Lecture 20

- Variance of the Sample Mean
  (1) Sample with Replacement: N >> n
  (2) Sample without Replacement: N is small

Case (1): with Replacement

Question: $Var(\hat{X})$?

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

also a RV

$$Var(\hat{X}) = E\left[(\hat{X})^2\right] - E^2\left[\hat{X}\right]$$

$$E\left[\hat{X}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n}\cdot\sum_{i=1}^{n}\bar{X} = \frac{1}{n}\cdot n\cdot\bar{X} = \bar{X}$$

True Mean

How about $E\left[(\hat{X})^2\right]$?

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

$$(\hat{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\cdot\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i\cdot X_j$$

$$E(\hat{X})^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[X_i X_j\right], \text{ where}$$

$$E\left[X_i X_j\right] = \begin{cases} E\left[X^2\right], & \text{if } i=j \\ \underbrace{E\left[X_i\right]E\left[X_j\right]}_{E^2[X]}, & \text{if } i\neq j \end{cases}$$

assume that X_i and X_j are uncorrelated

$$E(\hat{X})^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E[X_i X_j] = \frac{1}{n^2}\left\{nE[X^2] + \underbrace{(n^2-n)E^2[X]}_{n^2E^2[X] - nE^2[X]}\right\}$$

$$= \frac{1}{n^2}\left\{n\,Var[X] + n^2E^2[X]\right\} = \frac{Var[X]}{n} + E^2[X]$$

$$Var(\hat{X}) = E\left[(\hat{X})^2\right] - E^2[\hat{X}]$$

$$= \frac{Var[X]}{n} + E^2[X] - E^2[X]$$

$$\boxed{Var(\hat{X}) = \frac{Var[X]}{n}}$$

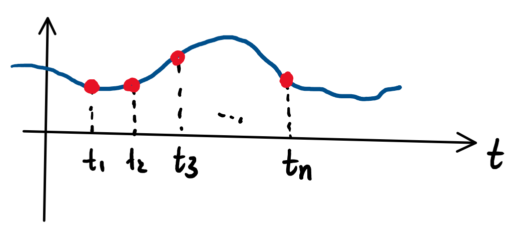
$$n\uparrow \;\Rightarrow\; Var(\hat{X})\downarrow$$

Example :

Assume that

True Mean = 10

True Variance = 9



$\rightarrow n$?

Question: How many samples we should take to estimate the true mean value such that

$$std(\hat{X}) = 1\% \text{ of the true mean value ?}$$

$\downarrow$

$$Var(\hat{X}) = (1\% \text{ of } 10)^2 = (0.01 \times 10)^2 = 0.01 = \frac{Var[X]}{n} = \frac{9}{n}$$

Thus $n = \frac{9}{0.01} = 900$

(2) Sample without Replacement: N is small

$$\boxed{Var(\hat{X}) = \frac{Var[X]}{n}\left(\frac{N-n}{N-1}\right)}$$

N: Population Size

n: Sample Size

Example 2:

$N = 100$ transistors

Estimate the mean value of the current gain β

$$\begin{cases} \text{True population variance} = 25 \\ \text{True population mean} = 120 \end{cases}$$

Question: How many samples we should take to estimate the true mean value such that

Sample mean has a STD that is 1% of the True Mean?

$$Var(\hat{\beta}) = (1\% \text{ of } 120)^2 = (0.01 \times 120)^2 = 1.44$$

$$1.44 = \frac{Var(\beta)}{n}\left(\frac{N-n}{N-1}\right) = \frac{25}{n}\cdot\left(\frac{100-n}{100-1}\right) = \frac{25}{n}\cdot\frac{100-n}{99}$$

$$1.44 \times 99n = 25(100-n) = 2500 - 25n$$

$$(1.44 \times 99 + 25)n = 2500$$

$$n = \frac{2500}{1.44 \times 99 + 25} = 15$$

```
>> 2500/(1.44*99 + 25)
ans =
14.9200
```

→ round it up

Sum of many independent, identical RV's

Example: $X_1, X_2, \ldots, X_k$ are all $N(0, \sigma)$

$\uparrow$ Mean $\uparrow$ Standard deviation

$$Z = \sum_{i=1}^{k} X_i$$

Question: PDF of Z ?

Answer: $\phi_Z(u) = \phi_{X_1}(u) \cdot \phi_{X_2}(u) \cdot \cdots \cdot \phi_{X_k}(u) = e^{-\frac{u^2\sigma^2}{2}} \cdot e^{-\frac{u^2\sigma^2}{2}} \cdots e^{-\frac{u^2\sigma^2}{2}}$

$$= e^{-u^2\frac{k\sigma^2}{2}} \rightarrow N(0, \sqrt{k\sigma^2})$$

Central Limit Theorem

$X_1, X_2, \ldots, X_n$ are independent RV's, having the same distribution (Mean $= m$, Variance $= \sigma^2$)

$$Y = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - m), \quad \text{where } E[Y] = 0, \ Var[Y] = \sigma^2$$

Then Y has a PDF $f_Y(y)$ that approaches a Gaussian distribution as n increases, regardless of the PDF of X_k's.

http://www.ece.uah.edu/~dwpan/course/ee385/code/clt_hist.m

```
function [m, v] = clt_hist(n)
 x = random('exp', 1, 1, 1000000);
 mean(x)
 var(x)
hist(x, 100);
for i = 2: n
   y = random('exp', 1, 1, 1000000);
   x = x - mean(x) + (y - mean(y));
end
S = x/sqrt(n);
m = mean(S);
v = var(S);
figure; hist(S,100);
```