1. integrate over the other variable
2. fix a line corresponding to a value of desired marginal pdf
3. find the limit of integral: intersecting points between this line and support of the joint pdf
4. check if the boundary of support is straight line
5. remember to specify the support of the resulting marginal pdf

Definition 1.2.2: Conditional Distribution discrete case

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

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

Definition 1.2.3: Conditional Distribution cont. case

If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$ respectively, then the conditional distribution function of Y_1 given $Y_2 = y_2$ is

$$\begin{aligned} F(y_1|y_2) &= P(Y_1 \leqslant y_1 | Y_2 = y_2) \\ &= \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_2(y_2)} dt_1 \end{aligned}$$

We will call the integrand of this expression the conditional density function of Y_1 given $Y_2 = y_2$ and we will denote it by $f(y_1|y_2)$.
Note for any $f_2(y_2) > 0$, 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 $f_1(y_1) > 0$, the conditional density of Y_2 given $Y_1 = y_1$ is gievn by

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

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

Note:-

$$P(a < Y_1 < b | Y_2 = y_2) = \int_a^b f(y_1|y_2) dy_1$$

also for a valid pmf/pdf for Y_1 or Y_2 as univariate r.v.

1. $f(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)} \geqslant 0$
2. $\int_{-\infty}^{\infty} f(y_1|y_2) dy_1 = \int_{-\infty}^{\infty} \frac{f(y_1,y_2)}{f_2(y_2)} dy_1 = 1$

1.2.1 Examples

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2, & y_1 - 1 \leqslant y_2 \leqslant 1 - y_1,\ 0 \leqslant y_1 \leqslant 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Question 1:

a show that the marginal density of Y_1 is a beta density with $\alpha = 2$ and $\beta = 4$
b Derive the marginal density of Y_2
c Derive the conditional density of Y_2 given $Y_1 = y_1$
d Find $P(Y_2 > 0 | Y_1 = 0.75)$

Solution:
a

$$f_1(y_1) = \begin{cases} \int_{y_1-1}^{1-y_1} 30y_1y_2^2 dy_2, & 0 \leqslant y_1 \leqslant 1 \\ 0, & \text{elsewhere.} \end{cases}$$

$\sim$ Beta(2,4) after integration
b

$$f_2(y_2) = \begin{cases} \int_0^{1-y_2} 30y_1y_2^2 dy_1, & 0 \leqslant y_2 \leqslant 1 \\ \int_0^{y_2+1} 30y_1y_2^2 dy_1, & -1 \leqslant y_2 \leqslant 0 \\ 0, & \text{elsewhere.} \end{cases}$$

c

$$f(y_2|y_1) = \frac{f(y_1, y_1)}{f_1(y_1)} = \frac{30y_1y_2^2}{20y_1(1-y_1)^3} = \frac{3y_2^2}{2(1-y_1)^3} \quad y_1 - 1 \leqslant y_2 \leqslant 1 - y_1, 0 \leqslant y_1 \leqslant 1$$

d

$$\int_0^{(1-0.75)} \frac{3y_2^2}{2(1-0.75)^3} dy_2 = \frac{1}{2}$$

1.3 Independence

Definition 1.3.1: The r.v. Y_1, Y_2 are independent if and only if

discrete:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2) \quad \forall y_1, y_2 \in \mathbb{R}^2$$

coninuous:

$$f(y_1, y_2) = f_1(y_1)f_2(y_2) \quad \forall y_1, y_2 \in \mathbb{R}^2$$

CDF:

$$F(y_1, y_2) = F_1(y_1)F_2(y_2) \quad \forall y_1, y_2 \in \mathbb{R}^2$$

Theorem 1.3.1 Determining independence

Given two r.v.'s Y_1, Y_2 Step 1: is the support of joint pmf/pdf rectangular with sides parallel to the axis?

- "No" $\implies$ Not independent
- "Yes" $\implies$ Move to step 2

Step 2: Can the joint pmf/pdf be written as the product of the form $p(y_1, y_2) = h(y_a)g(y_b), a, b \in \{1, 2\}$

- "No" $\implies$ Not independent
- "Yes" $\implies$ Independent

tip: To rigorously prove "cannot be factorized", one way is to find one pair of $(y_1, y_2) \in \mathbb{R}^2$ such that

$$f_1(y_1)f_2(y_2) \neq f(y_1, y_2)$$

1.4 Expected Value

Definition 1.4.1

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

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

Note:-

In general, $E(Y_1Y_2) \neq E(Y_1)E(Y_2)$ except for when Y_1 and Y_2 are independent
Important: Independence $\Longrightarrow E(Y_1Y_2) = E(Y_1)E(Y_2)$ but $E(Y_1Y_2) = E(Y_1)E(Y_2) \not\Longrightarrow$ Independence

Consider two random variables Y_1, Y_2 with density function $f(y_1, y_2)$.

$$E(Y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1, y_2) dy_2 dy_1$$

1.5 Covariance

The covariance between two random variables Y_1 and Y_2, a measure of dependence between Y_1, Y_2, is defined as

$$cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] \quad \mu_i = E(Y_i) \quad i = 1, 2$$

which can also be written in shortcut form as

$$cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$$

note if $Y_1 = Y_2 = Y$ then $cov(Y, Y) = V(Y)$

Corollary 1.5.1
if Y_1, Y_2 are independent, then $cov(Y_1, Y_2) = 0$ because

$$E(Y_1Y_2) = E(Y_1)E(Y_2)$$

but $cov(Y_1Y_2) = 0$ does not imply independence

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 increases as Y_2 increases; negative values indicate that Y_1 decreases as Y_2 increases. A zero value of the covariance indicates that the variables are uncorrelated and that there is no linear dependence between Y_1 and Y_2

1.5.1 Correlation Coefficient

correlation coefficient ρ is a measurement of dependency and is defined as

$$\rho = \frac{Cov(Y_1Y_2)}{\sigma_1\sigma_2}$$

where σ_1, σ_2 are the standard deviation of Y_1, Y_2 respectively.
Note

$$-1 \leqslant \rho_{Y_1Y_2} \leqslant 1$$

- $\rho = 0 \iff Y_1, Y_2$ uncorrelated
- $\rho = 1 \iff$ Perfect Positive linear dependence $Y_1 = aY_2 + b, \quad a > 0$
- $\rho = -1 \iff$ Perfect Negative linear dependence $Y_1 = -aY_2 + b, \quad a > 0$

1.6 Linear Combination, Conditional Expectation

Theorem 1.6.1

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

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

1.7 Conditional Expectation

Definition 1.7.1

If Y_1 and Y_2 are any two random variables, 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

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

if Y_1 and Y_2 are jointly discrete.

Example 1.7.1

Consider random variable Y_1, Y_2 with joint density function

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

Find the conditional expectation of the amount of sales, Y_1, given that $Y_2 = 1.5$

Solution: Note if $0 < y_2 \leqslant 2$,

$$f(y_1|y_2) = \begin{cases} \frac{1}{y_2}, & 0 < y_1 \leqslant y_2 \\ 0, & \text{elsewhere.} \end{cases}$$

so for any value of y_2 such that $0 < y_2 \leqslant 2$,

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

$$= \int_{0}^{y_2} y_1 \frac{1}{y_2} dy_1 = \frac{y_2}{2}$$

Theorem 1.7.1 The Towering Rule

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: Suppose that Y_1 and Y_2 are jointly continuous with joint density function $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. Then

$$\begin{aligned} E(Y_1) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1, y_2)\, dy_1\, dy_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1 \mid y_2) f_2(y_2)\, dy_1\, dy_2 \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} y_1 f(y_1 \mid y_2)\, dy_1\right\} f_2(y_2)\, dy_2 \\ &= \int_{-\infty}^{\infty} E(Y_1 \mid Y_2 = y_2) f_2(y_2)\, dy_2 = E[E(Y_1 \mid Y_2)]. \end{aligned}$$

The proof is similar for the discrete case.

😀

The conditional variance of Y_1 given $Y_2 = y_2$ is defined by analogy with an ordinary variance, again using the conditional density or probability function of Y_1 given $Y_2 = y_2$ in place of the ordinary density or probability function of Y_1. That is,

$$V(Y_1|Y_2 = y_2) = E(Y_1^2|Y_2 = y_2) - [E(Y_1|Y_2 = y_2)]^2$$

Theorem 1.7.2

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: As previously indicated, $V(Y_1 \mid Y_2)$ is given by

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

and

$$E[V(Y_1 \mid Y_2)] = E[E(Y_1^2 \mid Y_2)] - E\left\{[E(Y_1 \mid Y_2)]^2\right\}.$$

By definition,

$$V[E(Y_1 \mid Y_2)] = E\left\{[E(Y_1 \mid Y_2)]^2\right\} - \{E[E(Y_1 \mid Y_2)]\}^2.$$

The variance of Y_1 is

$$\begin{aligned} V(Y_1) &= E[Y_1^2] - [E(Y_1)]^2 \\ &= E\{E[Y_1^2 \mid Y_2]\} - \{E[E(Y_1 \mid Y_2)]\}^2 \\ &= E[Y_1^2 \mid Y_2] - \{E[E(Y_1 \mid Y_2)]\}^2 + E\{[E(Y_1 \mid Y_2)]^2\} - \{E[E(Y_1 \mid Y_2)]\}^2 \\ &= E[V(Y_1 \mid Y_2)] + V[E(Y_1 \mid Y_2)]. \end{aligned}$$

😀

Conditional Variancec

$$V[Y_1|Y_2 = y_2] = E(Y_1^2|Y_2 = y_2) - (E(Y_1|Y_2 = y_2))^2$$

Chapter 2

Functions of Random Variables

2.1 Method of distribution functions

Steps:

1. identify $u = g(Y)$ or $u = g(Y_1, Y_2)$
2. cdf: $F_u(u) = P(U \leqslant u)$

$$= P(g(Y) \leqslant u) = \begin{cases} \sum_{y_1, y_2 : g(y_1, y_2) \leqslant u} p(y_1 y_2) \\ \int\int_{(y_1, y_2) : g(y_1, y_2) \leqslant u} f(y_1, y_2) dy_1 dy_2 \end{cases}$$

3. if u is continuous, pdf $f_u(u) = \frac{dF_u(u)}{du}$

Corollary 2.1.1 Summary of the Distribution Function Method

Let U be a function of the random variables $Y_1, Y_2, \ldots, Y_n$.

1. Find the region $U = u$ in the $(y_1, y_2, \ldots, y_n)$ space.
2. Find the region $U \leqslant u$.
3. Find $F_U(u) = P(U \leqslant u)$ by integrating $f(y_1, y_2, \ldots, y_n)$ over the region $U \leqslant u$.
4. Find the density function $f_U(u)$ by differentiating $F_U(u)$. Thus,

$$f_U(u) = \frac{dF_U(u)}{du}.$$

2.2 Method of transformation

The method of transformation is a special case of the method of distribution, where $U = (Y)$ with function h strictly monotone on support of Y

Suppose that $h(y)$ is an increasing function of y and that $U = h(Y)$, where Y has density function $f_Y(y)$. Then $h^{-1}(u)$ is an increasing function of u: If $u_1 < u_2$, then $h^{-1}(u_1) = y_1 < y_2 = h^{-1}(u_2)$. We see from Figure 6.8 that the set of points y such that $h(y) \leqslant u$ is precisely the same as the set of points y such that $y \leqslant h^{-1}(u)$. Therefore (see Figure 6.8),

$$P(U \leqslant u) = P(h(Y) \leqslant u) = P(h^{-1}[h(Y)] \leqslant h^{-1}(u)) = P(Y \leqslant h^{-1}(u)).$$

or

$$F_U(u) = F_Y(h^{-1}(u)).$$

Then differentiating with respect to u, we have

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dF_Y(h^{-1}(u))}{du} = f_Y(h^{-1}(u))\frac{d[h^{-1}(u)]}{du}.$$

To simplify notation, we will write dh^{-1}/du instead of $d[h^{-1}(u)]/du$ and

$$f_U(u) = f_Y(h^{-1}(u))\frac{dh^{-1}}{du}.$$

Thus, we have acquired a new way to find $f_U(u)$ that evolved from the general method of distribution functions. To find $f_U(u)$, solve for y in terms of u; that is, find $y = h^{-1}(u)$ and substitute this expression into $f_Y(y)$. Then multiply this quantity by $\frac{dh^{-1}}{du}$.

Corollary 2.2.1

Let Y have probability density function $f_Y(y)$. If $h(y)$ is either increasing or decreasing for all y such that $f_Y(y) > 0$, then $U = h(Y)$ has density function

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|,$$

where

$$\frac{dh^{-1}}{du} = \frac{d[h^{-1}(u)]}{du}.$$

Note:-

Direct application of the method of transformation requires that the function $h(y)$ be either increasing or decreasing for all y such that $f_Y(y) > 0$. If you want to use this method to find the distribution of $U = h(Y)$, you should be very careful to check that the function $h(\cdot)$ is either increasing or decreasing for all y in the support of $f_Y(y)$. If it is not, the method of transformations cannot be used, and you should instead use the method of distribution functions

Summary of the Transformation Method

Let $U = h(Y)$, where $h(y)$ is either an increasing or decreasing function of y for all y such that $f_Y(y) > 0$.

1. Find the inverse function, $y = h^{-1}(u)$.
2. Evaluate
$$\frac{dh^{-1}}{du} = \frac{d[h^{-1}(u)]}{du}.$$
3. Find $f_U(u)$ by
$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|.$$

2.3 Moment Generating Functions

Theorem 2.3.1

Let $m_X(t)$ and $m_Y(t)$ denote the moment-generating functions of random variables X and Y , respectively. If both moment-generating functions exist and $m_X(t) = m_Y(t)$ for all values of t, then X and Y have the same probability distribution.

For a random variable Y , all the moments of Y are encapsulated in its moment generating function, denoted as $m_Y(t)$.

$$m_Y(t) = E(e^{tY})$$

$$= P(g(Y) \leqslant u) = \begin{cases} \sum_y e^{ty} p(y) \\ \\ \int_{-\infty}^{\infty} e^{ty} f(y) dy \end{cases}$$

The method of moment-generating functions is often very useful for finding the distributions of sums of independent random variables.

Theorem 2.3.2

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with moment-generating functions $m_{Y_1}(t), m_{Y_2}(t), \ldots, m_{Y_n}(t)$, respectively. If $U = Y_1 + Y_2 + \cdots + Y_n$, then

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \cdots \times m_{Y_n}(t).$$

We know that, because the random variables $Y_1, Y_2, \ldots, Y_n$ are independent,

$$m_U(t) = \mathbb{E}\left[e^{t(Y_1+\cdots+Y_n)}\right] = \mathbb{E}\left(e^{tY_1} e^{tY_2} \cdots e^{tY_n}\right)$$

$$= \mathbb{E}\left(e^{tY_1}\right) \times \mathbb{E}\left(e^{tY_2}\right) \times \cdots \times \mathbb{E}\left(e^{tY_n}\right).$$

Thus, by the definition of moment-generating functions,

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \cdots \times m_{Y_n}(t).$$

Example 2.3.1

Binomial MGF $Y \sim \text{bin}(n, 0)$

$$m_y(t) = E(e^{tY}) = \sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} e^{ty} \tag{2.1}$$

$$= \sum_{y=0}^{n} \binom{n}{y} (pe^t)^y (1-p)^{n-y} \tag{2.2}$$

$$= (pe^t + 1 - p)^n \tag{2.3}$$

2.4 MFG under linear transformation

Let $U = aY + b$, then $m_U(t) =$

$$\begin{aligned} &E(e^{tu}) = E(e^{t(aY+b)}) \\ &= E(e^{tb} e^{taY}) \\ &= e^{tb} E(e^{taY}) \\ &= e^{tb} m_Y(at) \end{aligned}$$

Let $S_n = \sum_{i=1}^n a_i Y_i$ and Y_i's are independent (may or may not have the same distribution) $m_{S_n}(t) =$

$$\begin{aligned} &E(e^{tS_n}) = E(e^{t\sum_{i=1}^n a_i Y_i}) \\ &= E(\prod_{i=1}^n e^{ta_i Y_i}) \\ &= \prod_{i=1}^n E(e^{ta_i Y_i}) \\ &= \prod_{i=1}^n m_{Y_i}(ta_i) \end{aligned}$$

Chapter 3

Sampling distribution

Theorem 3.0.1
Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2, then

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

is normally distributed with mean $\mu_{\bar{Y}=\mu}$ and variance $\sigma^2_{\bar{Y}} = \frac{\sigma^2}{n}$

then

$$Z = \frac{\bar{Y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}(\frac{\bar{Y} - \mu}{\sigma})$$

Theorem 3.0.2
Let $Y_1, Y_2, \ldots, Y_n$ be defined as in Theorem 7.1. Then $Z_i = \frac{(Y_i - \mu)}{\sigma}$ are independent, standard normal random variables, $i = 1, 2, \ldots, n$, and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2$$

has a χ^2 distribution with n degrees of freedom (df).

The χ^2 distribution plays an important role in many inferential procedures. For example, suppose that we wish to make an inference about the population variance σ^2 based on a random sample $Y_1, Y_2, \ldots, Y_n$ from a normal population. A good estimator of σ^2 is the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

The following theorem gives the probability distribution for a function of the statistic S^2.

Theorem 3.0.3
Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

has a χ^2 distribution with $(n-1)$ degrees of freedom (df). Also, $\overline{Y}$ and S^2 are independent random variables.

3.1 T distribution

Theorem 3.1.1

Let Z be a standard normal random variable and let W be a χ^2-distributed variable with ν degrees of freedom (df). Then, if Z and W are independent,

$$T = \frac{Z}{\sqrt{W/\nu}}$$

is said to have a t *distribution* with ν df.

If $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal population with mean μ and variance σ^2, Theorem 7.1 may be applied to show $Z = \sqrt{n}\frac{\overline{Y}-\mu}{\sigma}$ has a standard normal distribution. Theorem 7.3 tells us that $W = (n-1)\frac{S^2}{\sigma^2}$ has a χ^2 distribution with $\nu = n-1$ df and that Z and W are independent (because $\overline{Y}$ and S^2 are independent). Therefore, by Definition 7.2,

$$T = \frac{Z}{\sqrt{W/\nu}} = \frac{\sqrt{n}\frac{\overline{Y}-\mu}{\sigma}}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}} = \sqrt{n}\frac{\overline{Y}-\mu}{S}$$

has a t *distribution* with $(n-1)$ df.

3.2 F distribution

Definition 3.2.1

if $W_1 \sim \chi^2(\nu_1), W_2 \sim \chi^2(\nu_2)$ are independent, then

$$\frac{W_1/\nu_1}{W_2/\nu_2}$$

is said to have Snedecor's F-distribution with (ν_1, ν_2) degrees of freedom

$$W_1 = \frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1) \quad W_2 = \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$$

3.3 Central Limit Theorem

Theorem 3.3.1 Central Limit Theorem

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$. Define

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \quad \text{where} \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i.$$

Then the distribution function of U_n converges to the standard normal distribution function as $n \to \infty$. That is,

$$\lim_{n\to\infty} P(U_n \leqslant u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \quad \text{for all } u.$$

for **unknown** σ^2: Suppose $Y_1, \ldots, Y_n$ are iid random variables with mean μ and variance σ^2, both finite, then for any t

$$\lim_{n\to\infty} \mathbb{P}(\frac{\bar{Y} - \mu}{S/\sqrt{n}} \leqslant t) = \Phi$$

Chapter 4

Order Statistics

Suppose we have iid samples $Y_1, Y_2, \ldots, Y_n$ from some distribution. There are various statistics (which, you remember, are simply functions of the r.v.'s) that are only calculable if we sort the data in ascending order:

$$Y_{(1)} \leqslant Y_{(2)} \leqslant \ldots \leqslant Y_{(n)}$$

and $Y_{(1)}$ is the smallest of the set

Note:-

The k^{th} order statistics of a observed sample is the sample's k^{th} smallest value

$F_{(k)}(y) = \mathbb{P}(Y_{(k)} \leqslant y) = \mathbb{P}(\text{at least k of } Y_1, \ldots, Y_n \leqslant y) = \sum_{i=k}^{n} \binom{n}{i} F_Y(y)^i (1 - F_Y(y))^{n-i}$ for the pdf:

$$f_{(k)}(y) =$$

$$\frac{dF_{(k)}(y)}{dy} = \cdots = \begin{cases} \frac{n!}{(k-1)!(n-k)!} f_Y(y)[F_Y(y)]^{k-1}(1 - F_Y(y))^{n-k} \\ 0, \quad \text{elsewhere} \end{cases}$$

Note:-

Special Cases: $F_{(n)}(y) = \mathbb{P}(Y_{(n)} \leqslant y) = (F_Y(y))^n$
$f_{(n)}(y) = nF_Y(y)^{n-1} f_Y(y)$
$F_{(1)}(y) = \mathbb{P}(Y_{(1)} \leqslant y) = 1 - \mathbb{P}(Y_{(1)} > y) = 1 - (1 - F_Y(y))^n$
$f_{(1)}(y) = n(1 - F_Y(y))^{n-1} f_Y(y)$

Chapter 5

Finals review lecture

5.1 Bivariate Distribution Functions

- joint pdf: $f(y_1, y_2)$
- Marginal pdf: $f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_2$
- Conditional pdf: $f(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)}$, $\quad y_1 \in D(y_2)$, (remember to specify domain)
- Independence: if $f(y_1, y_2) = f_1(y_1)f_2(y_2)$ for all $y_1, y_2 \in \mathbb{R}^2$, also the two-step procedure.

5.2 Expectation and Covariance

- Expectation: $E(g(Y_1, Y_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(y_1, y_2)f(y_1, y_2)dy_1dy_2$
- if $Y_1 \perp Y_2$, $E(h(Y_1)g(Y_2)) = E(h(Y_1))E(g(Y_2))$

  Q: $E(Y_1/Y_2) = E(Y_1)/E(Y_2)$? $\quad$ not true in general $E(Y_1\frac{1}{Y_2}) = E(Y_1)E(\frac{1}{Y_2})$
- $Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$
- Correlation: $-1 \leqslant \rho = \frac{Cov(Y_1,Y_2)}{\sigma(Y_1)\sigma(Y_2) \leqslant 1}$

  independence $\implies$ uncorrelation but not the other way around

5.3 Conditional Expectation and Variance

- $E(Y_1|Y_2 = y_2) = \int_{-\infty}^{\infty} y_1 f(y_1|y_2)dy_1$ (is a function of y_2)

  then $E(Y_1|Y_2)$ is just a random variable because it is a function of r.v. Y_2
- Unconditional: $E(Y_1) = E(E(Y_1|Y_2))$, towering rule

  Q:$E(Y_1Y_2) = E(E(Y_1Y_2|Y_2)) = E(Y_2E(Y_1|Y_2))$, useful for $Cov(Y_1, Y_2)$
- $V(Y_1|Y_2 = y_2) = E(Y_1^2|Y_2 = y_2) - (E(Y_1|Y_2 = y_2))^2$

  $V(Y_1|Y_2)$ is a r.v.
- unconditional: $V(Y_1) = E(V(Y_1|Y_2)) + V(E(Y_1|Y_2))$

5.4 Distribution of Functions of Random Variables

- Method of distribution functions

  CDF: $F_u(u) = P(g(Y) \leqslant u)$

  PDF: $f_u(u) = \frac{dF_u(u)}{d_u}$

- Method of Transformation

  if g is monotone on the support of Y_1, then $y = g^{-1}(u) \implies \frac{dg^{-1}(u)}{du}$ (doesn't have to be monotone across entire domain, just the support of Y)

  $f_u(u) = f_Y(g^-1(u))|\frac{dg^-1(u)}{du}|$

- Method of MGF

  $u = \sum_{i=1}^{n} a_i Y_i + b$, $Y_i's$ are independent

  $m_u(0) = \prod_{i=1}^{n} m_{Y_i}(a_i t)e^{bt}$

  $M_Y(t) = E(e^{tY})$

  $E(Y^k) = |\frac{d^k m_Y(t)}{dt^k}|_{t=0}$

5.5 Sampling Distributions

- Exact Normality
- Approximation under Central Limit Theorem