Chapter 5
Random Sampling

Population and Sample

Definition: Population

A population is the totality of the elements under study. We are interested in learning something about this population.

Examples: Number of alligators in Texas, percentage of unemployed workers in cities in the U.S., stock returns of IBM.

A RV X defined over a population is called the population RV.

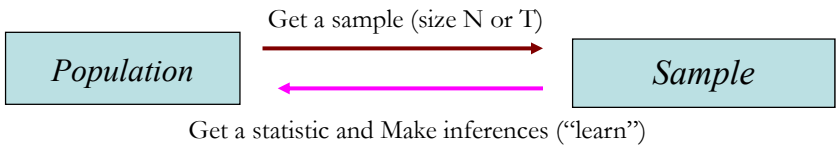
Usually, the population is large, making a complete enumeration of all the values in the population impractical or impossible. Thus, the descriptive statistics describing the population –i.e., the *population parameters*- will be considered unknown.

Population and Sample

• Typical situation in statistics: we want to make inferences about an unknown population parameter θ using a sample –i.e., a small collection of observations from the general population- $\{X_1, \ldots, X_N\}$.

• We summarize the information in the sample with a *statistic*, which is a function of the sample.

That is, any statistic summarizes the data, or reduces the information in the sample to a single number. To make inferences, we use the information in the statistic instead of the entire sample.



Population and Sample

Definition: Sample

The *sample* is a (manageable) subset of elements of the population.

Samples are collected to learn about the population. The process of collecting information from a sample is referred to as *sampling*.

<u>Definition</u>: Random Sample

A *random sample* is a sample where the probability that any individual member from the population being selected as part of the sample is exactly the same as any other individual member of the population

In mathematical terms, given a random variable X with distribution F, a *random sample* of length n is a set of n independent, identically distributed (*iid*) random variables with distribution F.

Sample Statistics

• A statistic (singular) is a single measure of some attribute of a sample (for example, its arithmetic mean value). It is calculated by applying a function (statistical algorithm) to the values of the items comprising the sample, which are known together as a set of data.

Definition: Statistic

A *statistic* is a function of the observable random variable(s), which does not contain any unknown parameters.

Examples: sample mean, sample variance, minimum, $(x_1+x_n)/2$, etc.

Note: A statistic is distinct from a population parameter. A statistic will be used to estimate a population parameter. In this case, the statistic is called an *estimator*.

Sample Statistics

• Sample Statistics are used to estimate population parameters

Example: $\bar{X}$ is an estimate of the population mean, μ.

Notation: Population parameters: Greek letters (μ, σ, θ, etc.)

Estimators: A hat over the Greek letter ($\hat{\theta}$)

• Problems:

– Different samples provide different estimates of the population parameter of interest.

– Sample results have potential variability, thus sampling error exits

Sampling error: The difference between a value (a statistic) computed from a sample and the corresponding value (a parameter) computed from a population.

Sample Statistics

• The definition of a sample statistic is very general. There are potentially infinite sample statistics.

For example, $(x_1+x_N)/2$ is by definition a statistic and we could claim that it estimates the population mean of the variable X. However, this is probably not a good estimate.

We would like our estimators to have certain desirable properties.

• Some simple properties for estimators, $\hat{\theta}$:

- $\hat{\theta}$ is *unbiased* estimator of θ if $E[\hat{\theta}] = \theta$
- $\hat{\theta}$ is *most efficient* if the variance of the estimator is minimized.
- $\hat{\theta}$ is BUE, or Best Unbiased Estimate, if it is the estimator with the smallest variance among all unbiased estimates.

Sampling Distribution

• A sample statistic is a function of RVs. Then, it has a statistical distribution.

• In general, in economics and finance, we observe only *one* sample mean (from our only sample). But, *many* sample means are possible.

• A sampling distribution is a distribution of a statistic over all possible samples.

• The sampling distribution shows the relation between the probability of a statistic and the statistic's value for all possible samples of size *n* drawn from a population.

Sampling Distribution: Sample Mean ($\bar{X}$)

Let $X_1, \ldots, X_n$ be n mutually independent random variables each having mean μ and standard deviation σ (variance σ^2).

Let $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}X_1 + \ldots + \frac{1}{n}X_n$$

Then $$\mu_{\bar{X}} = E\left[\bar{X}\right] = \frac{1}{n}E\left[X_1\right] + \ldots + \frac{1}{n}E\left[X_n\right] = \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu = \mu$$

Also $$\sigma^2_{\bar{X}} = Var\left[\bar{X}\right] = \left(\frac{1}{n}\right)^2 Var\left[X_1\right] + \ldots + \left(\frac{1}{n}\right)^2 Var\left[X_n\right]$$

$$= \left(\frac{1}{n}\right)^2 \sigma^2 + \ldots + \left(\frac{1}{n}\right)^2 \sigma^2 = n\frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Thus $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

Sampling Distribution: Sample Mean ($\bar{X}$)

- Summary for $\bar{X}$, when data is normal:

Sampling distribution: $\bar{X} \sim N(\mu, \sigma^2/N)$.

Mean: $E[\bar{X}] = \mu$

Variance: $Var[\bar{X}] = \sigma^2/N$.

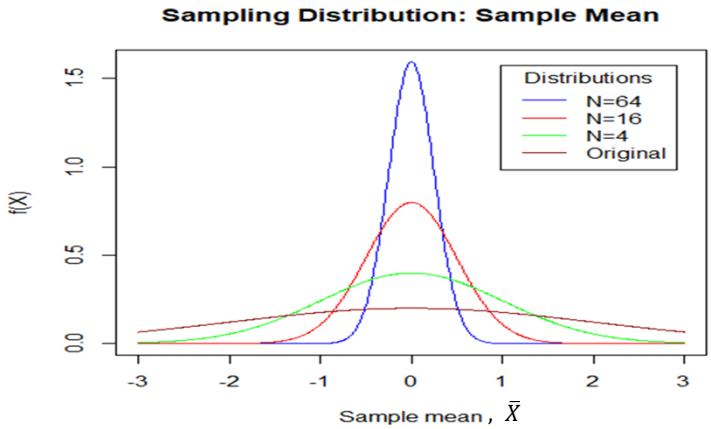
<u>Note</u>: If the data is not normal (& N is large), the CLT can be used to approximate the sampling distribution by the asymptotic one:

$$\bar{X} \xrightarrow{a} N(\mu, \sigma^2/N).$$

Sampling Distribution: Sample Mean ($\bar{X}$)

- Sampling Distribution for the Sample Mean

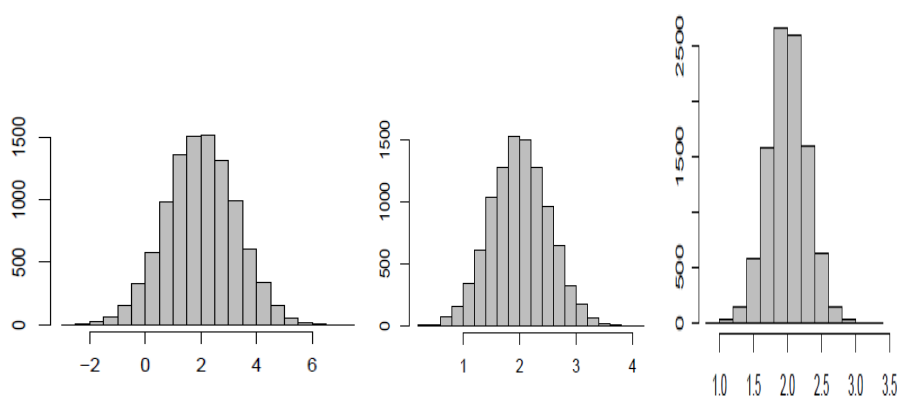
$$\bar{X} \sim N(\mu, \sigma^2/N)$$



Note: As $N \to \infty$, $\bar{X} \to \mu$ –i.e., the distribution becomes a spike at μ!

Sampling Distribution: Sample Mean ($\bar{X}$)

- 10,000 samples, for a N(2, 4) population. Different sample sizes, N:



$N = 10$ $N = 60$ $N = 200$

Note: As $N \to \infty$, $\bar{X}$ becomes more concentrated around μ=2. In the limit, a spike at μ!

12

Distribution of the sample variance: Preliminaries

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Use of moment generating functions (Again)

1. Using the moment generating functions of *X, Y, Z,* …determine the moment generating function of $W = h(X, Y, Z, \ldots)$.
2. Identify the distribution of W from its moment generating function.

   This procedure works well for sums, linear combinations etc.

Gamma Distribution: Theorem (Summation)

Let X and Y denote a independent random variables each having a gamma distribution with parameters (λ,α_1) and (λ,α_2). Then $W = X + Y$ has a gamma distribution with parameters $(\lambda, \alpha_1 + \alpha_2)$.

Proof:

$$m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1} \quad \text{and} \quad m_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2}$$

$$\text{Therefore } m_{X+Y}(t) = m_X(t)\, m_Y(t)$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1}\left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1+\alpha_2}$$

Recognizing that this is the moment generating function of the gamma distribution with parameters $(\lambda, \alpha_1 + \alpha_2)$ we conclude that $W = X + Y$ has a gamma distribution with parameters $(\lambda, \alpha_1 + \alpha_2)$.

Gamma Distribution: Theorem (Summation: n RV)

Let $x_1, x_2, \ldots, x_n$ denote n independent random variables each having a gamma distribution with parameters (α_i, λ), $i = 1, 2, \ldots, n$.

Then $W = x_1 + x_2 + \ldots + x_n$ has a gamma distribution with parameters $(\alpha_1 + \alpha_2 + \ldots + \alpha_n, \lambda)$.

Proof:

$$m_{x_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_i} \quad i = 1, 2 \ldots, n$$

$$m_{x_1+x_2+\ldots+x_n}(t) = m_{x_1}(t)\, m_{x_2}(t) \ldots m_{x_n}(t)$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1}\left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} \cdots \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_n} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1+\alpha_2+\ldots+\alpha_n}$$

This is the moment generating function of the gamma distribution with parameters $(\alpha_1 + \alpha_2 + \ldots + \alpha_n, \lambda)$. we conclude that $W = x_1 + x_2 + \ldots + x_n$ has a gamma distribution with parameters $(\alpha_1 + \alpha_2 + \ldots + \alpha_n, \lambda)$.

Gamma Distribution: Theorem (Scaling)

Suppose that X is a random variable having a gamma distribution with parameters (α, λ).
Then, $W = aX$ has a gamma distribution with parameters $(\alpha, \lambda/a)$

Proof:

$$m_x(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$$

$$\text{then } m_{ax}(t) = m_x(at) = \left(\frac{\lambda}{\lambda - at}\right)^{\alpha} = \left(\frac{\lambda/a}{\lambda/a - t}\right)^{\alpha}$$

Gamma Distribution: Theorem (Special Cases)

1. Let X and Y be independent random variables having a χ^2 distribution with n_1 and n_2 degrees of freedom, respectively. Then, $X + Y$ has a χ^2 distribution with degrees of freedom $n_1 + n_2$.

   <u>Notation</u>: $X + Y \sim \chi^2_{(n1+n2)}$ (a χ^2 distribution with df= $n_1 + n_2$.)

2. Let $x_1, x_2, \ldots, x_n$, be independent RVs having a χ^2 distribution with $n_1, n_2, \ldots, n_N$ degrees of freedom, respectively.
   Then, $x_1 + x_2 + \ldots + x_n \sim \chi^2_{(n1+n2+\ldots+nN)}$

<u>Note</u>: Both of these properties follow from the fact that a χ^2 RV with n degrees of freedom is a Gamma RV with $\lambda = ½$ and $\alpha = n/2$.

Gamma Distribution: Theorem (Special Cases)

• Two useful χ_n^2 results

1. Let $z \sim N(0,1)$. Then,

$$z^2 \sim \chi_1^2.$$

2. Let $z_1, z_2, \ldots, z_\nu$ be independent random variables each following a N(0,1) distribution. Then,

$$U = z_1^2 + z_2^2 + \ldots + z_\nu^2 \sim \chi_\nu^2$$

Gamma Distribution: Theorem

Suppose that U_1 and U_2 are independent random variables and that $U = U_1 + U_2$. Suppose that U_1 and U have a χ^2 distribution with degrees of freedom ν_1 and ν, respectively. ($\nu_1 < \nu$). Then,

$$U_2 \sim \chi_{\nu 2}^2,$$

where $\nu_2 = \nu - \nu_1$.

Proof:

Now $m_{U_1}(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t} \right)^{\frac{\nu_1}{2}}$ and $m_U(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t} \right)^{\frac{\nu}{2}}$

Also $m_U(t) = m_{U_1}(t) m_{U_2}(t)$

Gamma Distribution: Theorem

Proof (continuation):

Hence $m_{U_2}(t) = \dfrac{m_U(t)}{m_{U_1}(t)}$

$$= \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^{\frac{v}{2}} \Bigg/ \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^{\frac{v_1}{2}} = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^{\frac{v}{2} - \frac{v_1}{2}}$$

Q.E.D.

Distribution of the sample variance

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Properties of the sample variance (s^2)

We show that we can decompose the numerator of s^2 as:

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - a)^2 - n(\overline{x} - a)^2$$

Proof:

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - a + a - \overline{x})^2$$

$$= \sum_{i=1}^{n} \left[(x_i - a)^2 - 2(x_i - a)(\overline{x} - a) + (\overline{x} - a)^2 \right]$$

$$= \sum_{i=1}^{n} (x_i - a)^2 - 2(\overline{x} - a)\sum_{i=1}^{n} (x_i - a) + n(\overline{x} - a)^2$$

$$= \sum_{i=1}^{n} (x_i - a)^2 - 2n(\overline{x} - a)^2 + n(\overline{x} - a)^2$$

Properties of the sample variance (s^2)

Proof (continuation):

$$= \sum_{i=1}^{n} (x_i - a)^2 - 2n(\overline{x} - a)^2 + n(\overline{x} - a)^2$$

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - a)^2 - n(\overline{x} - a)^2$$

Properties of the sample variance (s^2)

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - a)^2 - n(\bar{x} - a)^2$$

Special Cases

1. Setting $a = 0$. (It delivers the computing formula.)

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

2. Setting $a = \mu$.

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

or $$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

Distribution of the s^2 of a normal variate

Let $x_1, x_2, \ldots, x_N$ denote a random sample from the *normal distribution* with mean μ and variance σ^2. Then,

$$(N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}, \qquad \text{where } s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$

Proof:

Let $z_1 = \dfrac{x_1 - \mu}{\sigma}, \ldots, z_n = \dfrac{x_n - \mu}{\sigma}$

Then $$z_1^2 + \ldots + z_n^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

Distribution of the s^2 of a normal variate

Now, recall Special Case 2:

$$\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{\sigma^2} + \frac{N\,(\bar{x}\,-\,\mu)^2}{\sigma^2}$$

or $U = U_1 + U_2$, where

$$U = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{\sigma^2} \sim \chi_N^2$$

We know that $\bar{x} \sim N(\mu, \sigma^2/N)$. Thus,

$$z = \frac{\sqrt{N}(\bar{x}-\mu)}{\sigma} \sim N(0,1) \;\Rightarrow\; U_1 = z^2 = \frac{N\,(\bar{x}-\mu)^2}{\sigma^2} \sim \chi_1^2$$

If we can show that U_1 and U_2 are independent. Then,

$$U_2 \sim \chi_{N-1}^2$$

$$U_2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{\sigma^2} = (N-1)\,s^2/\sigma^2 \sim \chi_{N-1}^2$$

To show that that U_1 and U_2 are independent (and to complete the proof) we need to show that:

$\sum_{i=1}^{n}(x_i-\bar{x})^2$ and $\bar{x}$ are independent RVs.

Let $u = N\,\bar{x}^2 = N\,(\Sigma_i\, x_i)^2/ = (1/\,N)(\Sigma_i\, x_i^2 + \Sigma_i\, \Sigma_j\, x_i\, x_j) = \mathbf{x}'M_1\mathbf{x}$, where

$$M_1 = \frac{1}{n}\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ . & . & \dots & . \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

Similarly $v = \sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = \mathbf{x}'\mathbf{x} - \mathbf{x}'M_1\mathbf{x}$

$= \mathbf{x}'\,(I - M_1)\,\mathbf{x} = \mathbf{x}'\,M_2\,\mathbf{x}$

Thus, u and v are independent if $M_1 M_2 = 0$.

$\Rightarrow M_1 M_2 = M_1\,(I - M_1) = M_1 - M_1^2 = 0$ (since M_1 is idempotent).

Distribution of the s^2 of a normal variate

• Summary

Let $x_1, x_2, \ldots, x_N$ denote a sample from the normal distribution with mean μ and variance σ^2. Then,

1. $\bar{x} \sim N(\mu, \sigma^2/N)$

2. $U = (N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}$

<u>Note</u>: If $X \sim \chi^2_v$, then $E[X] = v$

$Var[X] = 2\,v$

Then, $E[U] = N - 1 \Rightarrow E[(N-1)s^2/\sigma^2] = ((N-1)/\sigma^2)\, E[s^2] = N-1$

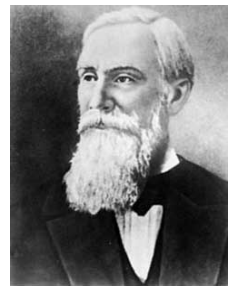
$\Rightarrow E[s^2] = \sigma^2$

$Var[U] = 2(N-1) \Rightarrow Var[(N-1)s^2/\sigma^2] = (N-1)^2/\sigma^4\, Var[s^2] = 2(N-1)$

$\Rightarrow Var[s^2] = 2\sigma^4/(N-1)$

The Law of Large Numbers

Law of Large Numbers: Chebyshev's Inequality



Pafnuty Chebyshev, Rusia (1821, 1894)

Theorem

Let b be a positive constant and $h(x)$ a non-negative function of the random variable X. Then,

$$P[\,|x - \mu| \geq b\,] \leq (1/b)\, E[h(x)]$$

Corollary (Chebishev's Inequality)

For any constant $c > 0$,

$$P[\,|x - \mu| \leq c\,] \leq \sigma^2/c^2,$$

where σ^2 is the Var(x).

This inequality can be expressed in two alternative forms:

1. $P[\,|x - \mu| \leq c\,] \geq 1 - \sigma^2/c^2$
2. $P[\,|x - \mu| \geq \eta\,\sigma\,] \leq 1/\eta^2$. $\quad (c = \eta\,\sigma)$

Law of Large Numbers: Chebyshev's Inequality

Proof: We want to prove $P[\,|X - \mu| \leq b\,] \leq (1/b)\, E[h(x)]$

$$E[h(x)] = \int_{-\infty}^{\infty} h(x) f(x)dx = \int_{h(x)\geq b} h(x) f(x)dx + \int_{h(x)<b} h(x) f(x)dx \geq$$

$$\geq \int_{h(x)\geq b} h(x) f(x)dx \geq b \int_{h(x)\geq b} f(x)dx = b\,P[h(x) \geq b]$$

Thus, $P[h(x) \geq b\,] \leq (1/b)\, E[h(x)]$. ■

<u>Corollary</u>: Let $h(x) = (x - \mu)^2$ and $b = c^2$.

$$P\,[(x - \mu)^2 \geq c^2] = P[\,|x - \mu| \geq c] \leq (1/c^2)\, E[(x - \mu)^2] = \sigma^2/c^2$$

Law of Large Numbers: Chebyshev's Inequality

• This inequality sets a weak lower bound on the probability that a random variable falls within a certain confidence interval.

Example: The probability that X falls within two standard deviations of its mean is at least (setting $c = 2\,\sigma$):

$$P[\,|X - \mu\,| \leq c\,] \geq 1 - \sigma^2/c^2 = 1 - 1/4 = 3/4 = 0.75.$$

Note: Chebyshev's inequality usually understates the actual probability. In the normal distribution, the probability of a random variable falling within two standard deviations of its mean is 0.95.

The Law of Large Numbers: Theorem

Long history

Gerolamo Cardano (1501-1576) stated it without proof.

Jacob Bernoulli published a rigorous proof in 1713.

Poisson described it under its current name "La loi des grands nombres."

Theorem (Weak Law)

Let $X_1, \ldots, X_N$ be N mutually independent random variables each having mean μ and a finite σ -i.e., the sequence $\{X_N\}$ is *i.i.d.*

Let $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{N}$

Then for any $\delta > 0$ (no matter how small)

$$P[\,|\bar{X} - \mu| < \delta\,] = P[\,\mu - \delta < \bar{X} < \mu + \delta] \to 1, \qquad \text{as } N \to \infty$$

The Law of Large Numbers: Theorem

Proof:

We will use Chebychev's inequality:

$$P[\,|\,X - \mu\,| \le c\,] \ge 1 - \sigma^2/c^2$$

Set $c = k\,\sigma$ or $\sigma^2/c^2 = 1/k^2$. Then,

$$P[\,\mu - k\,\sigma < X < \mu + k\,\sigma\,] \ge 1 - 1/k^2$$

Now set: $\mu_{\bar{x}} = \mu$ & $\sigma_{\bar{x}} = \sigma/\sqrt{N}$. Then,

$$P[\,\mu - \delta < \bar{X} < \mu + \delta] \ge 1 - 1/k^2$$

where $\delta = k\,\sigma_{\bar{x}}$ (or $k = \delta\,\sqrt{N}/\sigma$).

Thus, $P[\,\mu - \delta < \bar{X} < \mu + \delta\,] \ge 1 - 1/k^2 \to 1$, as $N \to \infty$

The Law of Large Numbers: Theorem

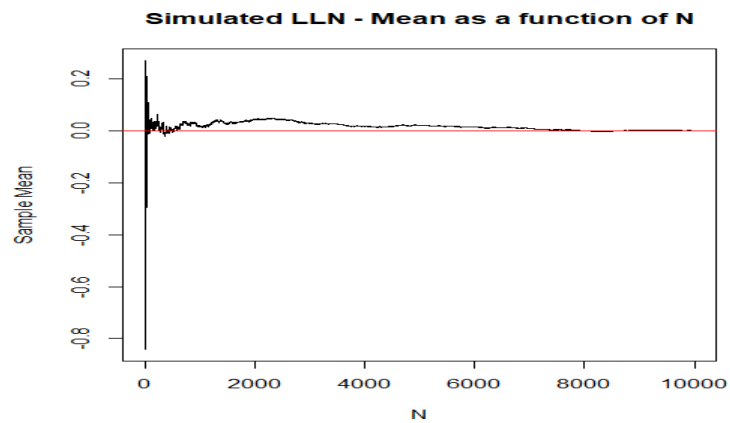
Then,

$$P[\,|\bar{X} - \mu| < \delta\,] = P[\,\mu - \delta < \bar{X} < \mu + \delta] \to 1, \quad \text{as } N \to \infty$$

Note: The proof assumed a finite variance –i.e., it relies on Chebychev's inequality. A finite variance is not needed to get the LLN.

The Law of Large Numbers: Illustration

- Below, we randomly generate 10,000 observations from a N(0, 1) and compute the sample mean as a function of N. As expected, as N increases, the sample mean gets closer to the population mean (μ =0).



37

The Law of Large Numbers: Proportions

A Special case: Proportions

Let $X_1, \ldots, X_N$ be n mutually independent random variables each having Bernoulli distribution with parameter p.

$$X_i = \begin{cases} 1 & \text{if repetition is } \mathbf{S} \text{ (prob } = p) \\ 0 & \text{if repetition is } \mathbf{F} \text{ (prob } = q = 1 - p) \end{cases}$$

$$\mu = \mathrm{E}[X_i] = p$$

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \hat{p} = \text{proportion of successes}$$

Thus, by LLN:

$$P[\hat{p} - \delta < \hat{p} < \hat{p} + \delta] \to 1 \quad \text{as } N \to \infty$$

The Law of Large Numbers: Proportions

•Thus, the LLN states that $\hat{p}$ (proportion of successes) converges to the probability of success p as $N \to \infty$.

<u>Misinterpretation</u>: If the proportion of successes is currently lower that p, then the proportion of successes in the future will have to be larger than p to counter this and ensure that the LLN holds true.

The Law of Large Numbers: Weak & Strong

Theorem (Khinchine's Weak Law of Large Numbers)

Let $X_1, \ldots, X_N$ be a sequence of n *i.i.d.* random variables each having mean μ. Then, for any $\delta > 0$,

$$\lim_{n\to\infty} P[\,|\bar{X}_N - \mu| < \delta\,] = 0$$

This is called *convergence in probability*.

<u>Note</u>: Khinchine's Weak Law of Large Numbers is more general. It allows for the case where only μ exists.

Theorem (Strong Law)

Let $X_1, \ldots, X_N$ be a sequence of N *i.i.d.* random variables each having mean μ. Then,

$$P[\,\lim_{N\to\infty} \bar{X}_N = \mu\,] = 1.$$

This is called *almost sure convergence*.

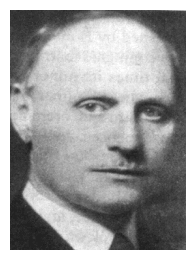
The Law of Large Numbers: Weak & Strong

LLN and SLN

• The *weak law* states that for a specified large N, the average $\bar{X}$ is likely to be near μ. Thus, it leaves open the possibility that $|\bar{X} - \mu| > \delta$ happens an infinite number of times, although at infrequent intervals.

• The *strong law* shows that this almost surely will not occur. In particular, it implies that with probability 1, we have that for any $\delta > 0$, the inequality $|\bar{X} - \mu| < \delta$ holds for all large enough N.

Famous Inequalities



Carlo Bonferroni (1892-1960)

8.1 Bonferroni's Inequality

Basic inequality of probability theory:

$$P[A \cap B] \geq P[A] + P[B] - 1$$

8.2 A Useful Lemma

Lemma 8.1: If $1/p + 1/q = 1$, then $1/p\ \alpha^p + 1/q\ \beta^q \geq \alpha\beta$

Almost all of the following inequalities are derived from this lemma.

Otto Hölder (1859-1937)

8.3 Holder's Inequality

For p, q satisfying Lemma 8.1, we have

$$|E[XY]| \leq E|XY| \leq (E|X|^p)^{1/p}\ (E|Y|^q)^{1/q}$$

8.4 Cauchy-Schwarz Inequality

(Holder's inequality with $p=q=2$)

$|E[XY]| \leq E|XY| \leq \{E|X|^2 E|Y|^2\}^{1/2}$

8.5 Covariance Inequality

(Application of Cauchy-Schwarz)

$E|(X - \mu_x)(Y - \mu_y)| \leq \{E(X - \mu_x)^2 E(Y - \mu_y)^2\}^{1/2}$

$\mathrm{Cov}(X,Y)^2 \leq \{\sigma_x^2 \sigma_y^2\}$

Andrey Markov (1856–1922)

8.6 Markov's Inequality

If $E[X] < 1$ and $t > 0$, then $P[|X| \geq t] \leq E[|X|]/t$

8.7 Jensen's Inequality

If $g(x)$ is convex, then $E[g(x)] \geq g(E[x])$

If $g(x)$ is concave, then $E[g(x)] \leq g(E[x])$

Johan Jensen (1859 – 1925)

Johan Jensen, Denmark (1859 – 1925)

Otto Hölder, Germany (1859–1937)

Carlo Bonferroni, Italy (1892-1960)

Andrey Markov, Rusia (1856–1922)

Pafnuty Chebyshev , Rusia (1821–1894)

Karl Schwarz, Germany (1843–1921)