Chapter VIII Sampling Distributions and the Central Limit Theorem

Functions of random variables are usually of interest in statistical application. Consider a set of observable random variables $X_1, X_2, \cdots, X_n$. For example, suppose the variables are a random sample of size n from a population.

Definition 8.0-1: A function of **observable** random variables, $U = g(X_1, X_2, \cdots, X_n)$, which does not depend on any *unknown* parameters, is called a ***statistic***.

It is required that the variables be **observable** because of the intended use of a statistic. The intent is to make inferences about the distribution of the set of random variables, and if the variables are not observable or if the function $g(X_1, X_2, \cdots, X_n)$ depends on unknown parameters, then U would not be useful in making such inferences.

Two important **statistics** are the sample mean $\overline{X}$ and the sample variance S^2. Of course, in a particular sample, say $x_1, x_2, \ldots, x_n$, we observed definite values of these **statistics**, $\overline{x}$ and s^2; however, we should recognize that each value is only one observation of respective random variable, $\overline{X}$ and S^2. That is, each $\overline{X}$ and S^2 is also a random variable with its own distribution.

8.1 Sampling Distributions Related to the Normal Distribution

Theorem 8.1-1: Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is normally distributed with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma_{\overline{X}}^2 = \sigma^2/n$, $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$.

Proof: Since the moment-generating function of each X is

$$M_X(t) = \exp\left(\mu t + \sigma^2 t^2/2\right),$$

the moment-generating function of $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is equal to

$$M_{\overline{X}}(t) = E\left(e^{(t/n)\sum X_i}\right) = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) = \left\{\exp\left[\mu\left(\frac{t}{n}\right) + \frac{\sigma^2 (t/n)^2}{2}\right]\right\}^n = \exp\left[\mu t + \frac{\left(\sigma^2/n\right)t^2}{2}\right].$$

However, the moment-generating function uniquely determines the distribution of the random variable. Since this one is that associated with the normal distribution $N(\mu, \sigma^2/n)$, the sample mean $\overline{X}$ is $N(\mu, \sigma^2/n)$. ◘

Theorem 8.2-2: Let $Z_1, Z_2, \cdots, Z_n$ have standard normal distributions, $N(0, 1)$. If these random variables are mutually independent, then $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ has a χ^2 distribution with n degrees of freedom.

Proof: The moment-generating function of W is given by

$$M_W(t) = E\left(e^{tW}\right) = E\left(\exp\left\{t\left(Z_1^2 + Z_2^2 + \cdots + Z_n^2\right)\right\}\right) = M_{Z_1^2}(t) \times M_{Z_2^2}(t) \times \cdots \times M_{Z_n^2}(t)$$

Since Z_i^2 is a χ^2 distributed random variable with 1 degree of freedom, we have

$$M_{Z_i^2}(t) = (1-2t)^{-1/2}, \quad i = 1, 2, \cdots, n.$$

Hence,

$$M_W(t) = (1-2t)^{-1/2} \times (1-2t)^{-1/2} \times \cdots \times (1-2t)^{-1/2} = (1-2t)^{-n/2}$$

The uniqueness of the moment-generating function implies that W is $\chi^2(n)$. ◘

Corollary 8.2-1: If $X_1, X_2, \cdots, X_n$ have mutually independent normal distributions $N\left(\mu_i, \sigma_i^2\right)$, $i = 1, 2, \cdots, n$, respectively, then the distribution of

$$W = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2}$$

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

Proof: We simply note that $Z_i = (X_i - \mu_i)/\sigma_i$ is $N(0,1)$, $i = 1, 2, \cdots, n$. ◘

Theorem 8.2-3: Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a normal distribution, $N(\mu, \sigma^2)$,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 ,$$

Then

(a) $\bar{X}$ and S^2 are independent.

(b) $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$ has a χ^2 distribution with $(n – 1)$ degrees of freedom.

Proof: The proof of part (a) is beyond this course; so we accept it without proof here. To prove part (b), note that

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n}\left[\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma}\right]^2 = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

since the cross-product term is equal to

$$2\sum_{i=1}^{n}\frac{(\bar{X} - \mu)(X_i - \bar{X})}{\sigma^2} = \frac{2(\bar{X} - \mu)}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}) = 0 .$$

But $Y_i = (X_i - \mu)/\sigma$, $i = 1, 2, \cdots, n$, are mutually independent standard normal variables. Hence $W = Y_1^2 + Y_2^2 + \cdots + Y_n^2$ is $\chi^2(n)$ by Theorem 8.2-2. Moreover, since $\bar{X} \sim N(\mu, \sigma^2/n)$, then $Z^2 = \left(\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = \dfrac{n(\bar{X} - \mu)^2}{\sigma^2}$ is $\chi^2(1)$. Thus,

$$W = \frac{(n-1)S^2}{\sigma^2} + Z^2 .$$

However, from part (a), $\bar{X}$ and S^2 are independent; thus Z^2 and S^2 are also independent. In the moment-generating function of W, this independence permits us to write

$$E\left(e^{t[(n-1)S^2/\sigma^2 + Z^2]}\right) = E\left(e^{t(n-1)S^2/\sigma^2} e^{tZ^2}\right) = E\left(e^{t(n-1)S^2/\sigma^2}\right)E\left(e^{tZ^2}\right).$$

Since W and Z^2 have χ^2 distributions, we can substitute their moment-generating functions to obtain

$$(1-2t)^{-n/2} = E\left(e^{t(n-1)S^2/\sigma^2}\right)(1-2t)^{-1/2}.$$

Equivalently, we have

$$E\left(e^{t(n-1)S^2/\sigma^2}\right) = (1-2t)^{-(n-1)/2}, \qquad t < 1/2.$$

This, of course, is the moment-generating function of a $\chi^2(n-1)$ variable, and accordingly $(n-1)S^2/\sigma^2$ has this distribution. ◘

Theorem 8.2-4: If Z is a standard normal distribution, $N(0,1)$, if U is a χ^2 distribution with v degrees of freedom, and if Z and U are independent, then

$$T = \frac{Z}{\sqrt{U/v}}$$

has a t distribution with v degrees of freedom. Its *p.d.f.* is

$$f(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)} \frac{1}{\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}, \qquad -\infty < t < \infty.$$

REMARK: This distribution was discovered by W. S. Gosset when he was working for an Irish brewery. Because Gosset published under the pseudonym Student, this distribution is sometimes known as Student's t distribution.

Proof: Since Z and U are independent, the joint *p.d.f.* of Z and U is

$$g(z,u) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \frac{1}{\Gamma(v/2)2^{v/2}} u^{v/2-1} e^{-u/2}, \qquad -\infty < z < \infty,\ 0 < u < \infty.$$

The distribution function $F(t) = \Pr(T \le t)$ of T is given by

$$F(t) = \Pr(Z/\sqrt{U/v} \le t) = \Pr(Z \le t\sqrt{U/v}) = \int_0^\infty \int_{-\infty}^{t\sqrt{u/v}} g(z,u)\,dz\,du.$$

That is,

$$F(t) = \frac{1}{\sqrt{\pi}\,\Gamma(v/2)} \int_0^\infty \left[\int_{-\infty}^{t\sqrt{u/v}} \frac{e^{-z^2/2}}{2^{(v+1)/2}} dz\right] u^{v/2-1} e^{-u/2} du.$$

The p.d.f. of T is the derivative of the distribution function; so applying the Fundamental Theorem of Calculus to the inner integral we see that

$$f(t) = F'(t) = \frac{1}{\sqrt{\pi}\,\Gamma(v/2)} \int_0^\infty \frac{e^{-(u/2)(t^2/v)}}{2^{(v+1)/2}} \sqrt{\frac{u}{v}}\, u^{v/2-1} e^{-u/2} du$$

$$= \frac{1}{\sqrt{\pi v}\,\Gamma(v/2)}\int_0^\infty \frac{u^{(v+1)/2-1}}{2^{(v+1)/2}}\, e^{-(u/2)\left(1+t^2/v\right)}du$$

In the integral make the change of variables $y = \left(1 + t^2/v\right)u$ so that $du/dy = 1/\left(1 + t^2/v\right)$. Thus we find that

$$f(t) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)}\frac{1}{\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}\int_0^\infty \frac{y^{(v+1)/2-1}}{\Gamma((v+1)/2)\,2^{(v+1)/2}}\, e^{-y/2}dy\ .$$

The integral is equal to 1 since the integrand is the p.d.f of a chi-square distribution with (v + 1) degrees of freedom. Thus the p.d.f. is as given in the theorem. ◘

Note that the distribution of T is completely determined by the number v. Its p.d.f. is symmetrical with respect the vertical axis $t = 0$ and is very similar to the graph of the p.d.f. of the standard normal distribution $N(0,1)$. It can be shown that $E(T) = 0$ for $v > 1$ and $Var(T) = v/(v-2)$ for $v > 2$.[1] When $v = 1$, the t distribution is the same as the standard Cauchy distribution in which the mean and the variance do not exist.

Theorem 8.2-5: If $X_1, X_2, \cdots, X_n$ denote a random sample from $N\left(\mu, \sigma^2\right)$, then

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

Proof: This follows from Theorem 8.2-4, since $\left(\bar{X}-\mu\right)/\left(\sigma/\sqrt{n}\right) \sim N(0,1)$ and by Theorem 8.2-3, $U = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, and $\bar{X}$ and S^2 are independent. ◘

Example 8.1-1: Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be independent random samples from populations with respectively distributions $X_i \sim N\left(\mu_x, \sigma_x^2\right)$ and $Y_j \sim N\left(\mu_y, \sigma_y^2\right)$. The distributions of $\bar{X}$ and $\bar{Y}$ are $N\left(\mu_x, \sigma_x^2/n\right)$ and $N\left(\mu_y, \sigma_y^2/m\right)$, respectively. Since $\bar{X}$ and $\bar{Y}$ are independent, the distribution $\bar{X}-\bar{Y}$ is $N\left(\mu_x-\mu_y, \sigma_x^2/n+\sigma_y^2/m\right)$, and

$$Z = \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_x-\mu_y\right)}{\sqrt{\sigma_x^2/n+\sigma_y^2/m}} \sim N(0,1).$$

Since $(n-1)S_x^2/\sigma_x^2 \sim \chi^2(n-1)$ and $(m-1)S_y^2/\sigma_y^2 \sim \chi^2(m-1)$, and both are independent,

$$U = \frac{(n-1)S_x^2}{\sigma_x^2}+\frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi^2(n+m-2).$$

---

[1] You can find these results by the following fact: If Y has a gamma distribution with parameter α and β, then

$$E\left(Y^k\right) = \frac{\beta^k\Gamma(\alpha+k)}{\Gamma(\alpha)} \qquad \text{for } (\alpha+k) > 0.$$

A random variable T with the t distribution having $v = n + m - 2$ degrees of freedom is given by

$$T = \frac{Z}{\sqrt{U/(n+m-2)}} = \frac{\left[(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)\right] \Big/ \sqrt{\sigma_x^2/n + \sigma_y^2/m}}{\sqrt{\left((n-1)S_x^2/\sigma_x^2 + (m-1)S_y^2/\sigma_y^2\right)/(n+m-2)}}.$$

In the statistical applications we sometimes assume that the two variances are the same, say $\sigma_x^2 = \sigma_y^2 = \sigma^2$, in which case

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_y)}{\sqrt{\left\{\left[(n-1)S_x^2 + (m-1)S_y^2\right] / (n+m-2)\right\} \left[(1/n) + (1/m)\right]}}$$

and neither T nor its distribution dependent on σ^2. ◘

Theorem 8.2-6: If U_1 and U_2 are independent chi-square variables with v_1 and v_2 degrees of freedom, respectively, then

$$F = \frac{U_1/v_1}{U_2/v_2}$$

Has an F distribution with v_1 and v_2 degrees of freedom. Its *p.d.f.* is

$$f(y) = \frac{\Gamma[(v_1 + v_2)/2]}{\Gamma(v_1/2)\Gamma(v_2/2)} \left(\frac{v_1}{v_2}\right)^{v_1/2} y^{(v_1/2)-1} \left(1 + \frac{v_1}{v_2} y\right)^{-(v_1+v_2)/2}, \quad 0 < y < \infty$$

REMARK: The symbol F was first proposed by George Snedecor to honor R. A. Fisher. Sometimes it also called Snedecor's F distribution.

Proof: Omitted. The p.d.f. can be derived in a manner similar to that of the t distribution as in Theorem 8.2-4. ◘

If F possesses an F distribution with v_1 numerator and v_2 denominator degrees of freedom, then

$E(F) = v_2/(v_2 - 2)$ if $v_2 > 2$ and $Var(F) = 2v_2^2(v_1 + v_2 - 2)/\{v_1(v_2 - 2)^2(v_2 - 4)\}$ if $v_2 > 4$.

You can have both $E(F)$ and $Var(F)$ in a manner similar to those of the t distribution. **Note** that the mean of an F-distributed random variable depends only on the number of denominator degrees of freedom, v_2.

Example 8.1-2: Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be independent random samples from populations with respectively distributions $X_i \sim N(\mu_x, \sigma_x^2)$ and $Y_j \sim N(\mu_y, \sigma_y^2)$. Since $(n-1)S_x^2/\sigma_x^2 \sim \chi^2(n-1)$ and $(m-1)S_y^2/\sigma_y^2 \sim \chi^2(m-1)$,

$$\frac{\left((n-1)S_x^2/\sigma_x^2\right)/(n-1)}{\left((m-1)S_y^2/\sigma_y^2\right)/(m-1)} = \frac{S_x^2\sigma_y^2}{S_y^2\sigma_x^2} \sim F((n-1), (m-1)).$$ ◘

8.2 Central Limit Theorem and Limiting Moment-Generating Functions

If $X_1, X_2, \cdots, X_n$ is a random sample of size n from a **normal** population, Theorem 8.1-1 tell us that $\overline{X}$ has normal sampling distribution with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n$, $\overline{X} \sim N(\mu, \sigma^2/n)$. But what can we say about the sampling distribution of $\overline{X}$ if the variables X_i are **NOT** normally distributed? Fortunately, $\overline{X}$ will have a sampling distribution that is approximately normal if the sample size is large. The formal statement of this result is called the ***central limit theorem***.

Theorem 8.2-1 (Central Limit Theorem): If $\overline{X}$ is the mean of a random sample $X_1, X_2, \cdots, X_n$ of size n from a distribution with $E(X_i) = \mu < \infty$ and $Var(X_i) = \sigma^2 < \infty$, then the distribution of

$$W = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$$

is $N(0,1)$ in the limit as $n \to \infty$.

Proof: This limiting result holds for random samples from any distribution with finite mean and variance, but the proof will be outlined under the stronger assumption that the moment-generating function of the distribution exists. The proof can be modified for the more general case by using a more general concept called a characteristic function, which we do not consider here.

We first consider

$$E[\exp(tW)] = E\left\{\exp\left[\left(\frac{t}{\sqrt{n}\,\sigma}\right)\left(\sum_{i=1}^{n} X_i - n\mu\right)\right]\right\}$$

$$= E\left\{\exp\left[\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_1 - \mu}{\sigma}\right)\right] \times \cdots \times \exp\left[\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_n - \mu}{\sigma}\right)\right]\right\}$$

$$= E\left\{\exp\left[\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_1 - \mu}{\sigma}\right)\right]\right\} \times \cdots \times E\left\{\exp\left[\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_n - \mu}{\sigma}\right)\right]\right\},$$

which follows from the mutual independence of $X_1, X_2, \cdots, X_n$. Then

$$E[\exp(tW)] = \left[M\left(t/\sqrt{n}\right)\right]^n, \qquad -h < t/\sqrt{n} < h,$$

where

$$M(t) = E\left\{\exp\left[t\left(\frac{X_i - \mu}{\sigma}\right)\right]\right\}, \qquad -h < t < h,$$

is the common moment-generating function of each

$$Y_i = \frac{X_i - \mu}{\sigma}, \qquad i = 1, 2, \cdots, n.$$

Since $E(Y_i) = 0$ and $E(Y_i^2) = 1$, it must be that

$$M(0) = 1, \qquad M'(0) = E\left(\frac{X_i - \mu}{\sigma}\right) = 0, \qquad M''(0) = E\left[\left(\frac{X_i - \mu}{\sigma}\right)^2\right] = 1.$$

Hence, using Taylor's formula with a remainder, we can find a number t_0 between 0 and t such that

$$M(t) = M(0) + M'(0)t + \frac{M''(t_0)t^2}{2} = 1 + \frac{M''(t_0)t^2}{2}.$$

By adding and subtracting $t^2/2$, we have that

$$M(t) = 1 + \frac{t^2}{2} + \frac{[M''(t_0) - 1]t^2}{2}.$$

Using this expression of $M(t)$ in $E[\exp(tW)]$, we can represent the moment-generating function of W by

$$E[\exp(tW)] = \left\{1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \frac{[M''(t_0) - 1]}{2}\left(\frac{t}{\sqrt{n}}\right)^2\right\}^n$$

$$= \left\{1 + \frac{t^2}{2n} + \frac{[M''(t_0) - 1]t^2}{2n}\right\}^n, \qquad -\sqrt{n}\,h < t < \sqrt{n}\,h$$

where now t_0 is between 0 and $t/\sqrt{n}$. Since $M''(t)$ is continuous at $t = 0$ and $t_0 \to 0$ as $n \to \infty$, we have that

$$\lim_{n\to\infty}[M''(t_0) - 1] = 1 - 1 = 0.$$

Thus, we have that

$$\lim_{n\to\infty} E[\exp(tW)] = \lim_{n\to\infty}\left\{1 + \frac{t^2}{2n} + \frac{[M''(t_0) - 1]t^2}{2n}\right\}^n = \lim_{n\to\infty}\left\{1 + \frac{t^2/2}{n}\right\}^n = e^{t^2/2},$$

for all real t. ◻

We have shown that the binomial distribution can be approximated by the Poisson distribution when n is sufficiently large and p fairly small in Chapter 4. This section will show this by taking the limit of a moment-generating function. Consider the moment-generating function of X, which is $B(n, p)$. We shall take the limit of this as $n \to \infty$ such that $np = \lambda$ is a constant; thus $p \to 0$. The moment-generating function of X is

$$M(t) = (1 - p + pe^t)^n.$$

Since $p = \lambda/n$, we have that

$$M(t) = \left[1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^t\right]^n = \left[1 + \frac{\lambda(e^t - 1)}{n}\right]^n.$$

Since

$$\lim_{n\to\infty}\left(1+\frac{b}{n}\right)^n = e^b ,$$

we have

$$\lim_{n\to\infty} M(t) = \exp\left\{\lambda\left(e^t - 1\right)\right\},$$

which exists for all real *t*. Hence a Poisson distribution seems like a reasonable approximation to the binomial one when *n* is large and *p* is small. This approximation is usually found to be fairly successful if $n \geq 20$ and $p \leq 0.05$ and very successful if $n \geq 100$ and $np \leq 10$.