Lecture 5. Sampling Distribution Theory (Chapter 6)

Ping Yu

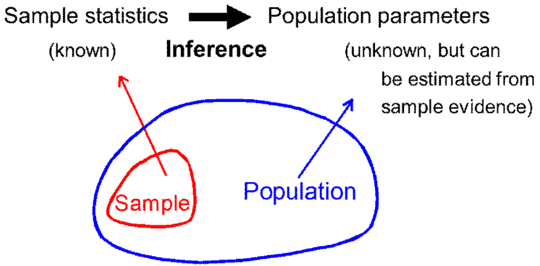
HKU Business School
The University of Hong Kong

Plan of This Lecture

- Sampling from a Population
- Sampling Distributions of Sample Means
- Sampling Distributions of Sample Proportions
- Sampling Distributions of Sample Variances
- Properties of Point Estimators (Section 7.1)

[Review] Descriptive and Inferential Statistics

- Descriptive statistics: collecting, presenting, and describing data.
- Inferential statistics: drawing conclusions and/or making decisions concerning a population based only on sample data. [figure here]
  - Estimation, e.g., estimate the population mean weight using the sample mean weight.
  - Hypothesis Testing, e.g., use sample evidence to test the claim that the population mean weight is 120 pounds.



Sampling from a Population

Population and Simple Random Sample

- Statistical analysis requires that we obtain a proper sample from a population of items of interest that have measured characteristics.
- Recall that a population means all (say, N) items of interest.
  - If N is large enough, N can be treated as ∞.
  - A population is generated by a process that can be modeled as a series of random experiments (see Lecture 2).
- A (simple) random sample is a sample of n objects drawn randomly.
  - Recall the definition of random sampling in Lecture 1.
- Random sampling with replacement means drawing a member from the population by chance (i.e., with probability $1/N$), putting it back to the population, and then independently drawing the next one.
  - This is the random sampling in Lecture 1.

continue

- Random sampling without replacement means randomly drawing each group of n **distinct** items with probability $1/C_n^N$, which seems easier in practice.
  - The first item is sampled with probability $1/N$; conditional on the first item was chosen, the second item is sampled with probability $1/(N-1)$, etc.

Why Sample?

- Less time consuming than a census.
- Less costly to administer than a census.
- It is possible to obtain statistical results of a sufficiently high precision based on samples.

- Samples can be obtained from a table of random numbers or computer random number generators.

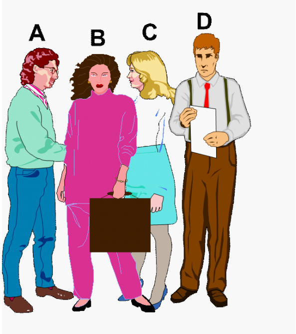
Sampling Distributions

- The randomness of a random sample comes from the random drawing, i.e., not all items are drawn ($n < N$) so the identities of the random sample are not determined beforehand.
- Let X be the population r.v. taking each value in $\{x_i\}_{i=1}^N$ with probability $1/N$, and $\{x_i\}_{i=1}^n$ be a random sample.[1]
- The population mean $\mu = E[X] = \frac{1}{N}\sum_{i=1}^N x_i$, and the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ is a natural estimator of μ.
- The population variance $\sigma^2 = E\left[(X-\mu)^2\right] = \frac{1}{N}\sum_{i=1}^N (x_i-\mu)^2$, and the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2$ is a natural estimator of σ^2. [The reason of $n-1$ instead of n will be explained below]
  - The population counterpart of s^2 should be $S^2 = \frac{1}{N-1}\sum_{i=1}^N (x_i-\mu)^2$.
  - The sample standard deviation is $s = \sqrt{s^2}$.
- The sampling distribution of a statistic such as the sample mean and sample variance is the probability distribution obtained from all possible samples of the same number of observations drawn from the population.

[1] The textbook uses $\{X_i\}_{i=1}^n$ to emphasize the randomness of x_i, but the notations are not consistent. In my lectures, you can tell from the context whether $\{x_i\}_{i=1}^n$ are random or just realizations.

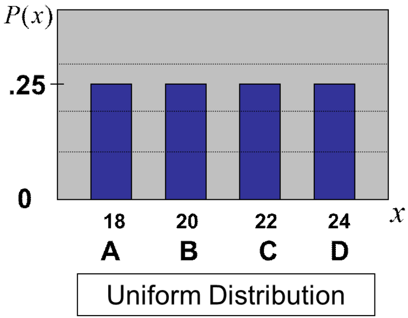
Development of a Sampling Distribution

- Assume there is a population$\cdots$
- Population size $N = 4$.
- Random variable, X, is age of individuals.
- Values of X: 18, 20, 22, 24 (years).



continue

- In this example the Population Distribution is uniform:



continue

- Now consider all possible samples of size $n = 2$:

1st Obs	2nd Observation: 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation: 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

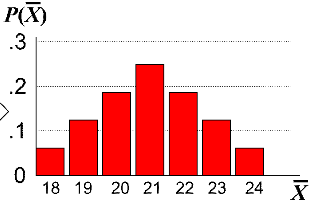
continue

- Sampling Distribution of All Sample Means:

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Distribution of Sample Means



(no longer uniform)

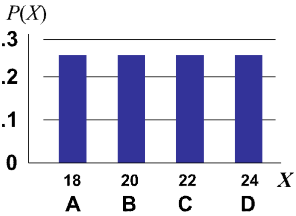
- Notation: $P\left(\bar{X}\right)$ should be $p\left(\bar{x}\right)$ in our notations.

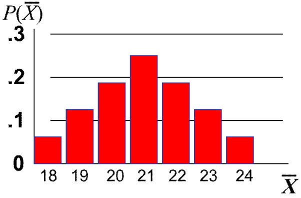
- When N is large, it is impossible to list all possible outcomes, so the abstract analysis in the next section is helpful.

Comparing the Population with Its Sampling Distribution

Population
$N = 4$
$\mu = 21, \sigma = 2.236$

Sample Means Distribution
$n = 2$
$\mu_{\bar{x}} = 21, \sigma_{\bar{x}} = 1.58$





- $\mu = \frac{\sum x_i}{N} = \frac{18+20+22+24}{4} = 21$, and $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} = 2.236$.
- $\mu_{\bar{x}} = E[\bar{x}] = \frac{\sum \bar{x}_i}{N} = \frac{18+19+21+\cdots+24}{16} = 21 = \mu$, and
  $\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{N}(\bar{x}_i - \mu_{\bar{x}})^2}{N}} = \sqrt{\frac{(18-21)^2+(19-21)^2+\cdots+(24-21)^2}{16}} = 1.58 = \frac{2.236}{\sqrt{2}} = \frac{\sigma}{\sqrt{n}}$.

Sampling Distributions of Sample Means

Mean of the Sample Means

- For random sampling with replacement,

$$E[\bar{x}] = E\left[\frac{1}{n}(x_1 + \cdots + x_n)\right] = \frac{n\mu}{n} = \mu.$$

- In random sampling with replacement, x_i (for each i) and X have the same distribution because x_i takes each value in $\{x_j\}_{j=1}^N$ with probability $1/N$, which is exactly the distribution of X. [check the $N = 4$ and $n = 2$ example above]
- This means that if we draw n samples repeatedly, and for each draw we calculate $\bar{x}$, then the average of these $\bar{x}$'s is the population mean.
- A particular $\bar{x}$ value can be considerably far from μ.
- Here, x_i is treated as a r.v. rather than a realization.

- (*) For random sampling without replacement,

$$E[\bar{x}] = \frac{1}{n}\frac{1}{C_n^N}\sum_{j=1}^{C_n^N}\left(\sum_{i=1}^{n} x_i^j\right) = \frac{1}{n}\frac{1}{C_n^N}C_{n-1}^{N-1}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{i=1}^{N} x_i = \mu,$$

where x_i^j is the ith draw in the jth sampling, and the second equality is from $nC_n^N = NC_{n-1}^{N-1}$ (intuition? for rigorous analysis, see the next[2] slide).

Variance of the Sample Means

- For random sampling with replacement,

$$Var\left(\bar{x}\right)=Var\left(\frac{1}{n}x_1+\cdots+\frac{1}{n}x_n\right)=\sum_{i=1}^{n}\left(\frac{1}{n}\right)^2\sigma_i^2=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}.$$

  - $Var\left(\bar{x}\right)$ decreases with n, i.e., larger sample sizes result in more concentrated sampling distributions.
  - Denote $Var\left(\bar{x}\right)$ as $\sigma_{\bar{x}}^2$; then the standard deviation of $\bar{x}$ is $\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$.

- (*) For random sampling without replacement,

$$Var\left(\bar{x}\right)=\frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}=\frac{S^2}{n}\cdot\frac{N-n}{N}=S^2\left(\frac{1}{n}-\frac{1}{N}\right).$$

  - Why? the variances of a hypergeometric distribution and a binomial distribution are $np\left(1-p\right)\frac{N-n}{N-1}$ and $np\left(1-p\right)$, respectively. The difference term $\frac{N-n}{N-1}$ appears due to the same reason as here. [for more details, see the next slide]

(**) Rigorous Analysis for Random Sampling Without Replacement

- Note that $\bar{x} = \frac{1}{n}\sum_{i=1}^{N} R_i x_i$, where R_i's are exchangeable Bernoulli r.v.'s, and $\sum_{i=1}^{N} R_i = n$, so

$$P\left((R_1,\cdots,R_N) = (r_1,\cdots,r_N)\right) = \frac{1}{C_n^N} = \frac{n!(N-n)!}{N!}$$

  for any $(r_1,\cdots,r_N)$ such that $\sum_{i=1}^{N} r_i = n$.
- Conditional on n, $E[R_i] = \frac{n}{N}$ because $\sum_{i=1}^{N} E[R_i] = N \cdot E[R_i] = n$, and $Var(R_i) = \frac{n}{N}\left(1 - \frac{n}{N}\right)$.
- However, R_i's are not independent. To study their covariances, note that for $1 \leq j < k \leq N$,

$$0 = Var\left(\sum_{i=1}^{N} R_i\right) = N \cdot Var(R_i) + N(N-1)\,Cov\left(R_j, R_k\right),$$

  so

$$Cov\left(R_j, R_k\right) = -\frac{Var(R_i)}{N-1} = -\frac{n}{N(N-1)}\left(1 - \frac{n}{N}\right) < 0,$$

  and

$$E\left[R_j R_k\right] = Cov\left(R_j, R_k\right) + E\left[R_j\right]E\left[R_k\right] = \frac{n(n-1)}{N(N-1)}, \qquad (1)$$

  where $E\left[R_j R_k\right]$ will be used in the future.

(**) continue

- First,

$$E\left[\bar{x}\right]=\frac{1}{n}\sum_{i=1}^{N}E\left[R_i\right]x_i=\frac{1}{N}\sum_{i=1}^{N}x_i=\mu.$$

- After some calculation, we can show

$$NS^2=\sum_{i=1}^{N}x_i^2-\frac{2}{N-1}\sum_{i=1}^{N}\sum_{j=i+1}^{N}x_ix_j. \tag{2}$$

- Therefore,

$$\begin{aligned}Var\left(\bar{x}\right) &= \frac{1}{n^2}\sum_{i=1}^{N}Var\left(R_i\right)x_i^2+\frac{2}{n^2}\sum_{i=1}^{N}\sum_{j=i+1}^{N}Cov\left(R_i,R_j\right)x_ix_j \\ &= \frac{1}{n^2}Var\left(R_1\right)\left(\sum_{i=1}^{N}x_i^2-\frac{2}{n-1}\sum_{i=1}^{N}\sum_{j=i+1}^{N}x_ix_j\right) \\ &= \frac{1}{n^2}Var\left(R_1\right)NS^2=\frac{S^2}{n}\left(1-\frac{n}{N}\right).\end{aligned}$$

Finite Population Correction Factor

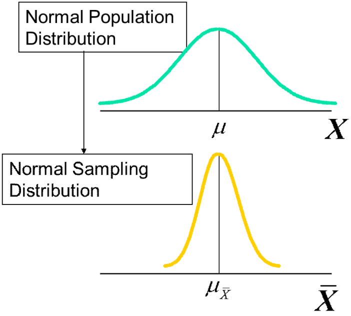
- $\frac{N-n}{N-1}$ is often called a finite population correction factor.
  - When N is large, the differences between the two random sampling schemes can be neglected: $\frac{N-n}{N-1} \to 1$ as $N \to \infty$ and $\frac{n}{N} \to 0$.
  - In business applications such as auditing, N is indeed not large.
  - n rather than the fraction of the sample in the population, $\frac{n}{N}$, is the dominant factor of $Var(\bar{x})$.
- Without special mention, we always mean random sampling with replacement or without replacement but N is large enough and n is a small proportion of N.

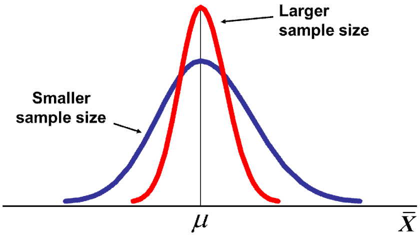
Sampling Distribution of the Sample Means

- If the population follows the normal distribution, then $\bar{x}$ follows a normal distribution $N\left(\mu, \frac{\sigma^2}{n}\right)$ since it is a linear combination of x_i's which follow the same normal distribution as the population.
  - Implicitly, $N = \infty$ because the normal distribution is continuous.
  - Recall that a normal distribution is determined only by its mean and variance.
  - Both X and $\bar{x}$ follow the normal distribution with the same mean, but $\bar{x}$ has a smaller variance (more so as n gets larger). [figure here]
- The standardized normal random variable

$$z = \frac{\bar{x} - E\left[\bar{x}\right]}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1). \quad (3)$$

- Terminology: the standard error (SE) of a statistic (usually an estimate of a parameter) is the standard deviation of its sampling distribution or an estimate of that standard deviation.
  - In our case, $\sigma / \sqrt{n}$ and $s / \sqrt{n}$ are both called the SE of $\bar{x}$.
  - Usually, only the latter is called the SE of $\bar{x}$ because it is feasible and the former already has a name – standard deviation.


Normal Population Distribution
μ
X
Normal Sampling Distribution
$\mu_{\bar{X}}$
$\bar{X}$


Larger sample size
Smaller sample size
μ
$\bar{X}$

Example 6.3: Spark Plug Life

- A spark plug manufacturer claims that the lives of its plugs follow $N\left(60,000, 4000^2\right)$. If we observed that the sample mean of a random sample of size 16 is 58,500 miles. Do you think the manufacturer's claim is credible?
- Since

$$P\left(\bar{x} \leq 58,500\right) = P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq \frac{58,500-60,000}{4000/\sqrt{16}}\right) = P\left(z \leq -1.50\right) = .0668,$$

which is quite small, so the claim of the manufacturer is skeptical.

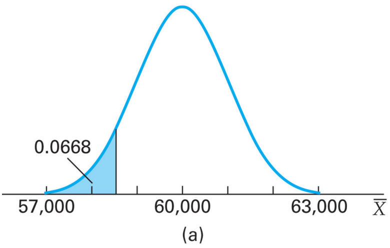


Figure: (a) $P\left(\bar{x} \leq 58,500\right)$; (b) $P\left(z \leq -1.50\right)$

The Law of Large Numbers (LLN)

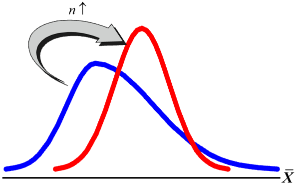
- Without normality, what is the distribution of $\bar{x}$? When n is fixed, there is no tractable description in general, while when n is large, we can say something.
- First, the distribution of $\bar{x}$ will degenerate at μ.
- LLN: If x_i, $i = 1, \cdots, n$, are independent and identically distributed (i.i.d.) with mean μ (as in a random sample with replacement), then $\bar{x}$ approaches μ as $n \to \infty$.
  - Not $N \to \infty$.
  - Only requires $E[x_i] = \mu < \infty$, regardless of what the distribution of x_i is.
  - This is different from $E[\bar{x}] = \mu$ (which fixes n and repeatedly samples $\{x_i\}_{i=1}^n$): it claims that if for **any** random sample $\{x_i\}_{i=1}^\infty$, we calculate $\bar{x}$ for the first n samples to obtain a sequence of numbers, say $\bar{x}_n$, then $\bar{x}_n \to \mu$ as $n \to \infty$.
  - Intuitively, $\mu = \frac{1}{N}\sum_{i=1}^N x_i$ involves all values of $\{x_i\}_{i=1}^N$; in $E[\bar{x}] = \mu$, although n is fixed, we repeatedly sampled so that all values of $\{x_i\}_{i=1}^n$ would be sampled; in $\bar{x}_n \to \mu$, by letting $n \to \infty$, we potentially sampled all values $\{x_i\}_{i=1}^N$.
- (*) Rigorously, "approach μ" means "is consistent to μ", where consistency is defined in the Appendix of Chapter 7, Page 330.
- Jacob Bernoulli proved the first LLN with $\{x_i\}_{i=1}^n$ being Bernoulli trials (i.e., $x_i \overset{iid}{\sim} \text{Bernoulli}(p)$); the current form of LLN is attributed to Khinchin [figure here], so is called the Khinchin LLN.

History of the LLN

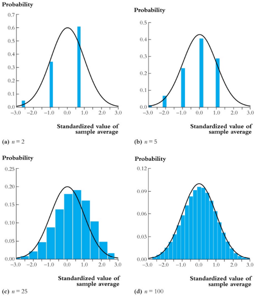
Aleksandr Khinchin (1894-1959),
Moscow State University

The Central Limit Theorem (CLT)

- CLT: If x_i, $i = 1, \cdots, n$, are i.i.d. with mean μ and variance σ^2, then the distribution of $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ approaches that of $N(0,1)$ as $n \to \infty$.
  - The result of CLT is stronger than that of LLN since it not only claims that $\bar{x}$ approaches μ, but also claims that the variance of $\bar{x}$ approaches σ^2/n (which is expected), and the standardized $\bar{x}$ is eventually bell-shaped as $n \to \infty$ (which is surprising). That is, the distribution of $\bar{x}$ not only degenerates at μ, but degenerates to μ in the rate $\sqrt{n}$ and in the bell shape.
  - Require $Var(x_i) = \sigma^2 < \infty$ besides $E[x_i] = \mu < \infty$, i.e., a stronger result need stronger assumptions.
  - It does not require x_i to be normally distributed. [figure here]
  - Intuitively, when n is large enough, the claim for the normally distributed x_i in (3) is roughly correct, or $\bar{x} \sim N(\mu, \sigma^2/n)$.
  - How large n is required for satisfactory approximation? If x_i is symmetrically distributed, then $n = 20$ to 25 is enough; otherwise, n needs to be much larger, e.g., > 50.
- The De Moivre-Laplace theorem is a special case of the CLT with $\{x_i\}_{i=1}^n$ being Bernoulli trials; the current form of CLT is attributed to Lindeberg and Lévy [figure here], so is called the Lindeberg-Lévy CLT.
  - As mentioned in the De Moivre-Laplace theorem, we can use a continuous r.v. to approximate a discrete r.v. when n is large.



As n increases, the sampling distribution of $\bar{x}$ becomes almost normal regardless of shape of population ($n = 1$)

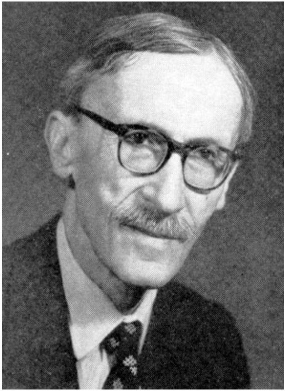


The Sampling Distribution of $\sqrt{n}(\bar{x} - \mu)/\sigma$ Compared with $N(0,1)$; x_i is discrete

Intuition: $\sqrt{n} \to \infty$, $\bar{x} - \mu \to 0$, but $\sqrt{n}(\bar{x} - \mu)$ will not diverge or degenerate!

History of the CLT

Jarl W. Lindeberg (1876-1932),
University of Helsinki

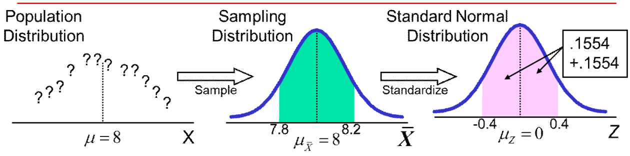


Paul P. Lévy (1886-1971),
École Polytechnique

[Example] Applying CLT

- Suppose a large population has mean $\mu = 8$ and $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected. What is the probability that the sample mean is between 7.8 and 8.2?
- Solution: Even if the population is not normally distributed, the central limit theorem can be used ($n > 25$), so the sampling distribution of $\bar{x}$ is approximately $N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(8, \frac{3^2}{36}\right)$.
- As a result,

$$
\begin{aligned}
P(7.8 < \bar{x} < 8.2) &= P\left(\frac{7.8-\mu}{\sigma/\sqrt{n}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-\mu}{\sigma/\sqrt{n}}\right) \\
&= P(-0.4 < z < 0.4) = 0.3108.
\end{aligned}
$$



(**) CLT for Random Sampling Without Replacement

- CLT for finite polulations: Consider a finite population $\{x_{N1}, \cdots, x_{NN}\}$ with N units. Define

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} x_{Ni},$$
$$S_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{Ni}-\mu_N)^2$$

and

$$m_N = \max_{1\le i\le N}(x_{Ni}-\mu_N)^2.$$

Assume that

$$\frac{1}{\min(n, N-n)}\frac{m_N}{S_N^2} \to 0 \qquad (4)$$

as $N \to \infty$. Then with $\bar{x} = \sum_{i=1}^{n} R_i x_{Ni}$ we have

$$\frac{\bar{x}-\mu_N}{\sqrt{S_N^2\left(\frac{1}{n}-\frac{1}{N}\right)}} \xrightarrow{d} N(0,1).$$

(**) Discussion

- Condition (4) implies $n \to \infty$ and $N - n \to \infty$ because $\frac{m_N}{S_N^2} \geq \frac{1}{N/(N-1)} = 1 - \frac{1}{N}$.
- Suppose $\lim_{N\to\infty} \frac{n}{N} =: \alpha \in [0,1]$, and we scale x_{Ni} as x_{Ni}/S_N such that $S_N^2 = 1$. Then when $\alpha = 0$, (4) is equivalent to $m_N/n \to 0$; when $\alpha \in (0,1)$, (4) is equivalent to $m_N/N \to 0$; when $\alpha = 1$, (4) is equivalent to $m_N/(N-n) \to 0$. When $\sup_i |x_{Ni}| \leq c < \infty$, then (4) is equivalent to $n \to \infty$ and $N - n \to \infty$.
- When $\alpha \in (0,1)$, and $\lim_{N\to\infty} S_N^2 = S_\infty^2$, then

$$\bar{x} - \mu_N \xrightarrow{d} N\left(0, (1-\alpha) S_\infty^2\right).$$

- We can estimate S_N^2 by

$$\hat{S}_N^2 = \frac{1}{n-1} \sum_{i=1}^{N} R_i \left(x_{Ni} - \bar{x}\right)^2,$$

  which is shown to be unbiased below.
- Actually, $\hat{S}_N^2 / S_N^2 \xrightarrow{p} 1$ as $N \to \infty$ under condition (4). So

$$\frac{\bar{x} - \mu_N}{\sqrt{\hat{S}_N^2 \left(\frac{1}{n} - \frac{1}{N}\right)}} \xrightarrow{d} N(0,1).$$

Sampling Distributions of Sample Proportions

Sampling Distribution of the Sample Proportion

- Everything is the same as in the last section except that x_i can only take 0 or 1 and follows the Bernoulli(p) distribution.
- Now, $X := \sum_{i=1}^{n} x_i \sim$Binomial$(n,p)$, and the sample proportion

$$\hat{p} = \frac{X}{n}.$$

- $E[\hat{p}] = p$, and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, where recall that $Var(x_i) = p(1-p)$ is a function of only p - its mean.
- As $n \to \infty$, $\hat{p}$ approaches p by the LLN, and

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

approaches $N(0,1)$ by the CLT. [figure here]
- Recall that the approximation of normality is good if $np(1-p) > 5$.[2]
- Note that $X - np = \sigma_{\hat{p}} \cdot nz = \sqrt{np(1-p)}z \sim N(0, np(1-p))$, where $np(1-p) \to \infty$, so the difference between the observed number of success and the expected number of success might increase with n.

[2]Since $p(1-p) \le \frac{1}{4}$, $n > 20$.

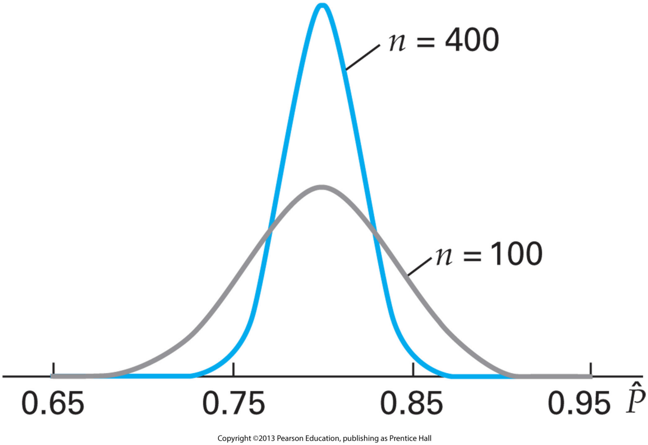


Figure: Density for $\hat{p}$ with $p = 0.80$

- $\sigma_{\hat{p}}$ decreases with n, and $\hat{p}$ is approximately normally distributed.

Example 6.8: Business Course Selection

- Suppose 43% of business graduates believe that a course in business ethics is very important. What is the probability of more than half of a random sample of 80 business graduates have this belief?
- Solution: Given that

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.43(1-0.43)}{80}} = 0.055,$$

we have

$$\begin{aligned} P(\hat{p} > 0.5) &= P\left(\frac{\hat{p}-p}{\sigma_{\hat{p}}} > \frac{0.5-0.43}{0.055}\right) \\ &= P(z > 1.27) \\ &= 1-\Phi(1.27) \\ &= 0.102, \end{aligned}$$

where $z \sim N(0,1)$ by the CLT.

Sampling Distributions of Sample Variances

Sampling Distribution of the Sample Variance

- Variance is important nowadays because consumers care about whether the particular item they bought works.
- Also, a smaller population variance reduces the variance of the sample mean: recall that $\sigma_{\bar{x}}^2 = \sigma^2/n$, where we assume random sampling with replacement or N is large.
- Recall that $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ is a natural estimator of σ^2.
- $E\left[s^2\right] = \sigma^2$, and if the population r.v. X is normally distributed, then

$$Var\left(s^2\right) = \frac{2\sigma^4}{n-1},$$

and

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \sim \chi_{n-1}^2,$$ [proof not required]

the chi-square distribution with $(n-1)$ degrees of freedom (df). [see the next slide for the definition of the chi-square distribution]

- Different from the CLT for the sample mean, the chi-square result is sensitive (i.e., not robust) to the normality assumption.

χ^2-Distribution

- If $Z_1, \cdots, Z_v$ are i.i.d. such that $Z_i \sim N(0,1)$, $i = 1, \cdots, v$, then

$$X = \sum_{i=1}^{v} Z_i^2 \sim \chi_v^2.$$

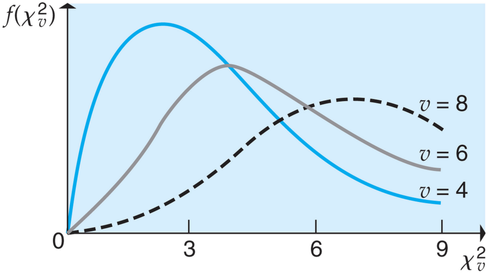


Figure: Density of the χ_v^2 Distribution with $v = 4, 6, 8$

- The χ^2 distribution can only take positive values (thinking of $\sum_{i=1}^{v} Z_i^2 > 0$ and $s^2 > 0$). [Appendix Table 7 contains chi-square probabilities]

History of the χ^2 Distribution

Friedrich R. Helmert (1843-1917), University of Berlin

- The χ^2 distribution was first described by Friedrich Robert Helmert in papers of 1875-6, and was independently rediscovered by Karl Pearson in 1900.

Mean and Variance of the Sample Variance

- $E\left[\chi_v^2\right]=v$ and $Var\left(\chi_v^2\right)=2v$ increase with v [refer to the figure in the last slide].
  - Why? $Var(Z_i)=E\left[Z_i^2\right]-E[Z_i]^2$ implies $E\left[Z_i^2\right]=E[Z_i]^2+Var(Z_i)=0^2+1$ $=1$, and $Var\left(Z_i^2\right)=E\left[Z_i^4\right]-E\left[Z_i^2\right]^2=3-1^2=2$.
- So $E\left[\frac{(n-1)s^2}{\sigma^2}\right]=n-1$ implies $E\left[s^2\right]=\sigma^2$.
  - $E\left[s^2\right]=\sigma^2$ even if X is not normally distributed, i.e., $\frac{(n-1)s^2}{\sigma^2}\not\sim\chi_{n-1}^2$. [(*) see the next slide which follows Appendix 3 of Chapter 6, Page 287]
- Also, $Var\left(\frac{(n-1)s^2}{\sigma^2}\right)=\frac{(n-1)^2}{\sigma^4}Var\left(s^2\right)=2(n-1)$ implies $Var\left(s^2\right)=\frac{2(n-1)\sigma^4}{(n-1)^2}$ $=\frac{2\sigma^4}{n-1}$, decreasing in n as in $Var(\bar{x})$.
- (*) Why lose one df in $\sum_{i=1}^n(x_i-\bar{x})^2$? Because the n values $\{(x_i-\bar{x})\}_{i=1}^n$ have only $(n-1)$ "independent" or "free" values: if we know $\{(x_i-\bar{x})\}_{i=1}^{n-1}$, then $(x_n-\bar{x})=-\sum_{i=1}^{n-1}(x_i-\bar{x})$ because $\sum_{i=1}^n(x_i-\bar{x})=\sum_{i=1}^n x_i-\sum_{i=1}^n\bar{x}=n\bar{x}-n\bar{x}=0$.
  - The df of $\{(x_i-\mu)\}_{i=1}^n$ is n, so we lose one df when we estimate μ by $\bar{x}$. [see more details in the next slide]
  - In general, the number of df lost equals the number of parameters estimated.

(*) The Mean of s^2 Without Normality

- Note that

$$\begin{aligned}&\sum_{i=1}^n (x_i-\bar{x})^2 = \sum_{i=1}^n [(x_i-\mu)-(\bar{x}-\mu)]^2 \\ &= \sum_{i=1}^n \left[(x_i-\mu)^2 - 2(\bar{x}-\mu)(x_i-\mu) + (\bar{x}-\mu)^2\right] \\ &= \sum_{i=1}^n (x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^n (x_i-\mu) + \sum_{i=1}^n (\bar{x}-\mu)^2 \\ &= \sum_{i=1}^n (x_i-\mu)^2 - 2n(\bar{x}-\mu)^2 + n(\bar{x}-\mu)^2 \\ &= \sum_{i=1}^n (x_i-\mu)^2 - n(\bar{x}-\mu)^2,\end{aligned}$$

so

$$E\left[\sum_{i=1}^n (x_i-\bar{x})^2\right] = E\left[\sum_{i=1}^n (x_i-\mu)^2\right] - nE\left[(\bar{x}-\mu)^2\right] = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.$$

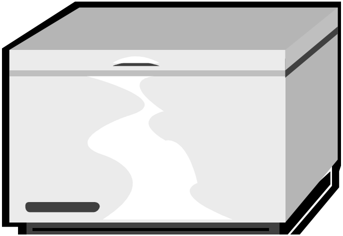
- As a result, $E\left[s^2\right] = E\left[\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2\right] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$.
- We lose one df in $\sum_{i=1}^n (x_i-\bar{x})^2$ because of the extra term $n(\bar{x}-\mu)^2$ since

$$\frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^n z_i^2 \sim \chi_n^2,$$

where $z_i = \frac{x_i-\mu}{\sigma}$ is the z-score of x_i and follows $N(0,1)$.

[Example] Applying the χ^2 Distribution

- A commercial freezer must hold a selected temperature with little variation. Specifications call for a standard deviation of no more than 4 degrees. A sample of 14 freezers is to be tested. What is the upper limit (K) for the sample variance such that the probability of exceeding this limit, given that the population standard deviation is 4, is less than 0.05?



continue

- Solution: The our target is to find K such that

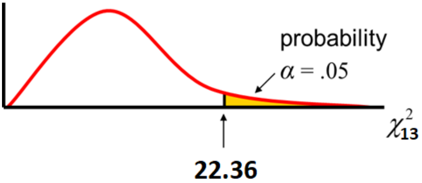
$$P\left(s^2 > K\right) = 0.05,$$

which implies

$$P\left(\frac{(n-1)s^2}{\sigma^2} > \frac{(n-1)K}{\sigma^2}\right) = P\left(\chi^2_{13} > \frac{13K}{16}\right) = 0.05,$$

so

$$\frac{13K}{16} = 22.36 \Longrightarrow K = \frac{22.36 \times 16}{13} = 27.25.$$



- If s^2 from the sample of size $n = 14$ is greater than 27.52, there is strong evidence to suggest the population variance exceeds 16.

(**) Further Results

- In random sampling without replacement, $E\left[s^2\right] = S^2 = \frac{N}{N-1}\sigma^2$. [see the next slide for details]
- So in random sampling with replacement, an unbiased estimator of $Var(\bar{x}) = \frac{\sigma^2}{n}$ is
$$\frac{s^2}{n},$$
and in random sampling without replacement, an unbiased estimator of $Var(\bar{x}) = \frac{S^2}{n} \cdot \frac{N-n}{N}$ is
$$\frac{s^2}{n} \cdot \frac{N-n}{N} = s^2\left(\frac{1}{n} - \frac{1}{N}\right),$$
where unbiasedness will be defined in the next section.
- In random sampling with replacement and X is not normally distributed,
$$Var\left(s^2\right) = \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4, \text{ [exercise*]}$$
which reduces to $\frac{2\sigma^4}{n-1}$ when X is normally distributed because the fourth central moment $\mu_4 := E\left[(X-\mu)^4\right] = 3\sigma^4$ now.[3]

---

[3] $\frac{3\sigma^4}{n} - \frac{(n-3)\sigma^4}{n(n-1)} = \frac{3(n-1)-(n-3)}{n(n-1)}\sigma^4 = \frac{2n\sigma^4}{n(n-1)} = \frac{2\sigma^4}{n-1}$.

(**) Mean of s^2 in Random Sampling Without Replacement

- First,

$$\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{N} R_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{N} R_i x_i\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{N} R_i x_i^2 - \frac{1}{n^2}\sum_{i=1}^{N} R_i x_i^2 - \frac{2}{n^2}\sum_{i=1}^{N}\sum_{j=i+1}^{N} R_i R_j x_i x_j.$$

- Therefore, from $E[R_i] = \frac{n}{N}$ and (1),

$$E\left[\hat{\sigma}^2\right] = \frac{1}{N}\frac{n-1}{n}\sum_{i=1}^{N} x_i^2 - \frac{2}{N(N-1)}\frac{n-1}{n}\sum_{i=1}^{N}\sum_{j=i+1}^{N} x_i x_j$$

$$= \frac{n-1}{n} S^2 \text{ by (2).}$$

- As a result,

$$E\left[s^2\right] = \frac{n}{n-1} E\left[\hat{\sigma}^2\right] = S^2.$$

Summary

Parameter	Estimator (Normalized◊)	Mean and Var of the Est.		Dist. of Normalized Est.	
		With Rep	Without Rep	Normality	$n \to \infty$
μ	$\bar{x}$	μ $\frac{\sigma^2}{n}$	μ $\frac{\sigma^2}{n}\frac{N-n}{N-1}$	-	-
	$(z = \frac{\bar{x}-E[\bar{x}]}{\sigma_{\bar{x}}})$	~~$(0,1)$~~	~~$(0,1)$~~	$N(0,1)$	$N(0,1)$†
σ^2	s^2	σ^2 $\frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4$‡	$S^2 = \frac{N}{N-1}\sigma^2$?	-	-
	$(\chi^2 = \frac{(n-1)s^2}{\sigma^2})$	~~$(n-1, \cdots)$~~	~~$\left(\frac{N}{N-1}(n-1), ?\right)$~~	χ^2_{n-1}	?

- (◊) When studying the distributions of normalized estimators, assume x_i's are iid, i.e., randomly sampling x_i with replacement.
- (†) If $\bar{x}$ is sample proportion, i.e., $x_i \sim$Bernoulli(p) (not normal), then $E[\bar{x}] = p$, and $\sigma_{\bar{x}} = \sqrt{\frac{p(1-p)}{n}}$.
- (‡) The variance reduces to $\frac{2\sigma^4}{n-1}$ if $x_i \sim N\left(\mu, \sigma^2\right)$.

Properties of Point Estimators

Estimator and Estimate

- We mentioned "estimator" above, and here we provide a rigorous definition.
- An estimator of a population parameter is a function of the sample whose value provides an approximation to this unknown parameter. If the sample is $\{x_i\}_{i=1}^n$, then an estimator is $f(x_1, \cdots, x_n)$ which is also a random variable given that x_i is random.
- An estimate is a realized value of the estimator. So an estimate is just a number.
- A point estimator of a population parameter is a function of the sample that produces a single number called a point estimate.
- An interval estimator of a population parameter is a function of the sample that produces an interval.
  - An example of the interval estimator is the confidence interval that will be discussed in Lecture 7.
- No single mechanism exists for the determination of a uniquely "best" point estimator in all circumstances.
- What is available instead is a set of criteria under which particular estimators can be evaluated.
- Two criteria discussed here are unbiasedness and efficiency.

Unbiasedness

- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of a population parameter θ if $E\left[\hat{\theta}\right] = \theta$ for any possible value of θ.
- We show above that $E\left[\bar{x}\right] = \mu$, $E\left[\hat{p}\right] = p$, and $E\left[s^2\right] = \sigma^2$, so $\bar{x}, \hat{p}$ and s^2 are unbiased estimators of μ, p and σ^2, respectively.

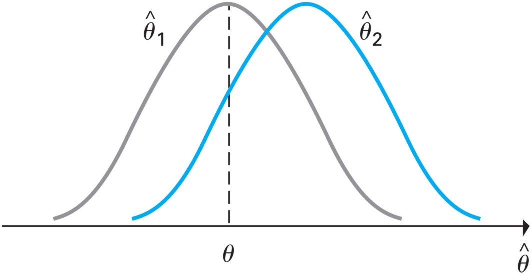


Figure: Density of an Unbiased Estimator $\hat{\theta}_1$ and a Biased Estimator $\hat{\theta}_2$

- Bias$\left(\hat{\theta}\right) = E\left[\hat{\theta}\right] - \theta$. So the bias of an unbiased estimator is 0.

Most Efficient

- There may be many unbiased estimators. To choose among them, we use variance as a criterion.
- The unbiased estimator with the smallest variance is preferred, and is called the most efficient estimator, or the minimum variance unbiased estimator (MVUE).
- For two unbiased estimators of θ based on the same sample, $\hat{\theta}_1$ and $\hat{\theta}_2$, $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$.
- The relative efficiency of $\hat{\theta}_1$ with respect to (w.r.t.) $\hat{\theta}_2$ is $Var\left(\hat{\theta}_2\right)/Var\left(\hat{\theta}_1\right)$, i.e., if $Var\left(\hat{\theta}_2\right) > Var\left(\hat{\theta}_1\right)$, then $\hat{\theta}_1$ is more efficient, so its relative efficiency w.r.t. $\hat{\theta}_2$ is greater than 1.
- Given a random sample $\{x_i\}_{i=1}^n$ with $x_i \sim N\left(\mu, \sigma^2\right)$. Both the sample mean $\bar{x}$ and sample median $x_{.5}$ are unbiased estimator of μ.
- But $Var\left(\bar{x}\right) = \frac{\sigma^2}{n}$, and $Var\left(x_{.5}\right) = \frac{\pi}{2}\frac{\sigma^2}{n} = \frac{1.57\sigma^2}{n}$ when n is large [proof not required], so the sample mean is more efficient than the sample median, and the relative efficiency of the former to the latter is

$$\text{relative efficiency} = \frac{Var\left(x_{.5}\right)}{Var\left(\bar{x}\right)} = 1.57.$$

Table 7.1 Properties of Selected Point Estimators

Population Parameter	Point Estimator	Properties
Mean, μ	$\overline{X}$	Unbiased, most efficient (assuming normality)
Mean, μ	Median	Unbiased (assuming normality), but not most efficient
Proportion, P	$\hat{p}$	Unbiased, most efficient
Variance, σ^2	s^2	Unbiased, most efficient (assuming normality)

- (**) Proof for the efficiency properties are not required.