Chapter 7 Sampling Distribution: Univariate Case

► Goldberger, Ch. 8

Wackerly et al. Chapter 7

Yale Note, Ch. 10

► Random Sample

► Sample Statistics

► Sampling Distribution
  - from Normal Distribution
  - from non-Normal Distribution

1. Random Sample

- X : random variable with $f(x)$.

► Let $X_1, X_2, \cdots, X_n$ be independent drawings from the population.

⇒ The outcome when the same experiment is repeated n times independently.

► $\mathbf{X} = \left(X_1, X_2, \cdots, X_n \right)'$; random sample of size n on the variable X.

(From the population X / From the probability distribution $f(x)$).

► $\mathbf{x} = \left(x_1, x_2, \cdots, x_n \right)'$; taken values of random vector $\mathbf{X} = \left(X_1, X_2, \cdots, X_n \right)'$.

• If $\mathbf{X}=\left(X_1, X_2, \cdots, X_n\right)'$ is a random sample on X, then the X_i's are ***independent and identically distributed****(i.i.d.)*.

► Joint density $g_n(\mathbf{x}) = g(x_1, x_2, \cdots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) = \prod_{i=1}^{n} f(x_i)$.

(Example)

① $X \sim Bernoulli(p)$,

☞ $g_n(\mathbf{x})$

► Here, $g_n(\mathbf{x}) = P\left(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\right)$: likelihood.

② $X \sim N\left(\mu, \sigma^2\right)$, $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{\left(x-\mu\right)^2}{2\sigma^2}\right)$ for $-\infty < x < \infty$.

☞ $g_n(\mathbf{x})$

2. Sample Statistics

- Let $T = h\left(X_1, X_2, \cdots, X_n\right) = h\left(\mathbf{X}\right)$ be a scalar function of the random sample.

Then, T is called *sample statistics*(표본통계량).

(Example)

① $\bar{X} = \dfrac{1}{n}\sum_i X_i$: sample mean.

② $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$: sample variance.

③ $M_r' = \dfrac{1}{n}\sum_{i=1}^{n} X_i^r$: sample raw moment.

④ $M_r = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^r$: sample central moment.

- It is possible to derive the ***sampling distribution*** of $T = h(X_1, X_2, \cdots, X_n) = h(\mathbf{X})$ from $f(x)$ and n (to make an inference on population parameter).

(1) Sample Mean

The sampling distribution of sample mean differs,

as ⓐ population distribution differs and

ⓑ sample size differs (small sample vs. large sample).

• Common properties of the sample mean:

① $E\left(\bar{X}\right)$

② $Var\left(\bar{X}\right)$

⦿ Sample Mean Theorem

In random sampling with sample size n, from any distribution with $E\left(X_i\right)=\mu$ and $V(X_i)=\sigma^2$, sample mean $\bar{X}$ has $E\left(\bar{X}\right)=\mu,\ V\left(\bar{X}\right)=\dfrac{\sigma^2}{n}$.

(2) Other Sample Moments

① Sample raw moment: $M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$

(Note) Population raw moment: $\mu_r' = E\left(X^r\right)$.

► M_r' has mean $E\left(M_r'\right) = \mu_r'$, $V\left(M_r'\right) = \frac{1}{n}\left(\mu_{2r}' - \left(\mu_r'\right)^2\right)$.

► Analogy principle

② Sample central moment (about population mean)

- $M_r^* = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^r$.

Population moment: $\mu_r = E\left[\left(X - \mu\right)^r\right]$.

► Sample mean theorem applies with $E\left(M_r^*\right) = \mu_r,\ V\left(M_r^*\right) = \dfrac{1}{n}\left(\mu_{2r} - \left(\mu_r\right)^2\right)$.

③ Sample central moment about *sample mean*

- $M_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^r$.

(Example) $r = 2,\ M_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n-1}{n}s^2, \quad \text{where } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Let $Y_i = (X_i - \overline{X})^r$, so $M_r = \overline{Y}$.

However, the Y_i's are NOT independent(so NOT i.i.d.)

► Let $U_1 = \left(X_1 - \overline{X}\right)$, $U_2 = \left(X_2 - \overline{X}\right)$, then $Cov\left(U_1, U_2\right) \neq 0$.

So, sample mean theorem can NOT be applied.

► However, the brute force can be applied to the expectation and variance.

(Example) Consider $M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

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

So, $E\left[M_2\right] = \frac{n-1}{n}\sigma^2$.

(Note) $E[M_2] = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$.

ⓐ As $n \to \infty$, $E[M_2] \to \sigma^2$.

ⓑ $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, then $E\left[s^2\right] = \sigma^2$; unbiased.

3. Sampling Distribution: the probability distribution of a sample statistic.

① <u>Chi-squares distribution</u>

- If $Z_1, Z_2, \cdots, Z_k \sim N(0,1)$, then $W = \sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$.

$$g_k(w) = \begin{cases} \dfrac{\left(\frac{1}{2}\right)\left(\frac{w}{2}\right)^{\frac{k}{2}-1} \exp\left(-\frac{w}{2}\right)}{\Gamma(k/2)} & \textbf{for } w > 0 \\ 0 & \textbf{otherwise} \end{cases}$$

where $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \ \Gamma(1) = 1, \ \Gamma(n) = (n-1)\Gamma(n-1)$.

► Single parameter k, the number of degrees of freedom.

► $\chi^2(k)$ random variable only takes positive values.

► Skewed to the right.

- Mean & Variance

Since $E(Z_i)=0,\ E(Z_i^{\,2})=1,\ E(Z_i^{\,4})=3,\ \text{so}\ V(Z_i^{\,2})=2$.

ⓐ $E(W)=k$

ⓑ $V(W)=2k$

② Student's t-distribution

If $Z_i \sim N\left(0,1\right)$, $W \sim \chi^2(k)$ with Z_i and W are independent, then

$$U = \frac{Z_i}{\sqrt{W/k}} = \frac{"N(0,1)"}{"\sqrt{X^2(k)/k}"} \sim t(k)$$

- The pdf of U is

$$g_k(u) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k}\,\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)}\left(1+\frac{u^2}{k}\right)^{-\frac{k+1}{2}}.$$

► Single parameter k, the number of degrees of freedom.

► The pdf is symmetric about zero.

► Similar in shape to a standard normal pdf.

► More area in the tails relative to a standard normal.

- <u>Mean & Variance</u>

ⓐ $E\left(U\right)=0$

ⓑ $Var(U)=\dfrac{k}{k-2}$.

► For future reference, $E\left(\dfrac{1}{W}\right)=\dfrac{1}{k-2}$ **for** $k>2$, $E\left(\dfrac{1}{W^2}\right)=\dfrac{1}{(k-2)(k-4)}$ **for** $k>4$.

► As $k\to\infty$, $V(U)\to 1$. So, $G_k(\cdot)\to\Phi(\cdot)$.

► Name of student-t distribution.

(Remark)

► $t(1)$ is called Cauchy distribution. Its mean and variance do not exist.

► $t(2)$ has mean 0 but no variance.

- <u>Some Important Applications</u>:

ⓐ Let $W_1, W_2, \cdots, W_k \sim$ **indep.** $\chi^2(1), \chi^2(2), \cdots, \chi^2(k)$,

then $W_1 + W_2 + \cdots + W_k \sim \chi^2(1 + 2 + \cdots + k)$.

ⓑ Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$, then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

ⓒ If $W_1 \sim \chi^2(r_1)$, $W_1 + W_2 \sim \chi^2(r)$ and independent with $r > r_1$, then

$W_2 \sim \chi^2(r - r_1)$.

4. Sampling Distribution from Normal Distribution

- $X_1, X_2, \cdots, X_n$: random sample from $N(\mu, \sigma^2)$.

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

① $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

② $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

③ $\bar{X}$ and S^2 are independent.

④ $\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{S} = \frac{\left(\bar{X} - \mu\right)}{S/\sqrt{n}} \sim t(n-1)$.

(note) If $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, ② $\frac{nS^{*2}}{\sigma^2} \sim \chi^2(n-1)$.

Then, ④ $\frac{\sqrt{n-1}\left(\bar{X}-\mu\right)}{S^*} = \frac{\left(\bar{X}-\mu\right)}{S^*/\sqrt{n-1}} \sim t(n-1)$.

(Proof of ② under ③)

5. Sampling Distribution <u>from non-Normal Distribution</u>

We can NOT derive the general conclusion when the population distribution is not normal.

▶ Asymptotic distribution(Large sample theory)