

The Central Limit Theorem

When n becomes larger we have more information, we expect the values of sample mean $\bar{Y}$ to be closer to population mean μ By using Chebyshev's theorem

For any positive constant c , prob. that $\bar{Y}$ will take on a value between $\mu - c$ & $\mu + c$ is at least $1 - \frac{\sigma^2}{nc^2}$

When $n \rightarrow \infty$, this probability approaches 0
This result is called law of large numbers

Central Limit Theorem (CLT)

It concerns limiting distributions of standardized mean of n random variables when $n \rightarrow \infty$

If $Y_1, Y_2, \dots, Y_n$ constitute a random sample from an infinite population with mean μ & variance σ^2 , moment generating func. $M_Y(t)$ then limiting distribution

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \quad \text{as } n \rightarrow \infty \quad \text{is standard normal distribution}$$

CLT is interpreted incorrectly as implying that distribution of Y approaches a normal distr. when $n \rightarrow \infty$, this is incorrect

This is incorrect since variance of $\bar{Y}$ $\text{var}(\bar{Y}) \rightarrow 0$ when $n \rightarrow \infty$

On the other hand CLT justifies approximating distribution of $\bar{Y}$ with a normal distr. having mean μ & variance σ^2/n when n is large

In practice, this approximation is used when $n \geq 30$ regardless of actual shape of population

For smaller values of n , the approximation is questionable

If $\bar{y}$ is a mean of a random sample of size n from a normal population with mean μ & variance σ^2 , its sampling distribution is a normal distribution with mean μ & variance σ^2/n

Order Statistics

Consider a normal sample of size n from a population with continuous density

Suppose that we arrange values of $y_1, y_2, \dots, y_n$ according to size

$$\begin{bmatrix} x_1 \\ 11 \\ 5 \\ 12 \\ 3 \\ 7 \\ 10 \end{bmatrix} \longrightarrow \begin{bmatrix} y_1 \\ 3 \\ 5 \\ 7 \\ 10 \\ 11 \\ 12 \end{bmatrix}$$

If we look upon the smallest of the x 's as a value of random variable y_1 , the next largest as a value of random variable y_2 , the next as $y_3, \dots$, the the largest as a value of random variable y_n .

We refer these y 's as order statistics

$y_1 \Rightarrow$ is 1st order statistics

$y_2 \Rightarrow$ is 2nd order statistics

$y_3 \Rightarrow$ is 3rd order statistics

$\vdots$

$y_n \Rightarrow$ is n^{th} order statistics

We are limiting this discussion to infinite populations with continuous density so that there is a zero prob. that any of two x 's will be alike

consider case $n=2$

$$\begin{aligned} y_1 &= x_1, & y_2 &= x_2 & \text{when } x_1 < x_2 \\ y_1 &= x_2, & y_2 &= x_1 & \text{when } x_2 < x_1 \end{aligned}$$

Consider a case of $n=3$

$$\begin{aligned} y_1 &= x_1, & y_2 &= x_2, & y_3 &= x_3 & \text{when } x_1 < x_2 < x_3 \\ y_1 &= x_1, & y_2 &= x_3, & y_3 &= x_2 & \text{when } x_1 < x_3 < x_2 \\ y_1 &= x_3, & y_2 &= x_2, & y_3 &= x_1 & \text{when } x_3 < x_2 < x_1 \end{aligned}$$

For prob. density of r th order stat for $r = 1, 2, \dots, n$ for random samples of size n from an infinite population that has values $f(x)$ at x , prob. density of the r th order stat. y_r is given by

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!} \left[\int_{-\infty}^{y_r} f(x) \cdot dx \right]^{r-1} f(y_r) \left[\int_{y_r}^{\infty} f(x) \cdot dx \right]^{n-r}$$

for $-\infty < y_r < \infty$

sampling distribution of y_1 , smallest value in a random sample of size n is

$$g_1(y_1) = n f(y_1) \left[\int_{y_1}^{\infty} f(x) \cdot dx \right]^{n-1}$$

sampling distribution of y_n , the largest value in a random sample of size n is

$$g_n(y_n) = f(y_n) \left[\int_{-\infty}^{y_n} f(x) \cdot dx \right]^{n-1}$$

In a random sample of size $n = 2m+1$, sample median, $\tilde{x}$, is y_{m+1} , whose sampling distr. is

$$h(\tilde{x}) = \frac{(2m+1)!}{m!m!} \left[\int_{-\infty}^{\tilde{x}} f(x) \cdot dx \right]^m f(\tilde{x}) \left[\int_{\tilde{x}}^{\infty} f(x) \cdot dx \right]^m$$

for $-\infty < \tilde{x} < \infty$

For a random sample of size $n = 2m$, the median is defined $\frac{1}{2}(y_m + y_{m+1})$

Sampling Distribution of Median

which holds when population density is continuous & non zero at population median, $\tilde{\mu}$, such that $\int_{-\infty}^{\tilde{\mu}} f(x) \cdot dx = \frac{1}{2}$

For large n sampling distribution of median for random samples of size $2n+1$ is approximately normal with mean $\tilde{\mu}$ and variance $\frac{1}{8[f(\tilde{\mu})^2 n]}$

For random samples of size $2n+1$ from a normal population we have $\mu = \tilde{\mu}$ so

$$f(\tilde{\mu}) = f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

and variance of median is approximately $\frac{\pi\sigma^2}{4n}$

If we compare this with variance of mean for random samples of size $2n+1$ from a infinite population, $\frac{\sigma^2}{2n+1}$,

we find that for large samples from normal populations the mean is more reliable than median for measuring central tendency.

Mean is subject to smaller chance fluctuations than median

Theory of Estimation

Statistical $\rightarrow$ Estimation
Inference $\rightarrow$ Hypothesis Testing
Estimation $\rightarrow$ point estimation
 $\rightarrow$ Interval estimation

purpose of point estimation is to use a sample of data plus any a priori (non sample) information we have about probability distribution in question

$\hookrightarrow$ To provide a value that is our best estimate of unknown parameter value

Estimator vs. Estimate
 $\hookrightarrow$ function $\hookrightarrow$ value

We are also concerned about reliability of estimate. It is sometimes useful when reporting results to give a range of possible values that incorporates a point estimate & measure of precision $\rightarrow$ interval estimate does this

Interval estimates are usually based on sampling distribution of an estimator.

Properties of Estimators

1- Unbiasedness

$\hat{\theta}$ is an unbiased estimator of θ if $E(\hat{\theta}) = \theta$

If $E(\hat{\theta}) \neq \theta \Rightarrow \hat{\theta}$ is a biased estimator of θ

$$E(\hat{\theta}) - \theta = \delta(\hat{\theta}) \Rightarrow \text{amount of biased}$$

Unbiasedness means that estimator on average will yield true parameter value

If underlying experiments are presented infinitely many times by drawing samples of size n , average value of estimate $\hat{\theta}$ from all these samples will be θ

Ex: Consider sample mean $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i$

Is it an unbiased estimator?

$$E(\hat{\theta}) = \sum_{i=1}^n \frac{1}{n} E(y_i) = \sum_{i=1}^n \theta/n = n(\theta/n) = \theta$$

Ex: If $x_1, x_2, \dots, x_n$ constitute a random sample from a population given by

$$f(x) = \begin{cases} e^{-(x-\delta)} & \text{for } x > \delta \\ 0 & \text{otherwise} \end{cases}$$

Show that $\bar{x}$ is a biased estimator of δ

mean of population

$$E(\bar{x}) = \mu = \int_{\delta}^{\infty} x e^{-(x-\delta)} \cdot dx = 1 + \delta$$

$$E(\bar{x}) = 1 + \delta \neq \delta \Rightarrow \bar{x} \text{ is a biased estimator}$$

Amount of Bias

$$E(\bar{x}) - \delta = 1 + \delta - \delta = 1$$

$$E(\bar{x} - 1) = \delta \Rightarrow \bar{x} - 1 \text{ is an unbiased estimator}$$

Bias & Precision

Both bias and variance must be taken into account when evaluating estimator.

↳ one way of doing this is to consider an estimator's mean square error

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

Average squared distance between $\hat{\theta}$ & true parameter value θ

$(\hat{\theta} - \theta)^2 \Rightarrow$ measures loss from using $\hat{\theta}$ as an estimator of θ

Taking mathematical expectation of this loss yields average loss or risk of using $\hat{\theta}$ to estimate θ

$$MSE(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2$$

$$= E\{[\hat{\theta} - E(\hat{\theta}) + [E(\hat{\theta}) - \theta]]^2\}$$

$$= E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta} - \theta)]^2 + \underbrace{2E[\hat{\theta} - E(\hat{\theta})E(\theta) - \theta]}_{\substack{|| \\ 0}}$$

Since

$$E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = [E(\hat{\theta} - \theta)][E(\hat{\theta}) - E(\hat{\theta})]$$

$$= [E(\hat{\theta}) - \theta] \underbrace{[E(\hat{\theta}) - E(\hat{\theta})]}_{\substack{|| \\ 0}} = 0$$

Thus,

$$\begin{aligned} MSE(\hat{\theta}) &= E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta} - \theta)]^2 \\ &= \text{Var}(\hat{\theta}) + [\text{bias}(\hat{\theta})]^2 \end{aligned}$$

Use of MSE allows us to compare estimators

A biased estimator may be preferred if it yields to a smaller MSE

2-Efficiency

Another way to take into account both bias and variance is to compare estimators that are unbiased

We check that whether a given unbiased estimator has the smallest possible variance

Minimum variance unbiased estimators

↳ best unbiased estimators

An estimator $\hat{\theta}$ is an efficient estimator of θ if $E(\hat{\theta}) = \theta$ (unbiasedness) and $\text{Var}(\hat{\theta}) \leq \text{Var}(\hat{\phi})$ where $\hat{\phi}$ is any other unbiased estimator of θ .

Cramer - Rao inequality

provides a sufficient but not necessary condition for an estimator to be efficient

$$\rightarrow \text{Var}(\hat{\theta}) \geq \frac{1}{n E \left[\left(\frac{\partial \ln f(x)}{\partial \theta} \right)^2 \right]}$$

This result states that $\hat{\theta}$ has variance equal to $\frac{1}{n E \left[\left(\frac{\partial \ln f(x)}{\partial \theta} \right)^2 \right]}$ then it is efficient. Since no lower variance is possible for unbiased estimator

$$\frac{1}{n E \left[\left(\frac{\partial \ln f(x)}{\partial \theta} \right)^2 \right]} \Rightarrow \text{Cramer - Rao Lower Bound (CRLB)}$$

This condition is sufficient but not necessary, it is possible that no unbiased estimator has variance as small as CRLB

Ex: Show that $\bar{x}$ is a minimum variance unbiased estimator of mean μ of a normal population

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} \quad \text{for } -\infty < x < \infty$$

$$\ln f(x) = -\ln \sigma \sqrt{2\pi} - \frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2$$

$$\frac{\partial \ln f(x)}{\partial \mu} = \frac{1}{\sigma} \left(\frac{x - \mu}{\sigma} \right) \Rightarrow E \left[\left(\frac{\partial \ln f(x)}{\partial \mu} \right)^2 \right] =$$

$$= \frac{1}{\sigma^2} E \left[\left(\frac{x - \mu}{\sigma} \right)^2 \right] = \frac{1}{\sigma^2} \cdot 1 = \frac{1}{\sigma^2}$$

$$\frac{1}{n E \left[\dots \right]} = \frac{1}{n \cdot \frac{1}{\sigma^2}} = \sigma^2 / n$$

Since $\bar{x}$ is unbiased and $\text{var}(\bar{x}) = \sigma^2/n$, it follows that $\bar{x}$ is a minimum variance unbiased estimator of μ

Asymptotic Efficiency

Assume sample size n is large $n \rightarrow \infty$
 suppose we have two estimators, $\hat{\theta}$ & $\tilde{\theta}$ of a parameter
 such that difference between

$$\sqrt{n} (\hat{\theta} - \theta) \sim N(0, G_1^2)$$

$$\sqrt{n} (\tilde{\theta} - \theta) \sim N(0, G_2^2)$$

If $G_2^2 \geq G_1^2$, then $\hat{\theta}$ is asymptotically efficient relative to $\tilde{\theta}$