# Chapter 6: Distribution of Sampling Statistics

ST2334 Probability and Statistics[1]
(Academic Year 2014/15, Semester 1)

Department of Statistics & Applied Probability (DSAP)
National University of Singapore (NUS)[1]

# Outline

# Introduction I

## Learning Outcomes

Questions to Address: ♦ View population & sample in a statistical problem as r.v.'s ♦ Mean & variance of the sample mean ♦ Using the empirical method to find a sampling distribution ♦ Using the central limit theorem to approximate the distribution of a sum/an average of i.i.d. r.v.'s ♦ Normal approximation to the binomial distribution with continuity correction ♦ Mean of the sample variance ♦ Some sampling distributions from a normal population

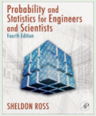
# Introduction II

## Learning Outcomes (continued)

**Concept & Terminology:** ♦ underlying distribution ♦ parametric/nonparametric inference problem ♦ statistic/sampling statistic ♦ sampling distribution of a statistic ♦ sample sum/mean/variance/standard deviation ♦ population mean/variance ♦ empirical method ♦ sampling distribution of the sample sum/mean/variance ♦ central limit theorem ♦ normal approximation to the binomial distribution with continuity correction ♦ joint sampling distribution of the sample mean & the sample variance from a normal population ♦ $t$-statistic constructed from a normal population

# Introduction III

## Mandatory Reading

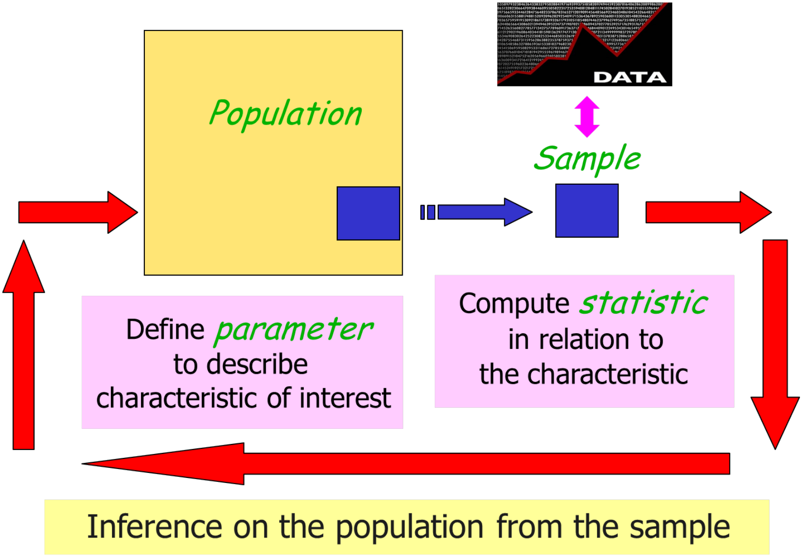



✏ Section 6.1 – Section 6.5

**The science of statistics deals with drawing conclusions from observed data**

- **Recall** (in Ch.0) the set–up of a *statistical problem*:
  - *population*: a large collection of items (each is associated with some values/measurements)
  - *sample/data*: only a subset of the population (*i.e.,* part of all measurements) is available

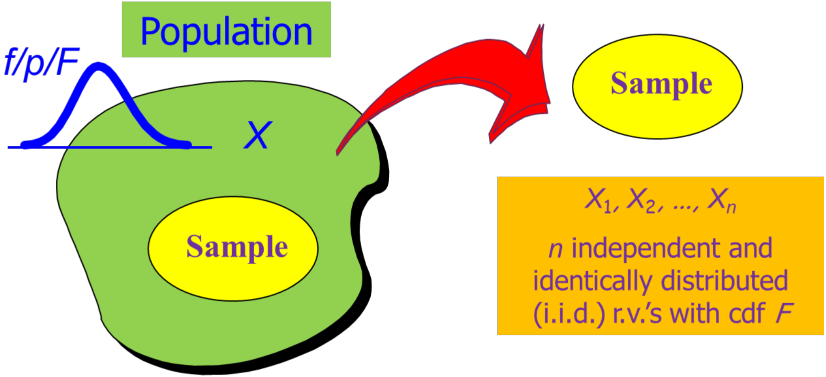
# Introduction IV



## *Set–up of Statistical Problems Via r.v.'s*

**Idea** :

- *population* $\Leftrightarrow$ r.v. $X$ with *unknown* cdf $F$ (& density $f/p$)
- *sample/data* $\Leftrightarrow$ i.i.d. r.v.'s, $X_1, \ldots, X_n$, with cdf $F$ (& density $f/p$)
- Refer to the distribution of $X$ or the cdf $F$ as the *underlying distribution*



# *Parametric Inference Problems I*

- **In practice**, we *often* select a particular *parametric family of probability distributions* *parametrized by some parameters* $\theta$ as the underlying distribution *F* of the population of interest, *e.g.*,
  1. Lifetime of a light bulb: an *exponential* r.v.
  2. annual revenue of Singapore Pools: a *normal* r.v.
  3. # of calls in an hour at a pizza delivery shop: a *Poisson* r.v.
- *Choosing with care appropriate values for* $\theta$ allows us to model/deal with most random phenomena or population of interest
- It *suffices to select values for* $\theta$ in the particularly selected family, *i.e.*,
  1. $\theta = \lambda \in (0, \infty)$ in the exponential case
  2. $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ in the normal case
  3. $\theta = \lambda \in (0, \infty)$ in the Poisson case

# *Parametric Inference Problems II*

**Definition**

A *parametric inference problem* has a set–up in which the underlying distribution of the population of interest is assumed to be from a particular *parametric family of probability distributions* which is *parametrized by some parameters* $\theta$

- When there is NOT any assumption made on the form of the cdf $F$, such a problem is called a *nonparametric inference problem*
- **Note** : This module *ONLY* looks at parametric inference problems

# *Sampling Statistics*

- **Recall** (in `Example 4 (Ch.0)`): In understanding $p$, the unknown proportion of the Singaporean population who have a a/c (parameter of interest $\theta = p$), we compute & study the proportion of people who have a a/c among a group of students (a sample)

## Definition

A *sampling statistic (or simply statistic)* summarizes relevant information from a sample in helping us to understand the population or the underlying distribution

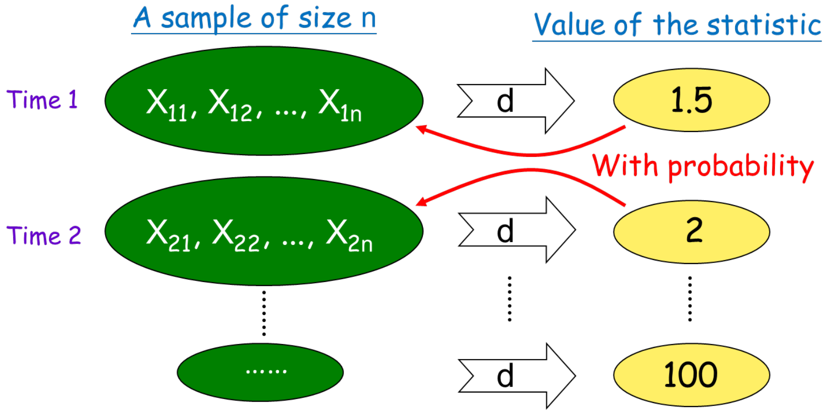
## Definition

In *parametric inference problems*, a *sampling statistic (or simply statistic)* summarizes relevant information to infer/guess the true but unknown value of the parameter $\theta$ of the underlying distribution

# *Sampling Statistics is a r.v.*

## Statistic as a r.v.

A statistic, defined by $d(X_1, \ldots, X_n)$, is a *r.v.* with possible values governed by the data/sample (*i.e.,* $X_1, \ldots, X_n$) through some *known/pre-determined function d* of $n$ i.i.d. r.v.'s



## *Sampling Distribution I*

**Definition**

The *probability distribution of a statistic* is called the *sampling distribution*

**In practice**, we observe

- $n$ known values, $x_1, \ldots, x_n$, as $n$ data points from a sample wherein $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ are $n$ i.i.d. realizations from the underlying distribution
- *ONLY 1 single value* $d(x_1, \ldots, x_n)$ called the *observed value* as a realization/draw from the sampling distribution of a statistic

# Sampling Distribution II

Suppose that we obtain 1.5 as the value of a statistic:

*How to interpret this number?*

*Is this number a good guess of the unknown value of the parameter? Close to it?*

**Address these questions by understanding the sampling distribution !**

# The Sample Mean $\overline{X}$ I

**Recall** (in `Example 19 (Ch.4)`): We define the sample mean $\overline{X}$ there. Here is an *alternative definition*

## Definition

1. A population of interest whose members or values can be regarded as being the values of a r.v. $X$ with cdf $F$
2. The mean $E(X) = \mu$ & variance $\text{Var}(X) = \sigma^2$ are called the *population mean* & the *population variance*, respectively
3. Let $X_1, \ldots, X_n$ be a sample from the population, *i.e., n i.i.d. r.v.'s* having cdf $F$
4. The *sample mean* is defined by

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

# *The Sample Mean $\overline{X}$ II*

- $\overline{X}$ is **1 of the most commonly used statistics** as it
  1. is defined as a function $d$ of $n$ i.i.d. r.v.'s from (1): $d$ is the average function
  2. serves as the *most appropriate* statistic which helps to *estimate/guess the unknown value of the mean $\mu$ of ANY population*

## Mean & Variance of The Sample Mean $\overline{X}$

The sample mean $\overline{X}$ is a r.v. with *mean* & *variance*

$$E(\overline{X}) = \mu \qquad \& \qquad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

- Shown in `Examples 19 & 26 (Ch.4)`
- $\overline{X}$ always has a "*centre*" at the population mean $\mu$ for all sample size $n$
- The *spread* of $\overline{X}$ $\downarrow$ as the sample size $n$ $\uparrow$

# Sampling Distribution of $\overline{X}$

- The *true sampling distribution of $\overline{X}$* is obtainable through *vigorous prob computations* that are rather tedious

This module would not consider it

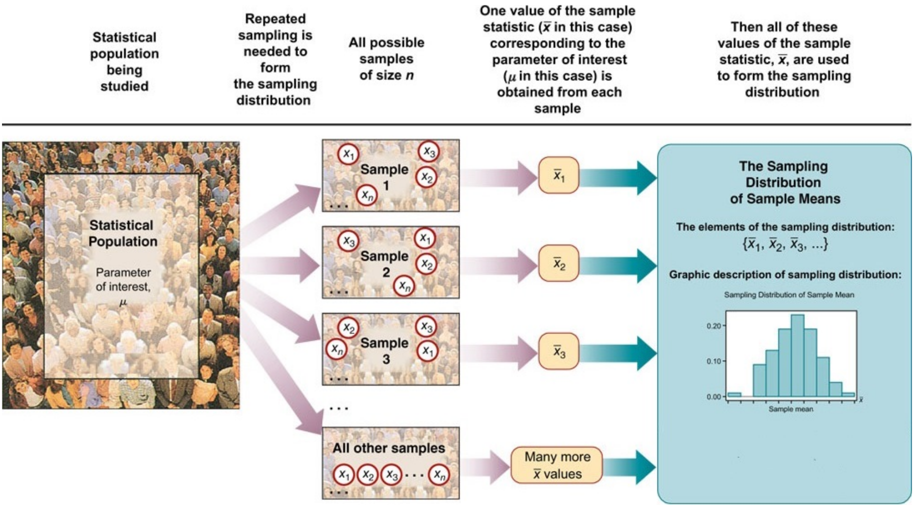
## Empirical Method

In principle, the sampling distribution of $\overline{X}$ (or, of any statistic) can be obtained by the *empirical method* (depicted at the next page):

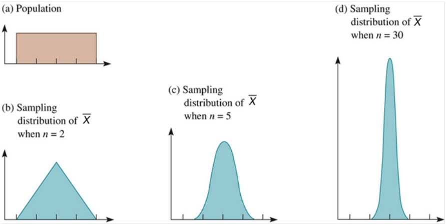
1. Collect *ALL possible samples of size n* from the population $X$
2. Compute a value $\overline{x}$ from each sample in ❶
3. Construct a *relative frequency histogram* of all $\overline{x}$ values in ❷

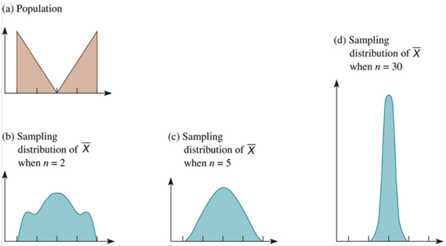
- The # of all possible samples of size $n$ in ❶ can be *astonomical* or even *uncountable infinite*
- An *approximate sampling distribution of $\overline{X}$*: Collect *a large # of samples of size n* in ❶, followed by ❷ & ❸

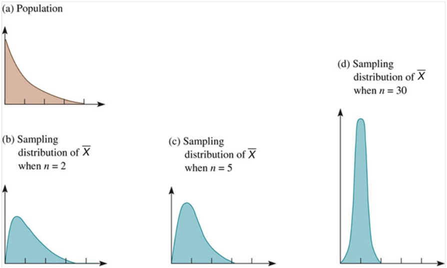
# The Empirical Method

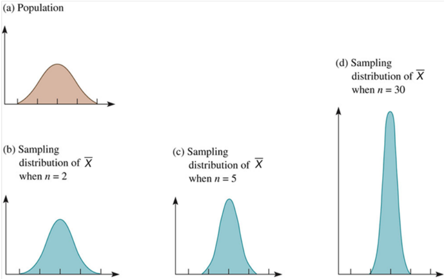


# Example 1: Sampling Distributions of $\overline{X}$



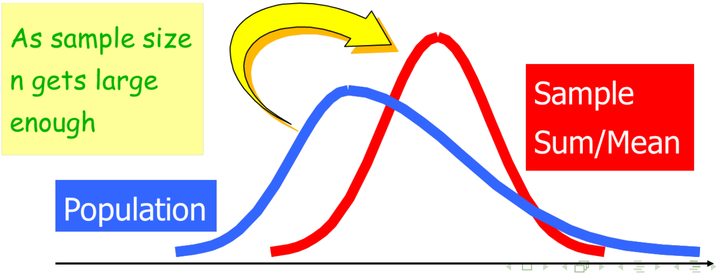



# *Central Limit Theorem I*

- The *central limit theorem (CLT)* is one of the *most remarkable results in probability theory*
- Loosely put, it states that the sum/average of a large # of indept r.v.'s has a distribution that is approximately normal
- It *not only* provides a simple method for computing approximate probs for sums of indept r.v.'s, but it *also* helps explain the remarkable fact that the empirical frequencies of so many natural populations exhibit bell–shaped (*i.e.,* normal) curves



# *Central Limit Theorem II*

## Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. r.v.'s having mean $\mu$ & variance $\sigma^2$. Then for $n \geq 30$, the distribution of the *sample sum*

$$X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

is *approximately normal with mean $n\mu$ & variance $n\sigma^2$*. Write

$$\sum_{i=1}^{n} X_i \dot{\sim} N(n\mu, n\sigma^2)$$

- $$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \dot{\sim} N(0,1) \Rightarrow \boxed{P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} < x\right) \approx \Phi(x)}$$

  *i.e.,* prob of a sum of i.i.d. r.v.'s can be *approximated by a standard normal prob when n is large*

# Approximate Sampling Distribution of $\overline{X}$

## Approximate Sampling Distribution of $\overline{X}$

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. r.v.'s having mean $\mu$ & variance $\sigma^2$. Then for $n \geq 30$, the distribution of

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is *approximately normal with mean $\mu$ & variance $\sigma^2/n$*. Write

$$\overline{X} \mathrel{\dot\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Translation of *approximate normality* through linear transformation: For $X \mathrel{\dot\sim} N(\mu, \sigma^2)$, $Y = a + bX \mathrel{\dot\sim} N(a + b\mu, b^2\sigma^2)$ for fixed constants $a$ & $b$

## Example 2: Central Limit Theorem

If a fair die is rolled 30 times, find the approximate prob that the sum obtained is between 100 & 110

**Solution:** Let $X_i$ be the # obtained at the $i$th roll of the fair die, for $i = 1, 2, \ldots, 30$. We have $\mu = E(X_i) = 7/2$, $E(X_i^2) = 91/6$, & therefore $\sigma^2 = \text{Var}(X_i) = 35/12$. The required prob is

$$P(100 \le X_1 + \cdots + X_{30} \le 110)$$

Apply CLT based on $X_1 + \cdots + X_{30} \dot\sim N(n\mu, n\sigma^2) = N\left(30 \times \frac{7}{2}, 30 \times \frac{35}{12}\right)$:

$$\begin{aligned}
& P(100 \le X_1 + \cdots + X_{30} \le 110) \\
\approx\ & P\left(\frac{100 - 30 \times (7/2)}{\sqrt{30 \times (35/12)}} \le Z \le \frac{110 - 30 \times (7/2)}{\sqrt{30 \times (35/12)}}\right) \\
=\ & P(-.53 \le Z \le .53) \\
=\ & .4038
\end{aligned}$$

# Example 3: Central Limit Theorem I

The # of students who enrol in a psychology class is a Poisson r.v. with mean 100. The professor in charge of the course decided that if the number of enrollment is $\geq 120$, he will teach the course in 2 separate sessions, whereas if the enrollment is under 120 he will teach all the students in a single session. What is the prob that the professor will have to teach 2 sessions?

**Solution:** Let $X$ be the enrollment in the psychology class. Given that $X \sim Poi(100)$ with $E(X) = 100 = \text{Var}(X)$. The required prob,

$$P(X \geq 120) = e^{-100} \sum_{i=120}^{\infty} \frac{100^i}{i!},$$

is not readily available by hand

*Remark*: Using a simple spreadsheet in computer, this prob is computed to be $1 - .9718 = .0282$

## Example 3: *Central Limit Theorem II*

Alternatively, realize that

$$X = X_1 + \cdots + X_{100}$$

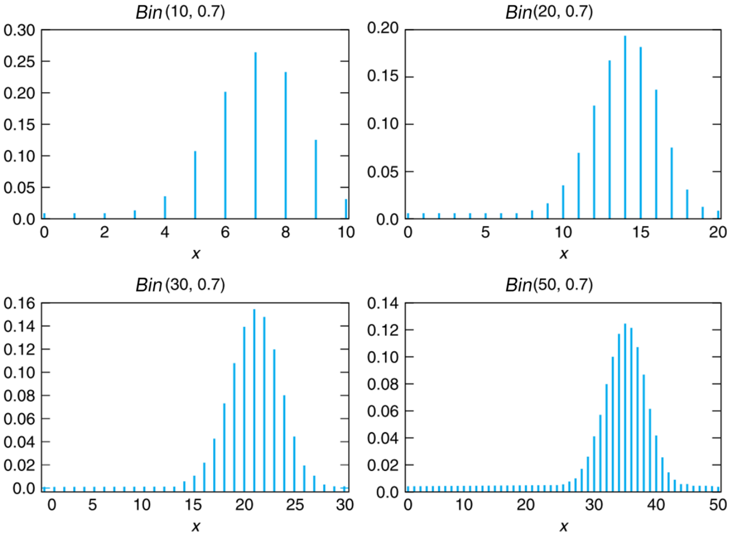
where $X_i$ are i.i.d. $Poi(1)$ r.v.'s with $\mu = \sigma^2 = 1$, & apply CLT to conclude that

$$X \mathrel{\dot\sim} N(n\mu = 100, n\sigma^2 = 100) \quad \Rightarrow \quad \frac{X - 100}{\sqrt{100}} \mathrel{\dot\sim} N(0, 1)$$

Then,

$$\begin{aligned} P(X \geq 120) &= P\left(\frac{X - 100}{\sqrt{100}} \geq \frac{120 - 100}{\sqrt{100}}\right) \\ &\approx P\left(Z \geq \frac{20}{\sqrt{100}}\right) \\ &= P(Z \geq 2) \\ &= .0228 \end{aligned}$$

# The Normal Approximation to the Binomial Distribution I




**Binomial pmf converging to a bell–shaped curve when $n \uparrow$**

# The Normal Approximation to the Binomial Distribution II

- Normal approximation to the binomial distribution
  - 1 extremely important application of CLT: A $Bin(n, p)$ r.v. is a sum of $n$ i.i.d. $Ber(p)$ r.v.'s when the *underlying distribution is a Bernoulli r.v.*
  - such an approximation by CLT is generally quite good for values of *$n$ satisfying $npq \geq 10$*, & it will be further improved if we incorporate *continuity correction (cc)*

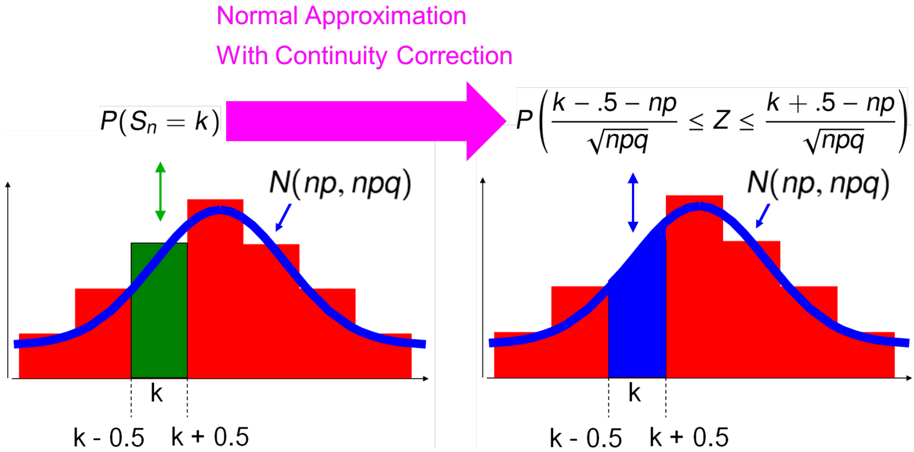
### Normal Approximation to Binomial Probabilities With Continuity Correction

If $S_n \sim Bin(n, p)$ with $q = 1 - p$, then, for $k = 0, 1, \ldots, n$,

$$P(S_n = k) = P\left(k - .5 \leq S_n \leq k + .5\right)$$
$$\approx P\left(\frac{k - .5 - np}{\sqrt{npq}} \leq Z \leq \frac{k + .5 - np}{\sqrt{npq}}\right)$$

- $P(S_n \geq k) = P\left(S_n \geq k - .5\right)$ & $P(S_n \leq k) = P\left(S_n \leq k + .5\right)$

# *The Normal Approximation to the Binomial Distribution III*



## Example 4: Normal Approximation

Let $X$ be a binomial r.v. with parameters 60 & .3

1. $P(12 \le X \le 26) \approx P\left(\frac{11.5-18}{\sqrt{12.6}} \le Z \le \frac{26.5-18}{\sqrt{12.6}}\right)$
   $\approx P(-1.83 \le Z \le 2.39) = .9916 - (1 - .9664) = .9580$
2. $P(12 < X \le 26) = P(13 \le X \le 26)$
   $\approx P\left(\frac{12.5-18}{\sqrt{12.6}} \le Z \le \frac{26.5-18}{\sqrt{12.6}}\right) \approx P(-1.55 \le Z \le 2.39)$
   $= .9916 - (1 - .9394) = .9310$
3. $P(12 \le X < 26) = P(12 \le X \le 25)$
   $\approx P\left(\frac{11.5-18}{\sqrt{12.6}} \le Z \le \frac{25.5-18}{\sqrt{12.6}}\right) \approx P(-1.83 \le Z \le 2.11)$
   $= .9826 - (1 - .9664) = .9490$
4. $P(12 < X < 26) = P(13 \le X \le 25)$
   $\approx P\left(\frac{12.5-18}{\sqrt{12.6}} \le Z \le \frac{25.5-18}{\sqrt{12.6}}\right) \approx P(-1.55 \le Z \le 2.11)$
   $= .9826 - (1 - .9394) = .9220$

# Example 5: Normal Approximation

Let $X$ be the # of times that a fair coin, flipped 40 times, lands heads. Suppose that we are interested in seeing heads half of the time. Compute the prob exactly & approximate it

**Solution:** *Exact answer*: $P(X=20)=\binom{40}{20}\left(\frac{1}{2}\right)^{40}=.1254$

*Normal approximation*:

$$\begin{aligned}
P(X=20) &= P(19.5 \le X \le 20.5) \\
&= P\left(\frac{19.5-20}{\sqrt{10}} \le \frac{X-20}{\sqrt{10}} \le \frac{20.5-20}{\sqrt{10}}\right) \\
&\approx P(-0.16 \le Z \le 0.16) = .1272
\end{aligned}$$

# Example 6: Normal Approximation

The ideal size of a first–year class at a particular college is 150 students. The college, knowing from the past experience that on the average only 30% of those accepted for admission to this class will actually attend, uses a policy of approving the applications of 450 students. Compute the prob that $>$ 150 students attend this class

**Solution:** Let $X$ denote the # of students that attend among the 450 students. Then, $X \sim Bin(450, .3)$ with mean & variance

$$E(X) = 450 \times .3 = 135 \qquad \& \qquad \mathrm{Var}(X) = 450(.3)(.7) = 9.721^2$$

Applying normal approximation with cc, the prob that $>$ 150 students attend this class is

$$\begin{aligned} P(X > 150) &= P(X \geq 150.5) \approx P\left(Z \geq \frac{150.5 - 135}{9.721}\right) \\ &= P(Z \geq 1.59) = .0559 \end{aligned}$$

## Example 7: Normal Approximation I

Refer to the previous example, if this college desires that the prob of $> 150$ students will attend this college should be at most .01. What is the largest # of students should this college admit?

**Solution:** Let $X$ denote the # of students that attend. Let $n$ be the # of students to be admitted. Then, $X \sim Bin(n, .3)$ with mean & variance

$$E(X) = .3n \qquad \& \qquad \mathrm{Var}(X) = .3n \times .7 = .21n$$

Applying normal approximation with cc yields the prob that $> 150$ students will attend the college,

$$\begin{aligned} P(X > 150) &= P(X \geq 150.5) \\ &\approx P\left(Z \geq \frac{150.5 - .3n}{\sqrt{.21n}}\right) \end{aligned}$$

# Example 7: *Normal Approximation II*

To have this probability to be at most .01, set it to be less than or equal to .01. This yields

$$P\left(Z \geq \frac{150.5 - .3n}{\sqrt{.21n}}\right) \leq P(Z \geq 2.33) = .01$$

Accordingly

$$\frac{150.5 - .3n}{\sqrt{.21n}} \geq 2.33 \qquad \Rightarrow \qquad n \leq 428.32$$

Hence, the largest # of students this college should admit is 428

# The Sample Variance I

## Definition

Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s having mean $\mu$ & variance $\sigma^2$, & $\overline{X} = \sum_{i=1}^{n} X_i/n$ be the *sample mean*. The *sample variance* is a statistic defined by

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$S = +\sqrt{S^2}$ is called the *sample standard deviation*

## Mean of the Sample Variance Equals the Population Variance $\sigma^2$

The expected value/mean of the sample variance $S^2$ always equals $\sigma^2$:

$$E(S^2) = \sigma^2$$

- Serves as an *appropriate statistic* in *estimating the unknown value of the variance of ANY population*

# The Sample Variance II

- Start with the following *important algebraic identity of $S^2$*:

$$
\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 + \sum_{i=1}^{n}(\overline{X}-\mu)^2 - 2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 + n(\overline{X}-\mu)^2 - 2(\overline{X}-\mu)n(\overline{X}-\mu) \\
&= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2
\end{aligned}
$$

Taking expectation of both sides yields

$$
\begin{aligned}
(n-1)E(S^2) &= \sum_{i=1}^{n} E[(X_i-\mu)^2] - nE[(\overline{X}-\mu)^2] \\
&= n\sigma^2 - n\mathrm{Var}(\overline{X}) = (n-1)\sigma^2 \quad \Rrightarrow \quad E(S^2) = \sigma^2
\end{aligned}
$$

# Sampling Distributions From a Normal Population I

Let's look at a *special case* in which the *underlying distribution is normal*

## Sampling Distribution of $\overline{X}$ From a Normal Population

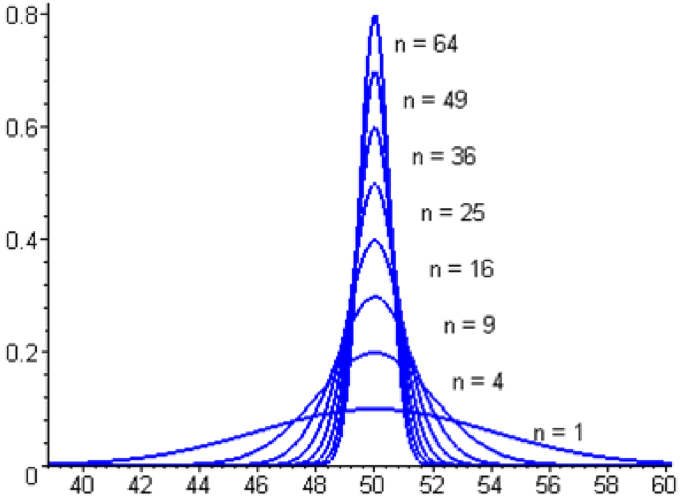
Let $X_1, \ldots, X_n$ be a sample from a *normal population* with mean $\mu$ & variance $\sigma^2$. Then,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{or} \qquad \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \tag{2}$$

- *Valid* for all sample sizes $n = 1, 2, 3, \ldots$
- Due to "linear transformation" result at `Page 29 (Ch.5)`: For $X \sim N(\mu, \sigma^2)$, $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$ for fixed constants $a$ & $b$

# *Sampling Distributions From a Normal Population II*

Sampling distribution of sample means from $N(50, 4^2)$:



# Sampling Distributions From a Normal Population III

## Joint Sampling Distribution of $\overline{X}$ & $S^2$ From a Normal Population

For a sample $X_1, \ldots, X_n$ from a *normal population* with mean $\mu$ & variance $\sigma^2$, the *joint sampling distribution of $(\overline{X}, S^2)$* is described as follows:

1. $\overline{X}$ & $S^2$ are indept r.v.'s
2. $\overline{X}$ has a distribution given in (2)
3. $\dfrac{(n-1)}{\sigma^2} S^2 \sim \chi^2_{n-1}$

- A brief discussion of ❶:
  - $S^2$ is a function of all $n$ deviations from $\overline{X}$, $X_i - \overline{X}$, which are all normally distributed r.v.'s
  - $\text{Cov}(X_i - \overline{X}, \overline{X}) = 0$ from Examples 25 (Ch.2)

    $\Rrightarrow$ indep of $X_i - \overline{X}$ & $\overline{X}$ ($\because$ normality of the 2 r.v.'s)

# Sampling Distributions From a Normal Population IV

- A streamline proof of ❸:
  - Note that
    $$W = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$$
    & that $W$ can be re-expressed as
    $$\frac{1}{\sigma^2}\sum_{i=1}^{n}[(X_i-\overline{X})+(\overline{X}-\mu)]^2 = \frac{(n-1)}{\sigma^2}S^2 + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = U+V$$
    by expanding the square in the LHS & using the fact that $\sum_{i=1}^{n}(X_i-\overline{X})=0$
  - $U$ & $V$ are indept ($\because$ $S^2$ & $\overline{X}$ are indept)
  - $W\sim\chi_n^2$ & $V\sim\chi_1^2$ $\Rrightarrow$ $U\sim\chi_{n-1}^2$

# Sampling Distributions From a Normal Population V

## A $t$-Statistic Constructed by $\overline{X}$ & $S^2$ From a Normal Population

Let $\overline{X}$ & $S^2$ be the sample mean & the sample variance of a sample from a *normal population* with mean $\mu$ & variance $\sigma^2$. Then,

$$\frac{\overline{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1} \tag{3}$$

- Follows from definition of a $t$-distribution by expressing the ratio as

$$\frac{\overline{X} - \mu}{\sqrt{S^2/n}} = \frac{\left(\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)}{\sqrt{S^2/\sigma^2}} = \frac{\left(\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}}$$

where the r.v. at the numerator is $N(0, 1)$ & the r.v. at the denominator is $\sqrt{\chi^2_{n-1}/(n-1)}$, & also the 2 r.v.'s are indept

# Example 8: Sampling Distributions From a Normal Population I

Suppose that the first monthly salary (in S$) of a undergraduate in a university is known to be approximately $N(2000, 900)$. For a selected group of 15 students, what are the probs that

1. the average first monthly salary is greater than S$2010?
2. the standard deviation of their first monthly salaries is greater than S$10?

**Solution:** Let $X_1, \ldots, X_{15}$ denote the first monthly salaries of the group of 15 students. Then, $X_1, \ldots, X_{15}$ are i.i.d. $N(2000, 900)$ r.v.'s

1. The required prob is given by

$$P(\overline{X} > 2010) = P(Z > \frac{2010 - 2000}{\sqrt{900/15}}) \approx P(Z > 1.29) = .0985$$

where $\overline{X} = \frac{1}{15}\sum_{i=1}^{15} X_i \sim N(2000, 900/15)$ by (2)

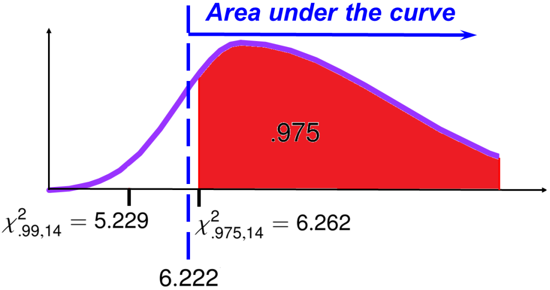
# Example 8: Sampling Distributions From a Normal Population II

2. The required prob is given by

$$P(S > 20) = P(S^2 > 400) = P\left(\chi^2_{14} > \frac{(15-1)(400)}{900}\right)$$

$$= P(\chi^2_{14} > 6.222) \in (.975, .99)$$

where $\dfrac{(n-1)}{\sigma^2}S^2 = \dfrac{(15-1)}{900}S^2 \sim \chi^2_{14}$



## Example 9: $t$ Statistic From a Normal Population

Consider the last example by assuming that the first monthly salary (in S$) of a undergraduate in a university is approximately $N(2000, \sigma^2)$ with an unknown $\sigma^2 > 0$. For a selected group of 15 students with sample standard deviation of their first monthly salaries as 35, what is the prob that their average first monthly salary is greater than S$2010?

**Solution:** Let $X_1, \ldots, X_{15}$ denote the first monthly salaries of the group of 15 students. Then,

1. $X_1, \ldots, X_{15}$ are i.i.d. $N(2000, \sigma^2)$ r.v.'s, and
2. $(\overline{X} - 2000)/\sqrt{S^2/15} \sim t_{15-1}$

The required prob is given by

$$P(\overline{X} > 2010) = P(t_{14} > \frac{2010 - 2000}{\sqrt{35^2/15}}) \approx P(t_{14} > 1.11) > .10$$

as $t_{.10,14} = 1.345$ from the $t$-table