Basic Statistics 03

Sampling & Central Limit Theorem

Sampling & C.L.T.

1. The distribution of the sample mean
2. The Central Limit Theorem
3. The application of CLT

1. Distribution of the Sample Means

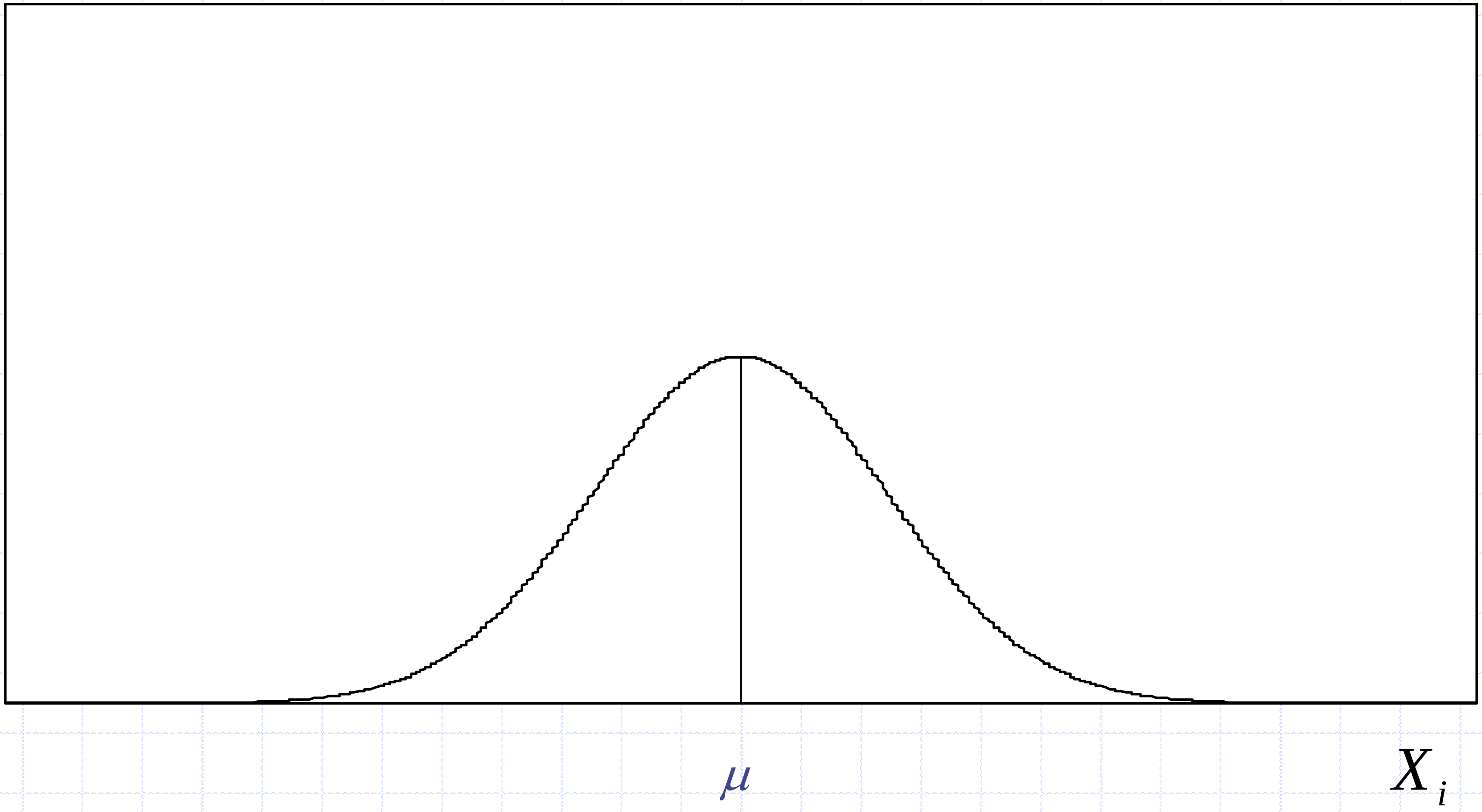
The ***sampling distribution of the sample mean*** is a probability distribution consisting of all possible sample means of a given sample size selected from a population.

$$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{n\mu}{n} = \mu$$

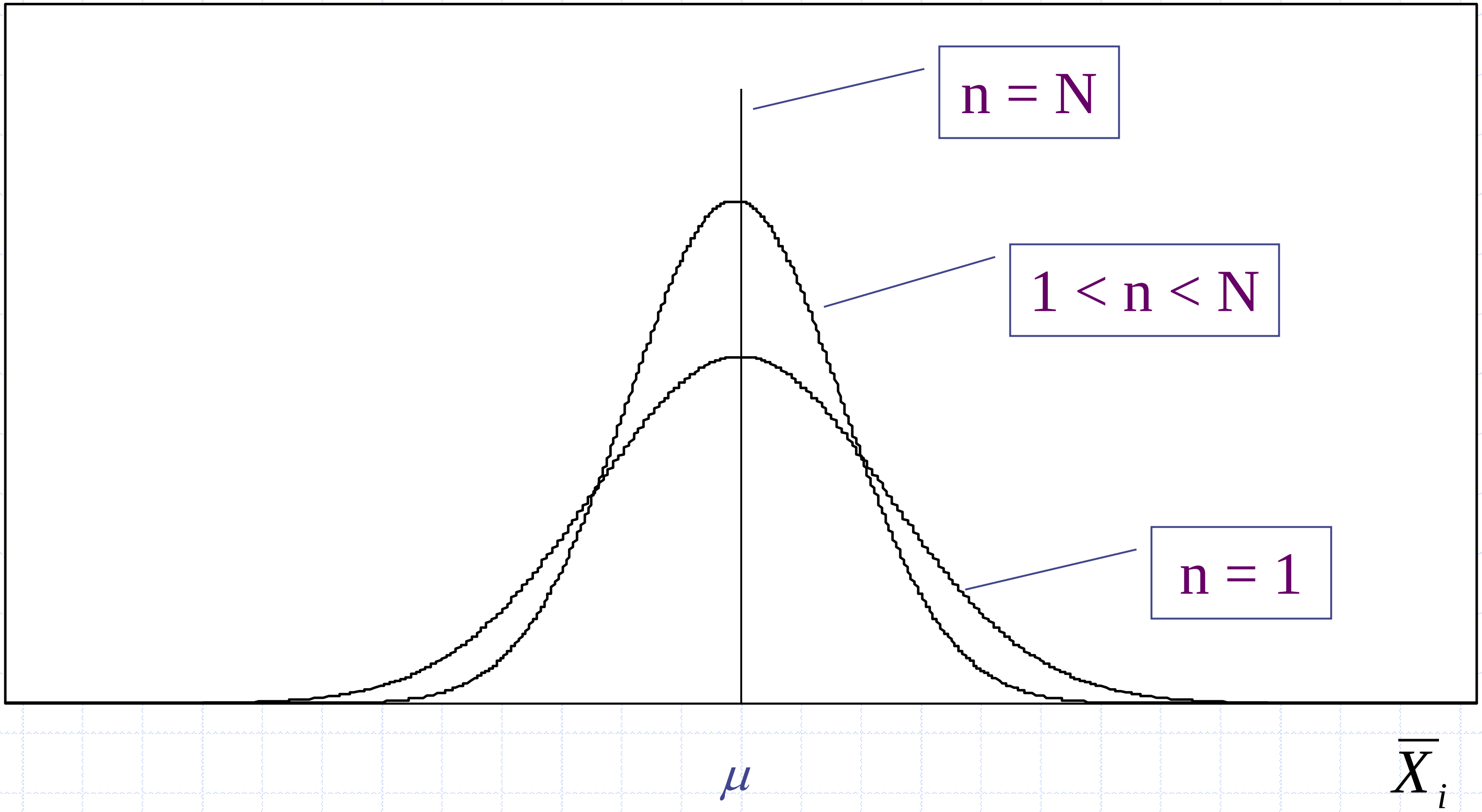
$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

*in the case of finite population, $Var(\bar{X}) = \frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}$

Population Distribution



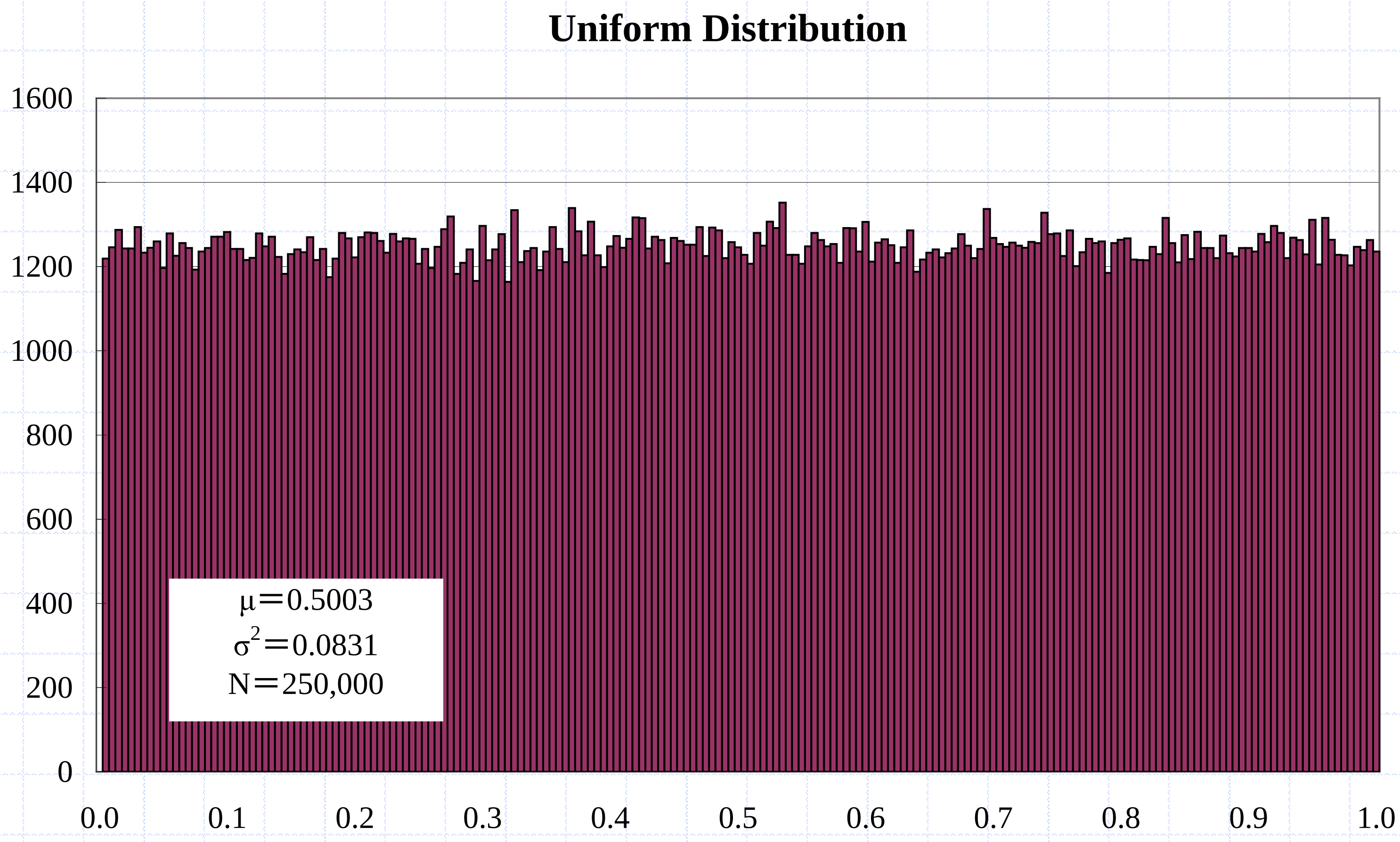
Sampling Dist. of the Sample Means



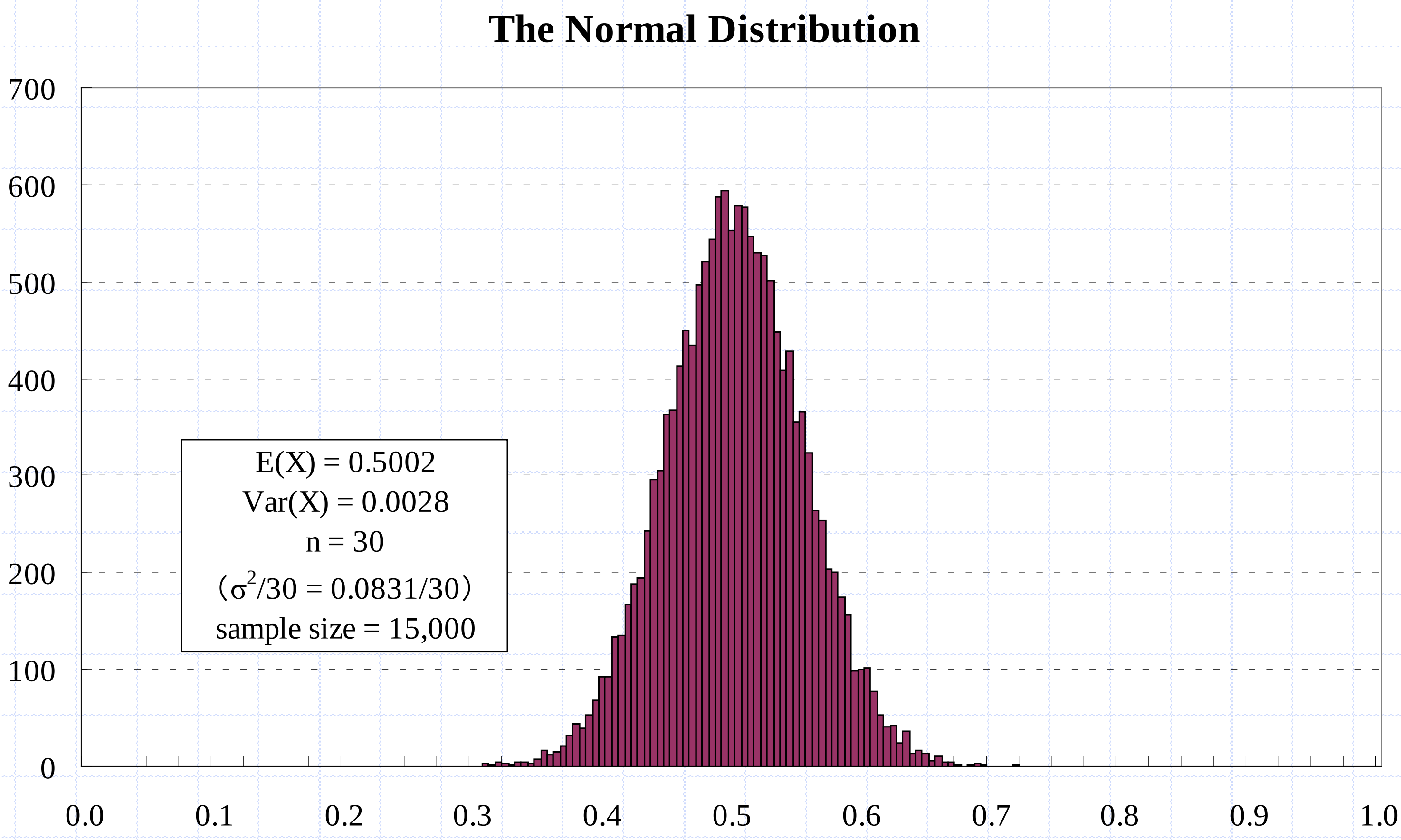
2. Central Limit Theorem

- For ***any*** population with a mean μ and a variance σ^2, the ***sampling distribution of the means*** of all possible samples of size n will be approximately normally distributed, with larger sample size n.
- The mean of the sampling distribution equal to μ and the variance equal to σ^2/n.

CLT *The Population Distribution*



CLT *The Sample Distribution of Sample mean*



CLT 0 *the case of $X \sim N(\mu, \sigma^2)$*

- If ***a population follows the normal distribution***, the sampling distribution of the sample mean will also follow the normal distribution for ***any sample size***.
- To determine the probability a sample mean falls within a particular region, use:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \qquad z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

CLT 1 *the case of X~ non-normal dist.(μ, σ^2)*

- If the population ***isn't normally distributed*** (with known σ^2) and ***sample size is large***, the sample means will follow the normal distribution. (See the above figure.)

$$\bar{X} \xrightarrow{d} N(\mu, \frac{\sigma^2}{n}) \qquad \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1)$$

CLT 2 *the case of $X \sim N(\mu$, unknown $\sigma^2)$*

- If ***the population follows the normal distribution but σ^2 is unknown***, the sample means will follow the ***t*** distribution.
  - But ***with larger sample size*** (*at least n >30*), the sample means will follow the normal distribution.

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(d.f.) \qquad t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \xrightarrow{d} N(0,1)$$

CLT 3 *the case of X~ non-N（μ, unknown σ^2）*

- If the population ***isn't normally distributed with unknown σ^2 and sample size is large***, the sample means will follow the ***t*** distribution. (with larger sample size, the normal distribution)

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \xrightarrow{d} N(0,1)$$