The Distribution of the Sample Mean

↳ inferential procedure: use $\overline{X}$ the sample mean to predict the population mean μ.

<u>Proposition:</u> Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

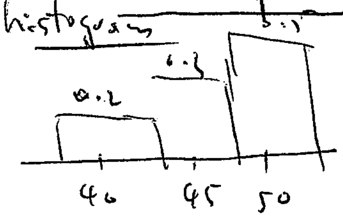
1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$
2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

In addition, with $T_o = X_1 + X_2 + \cdots + X_n$,

$E(T_o) = n\mu$, $V(T_o) = n\sigma^2$, $\sigma_{T_o} = \sqrt{n}\sigma$.

<u>Example 6.2</u>

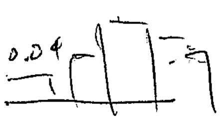
histogram



x	40	45	50
$p(x)$	0.2	0.3	0.5

$\mu = 46.5$, $\sigma^2 = 15.25$

$n = 2$



$\overline{x}$	40	42.5	45	47.5	50
$p(\overline{x})$	0.04	0.12	0.29	0.3	0.25

$\mu_{\overline{x}} = \mu$, $\sigma^2_{\overline{x}} = \dfrac{\sigma^2}{2}$

$n = 4$

$\overline{x}$	40	41.25	42.5	43.75	45	46.25	47.5	48.75	50
$p(\overline{x})$	0.0016	0.0096	0.0376	0.0936	0.1761	0.234	0.235	0.15	0.625

$\mu_{\overline{x}} = \mu$, $\sigma^2_{\overline{x}} = \dfrac{\sigma^2}{4}$

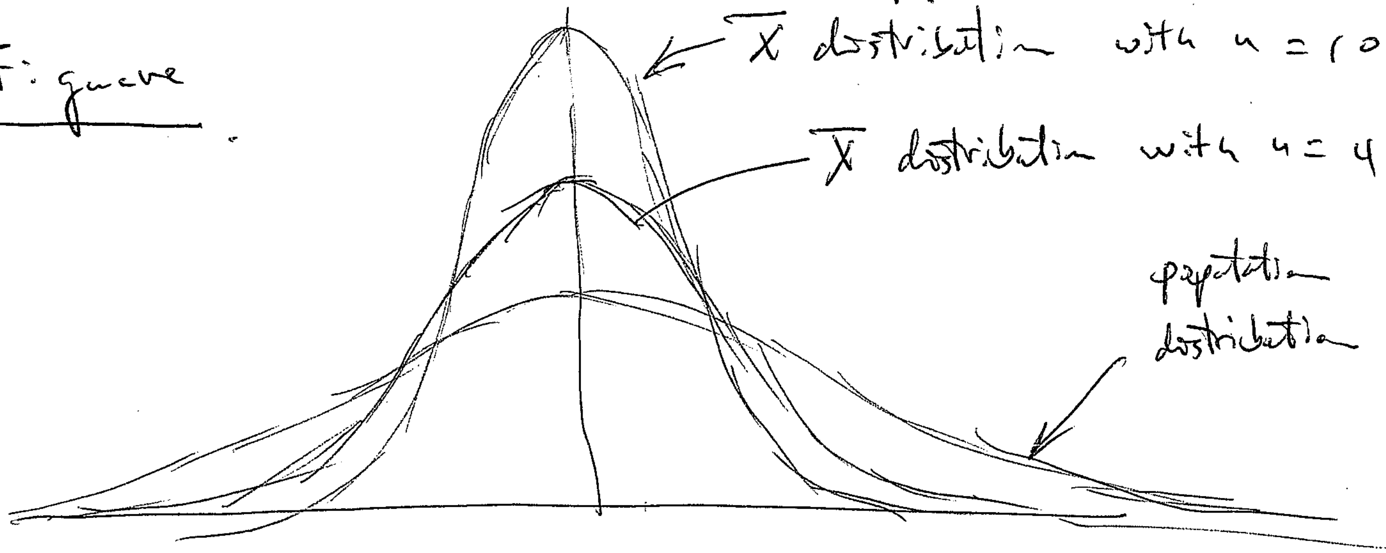
P 2 95

$n = 8$

The Case of a Normal Population Distribution

Proposition: Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal distribution with mean μ and standard deviation σ. Then for any n, $\overline{X}$ is normal distributed (with mean μ and standard deviation $\sigma/\sqrt{n}$), as is T_o (with mean $n\mu$ and standard deviation $\sqrt{n}\,\sigma$).

Figure.



Example: Let $X_1, X_2, \cdots, X_5$ be a random sample from $N(\mu=1.5,\ \sigma^2=0.35^2)$. Then $\overline{X} \sim N\left(\mu_{\overline{X}}=1.5,\ \sigma^2_{\overline{X}}=\frac{0.35^2}{5}\right)$

$T_o \sim N\left(\mu_{T_o}=1.5\times5 = 7.5,\ \sigma^2_{T_o}=0.35^2\times 5 = 0.6125\right)$ $\rightsquigarrow$ standardize T to compute its probability

$\sigma_{T_o} = 0.783$

$$P(6 \le T_o \le 8) = P\left(\frac{6-7.5}{0.783} \le Z \le \frac{8-7.5}{.773}\right) = \phi(0.64) - \phi(1.9$$

The Central Limit Theorem

- When X_i's are normally distributed, so is $\overline{X}$ for every sample size n.
- Even when the population distribution is highly non normal, averaging produces a distribution more bell-shaped than the one being sampled.
  → the formal statement of this result is the most important theorem of probability

Theorem. The Central Limit Theorem (CLT)

Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, in the limit as $n \to \infty$, the standardized versions of $\overline{X}$ and T_o have the standard normal distribution. That is

$$\lim_{n \to \infty} P\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \le z \right) = P(Z \le z) = \Phi(z)$$

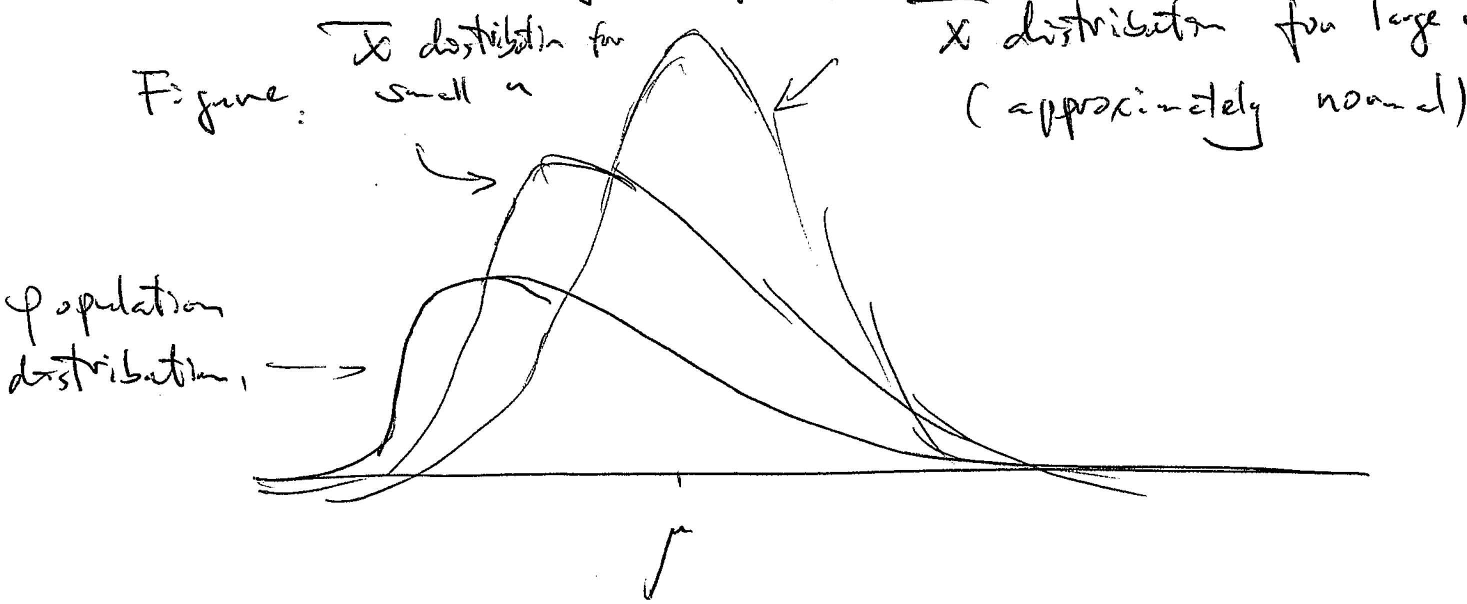
$$\lim P\left(\frac{T_o - n\mu}{\sqrt{n}\, \sigma} \le z \right) = P(Z \le z) = \Phi(z)$$

where Z is a standard normal r.v.

Remarks: 1°/. When n is sufficiently large, $\overline{X}$ has approximately a normal distribution with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \sigma^2/n$. Equivalently, for large n the sum T_0 has approximately a normal distribution with mean $\mu_{T_0} = n\mu$ and variance $\sigma^2_{T_0} = n\sigma^2$.

2°/ A partial proof of the CLT can be found in Chapter 6 - Appendix. It is shown that if the mgf exists, then the mgf of the standardized $\overline{X}$ and T_0 approach the standard normal mgf. With the aid of an advanced probability theorem, this implies the CLT statement about convergence of probabilities.

Figure:



<u>Rule of Thumb</u>: If $n > 30$, the Central Limit Theorem can be used.

<u>Example 6.9</u>: Let X = number of major defects for a new automobile model with mean $\mu = 3.2$ and standard deviation $\sigma = 2.4$.

Q: Among 100 randomly selected car of this model, how likely is that the sample average number of major defects exceeds 4?

A: $n = 100 > 30$, $\bar{X}$ does have approximately a normal distribution with mean $\mu_{\bar{X}} = 3.2$ and $\sigma_{\bar{X}} = \frac{1}{\sqrt{n}} \sigma = 0.24$.

$$P(\bar{X} > 4) \simeq P\left(Z > \frac{4 - 3.2}{0.24}\right) = 1 - \Phi(3.33)$$

$$= 0.0004.$$