# 5. Estimators

- statistics as estimators
- convergence
- properties of estimators
- sample mean and sample variance

# Descriptive statistics

if $x_1, x_2, \ldots, x_N$ are drawn independently from the same population

$\{x_i\}_{i=1,\ldots,N}$ is a **random sample** and said to be independent, identically distribuited (iid)

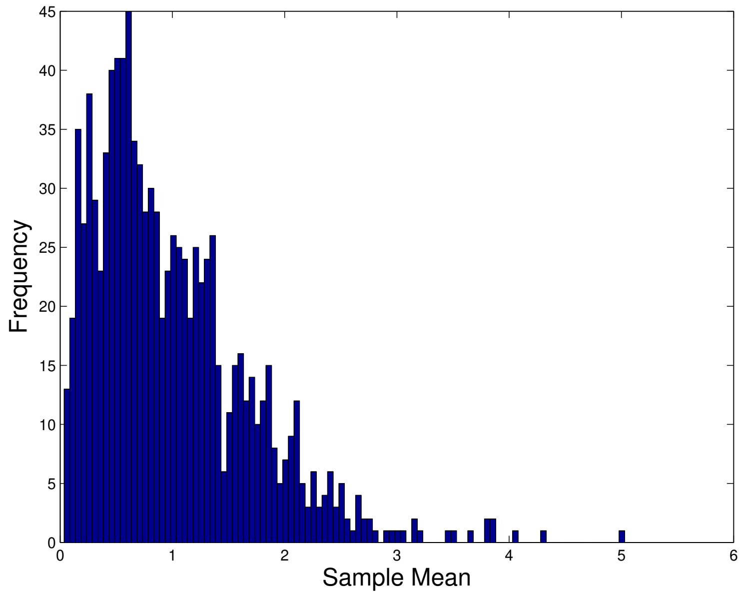
typical summary **statistics** used to describe the sample data

| statistic | description | what to describe |
|---|---|---|
| mean | $(1/N)\sum_{i=1}^N x_i$ | central tendency |
| median | middle ranked observation | central tendency |
| standard deviation | $\sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$ | dispersion |
| skewness | $\frac{(1/N)\sum_{i=1}^N (x_i - \bar{x})^3}{\mathrm{SD}^3}$ | asymmetry of pdf |
| kurtosis | $\frac{(1/N)\sum_{i=1}^N (x_i - \bar{x})^4}{\mathrm{SD}^4}$ | amount of heavy tails |

**definition:** a **statistic** is any function computed from the data in a sample

- a statistic is a function of random values, so it is also an RV
- the probability distribution of a statistic is called a **sampling distribution**

**example:** a histogram of $1000$ realizations of the sample mean of $\chi_1^2$




the sample mean is calculated on $4$ observations

# Estimation of parameters

Definition: an **estimator** is a rule for using data to estimate the model parameter

example: to estimate a population mean, one can use *sample mean* or *sample minimum*

- typically, one can compare an estimator with others from their properties
- such properties can be divided into
  - finite sample properties
  - asymptotic properties: when sample size is large

# Estimators

- statistics as estimators
- **convergence**
- properties of estimators
- sample mean and sample variance

# Convergence of deterministic sequences

**Definition:** a sequence of *deterministic* numbers $\{a_n : n = 1, 2, \ldots\}$ **converges** to $a$ if

$$\forall \epsilon > 0, \exists N \text{ such that if } n > N \text{ then } |a_n - a| < \epsilon$$

and we write

$$a_n \to a, \quad \text{as } n \to \infty$$

or

$$\lim_{n \to \infty} a_n = a$$

**Definition:** a sequence $a_n$ is **bounded** if there is some $M < \infty$ such that

$$|a_n| \leq M, \quad \text{for all } n$$

otherwise, we say that $a_n$ is **unbounded**

# Convergence in Probability

**Definition:** a sequence of *random variables* $\{X_n : n = 1, 2, \ldots\}$ **converges in probability** to a random variable $X$ if for all $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

and we write

$$X_n \xrightarrow{p} X$$

and say that $X$ is the **probability limit (plim)** of $X_n$ : $\textbf{plim}\, X_n = X$

**Definition:** $X_n$ is **bounded in probability** if for every $\epsilon > 0$, there exists $M_\epsilon < \infty$ and an integer $N_\epsilon$ such that

$$P(|X_n| \geq M_\epsilon) < \epsilon, \quad \forall n \geq N_\epsilon$$

**example:** $X_n$ is a Bernoulli where $P(X_n = 0) = 1 - 1/n$ and $P(X_n = 1) = 1/n$

$x \in \mathbf{R}^{20\times 5}$, contains $20$ samples of $X_n$ where $n = (1, 2, 3, 10, 100)$

```
x =

     1     1     1     0     0
     1     0     0     0     0
     1     0     0     0     0
     1     1     0     0     0
     1     0     0     0     0
     1     0     1     0     0
     1     1     0     0     0
     1     0     0     0     0
     1     1     1     0     0
     1     1     0     1     0
     1     0     0     0     0
     1     0     1     0     0
     1     1     0     0     0
     1     0     0     0     0
     1     0     1     0     0
     1     0     1     0     0
     1     1     0     0     0
     1     1     0     0     0
     1     1     0     0     0
     1     0     0     0     0
```

```
x =
     1     0     0     0     0
     1     0     1     0     0
     1     1     0     0     0
     1     1     0     0     0
     1     0     1     0     0
     1     0     1     0     0
     1     0     0     0     0
     1     1     0     0     0
     1     0     0     0     0
     1     0     0     0     0
     1     0     1     0     0
     1     1     1     0     0
     1     0     0     0     0
     1     0     0     0     0
     1     1     0     0     0
     1     0     0     0     0
     1     0     0     1     0
     1     1     0     0     0
     1     0     0     0     0
     1     1     0     0     0
```

$X_n$ converges in probability to $0$

# Rules for probability limits

if $X_n$ and $Y_n$ are RVs with $\mathbf{plim}\, X_n = x$ and $\mathbf{plim}\, Y_n = y$ then

- $\mathbf{plim}(X_n + Y_n) = x + y$ (sum rule)
- $\mathbf{plim}\, X_n Y_n = xy$ (product rule)
- $\mathbf{plim}\, X_n / Y_n = x/y$ if $y \neq 0$ (ratio rule)

(all the rules can be generalized to random matrices)

# Convergence with Probability One

**Definition:** a random sequence $X_n$ converges **with probability one** to a random variable $X$ if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

and denoted by $X_n \stackrel{as}{\to} X$

- aka **almost sure** or **strong consistency** for $X$
- almost sure implies convergence in probability (weak consistency for $X$)

# Laws of Large Numbers

theorems for convergence in probability for the sequence of **sample average**

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

where $X_i$ is a random variable

**weak law of large numbers:**

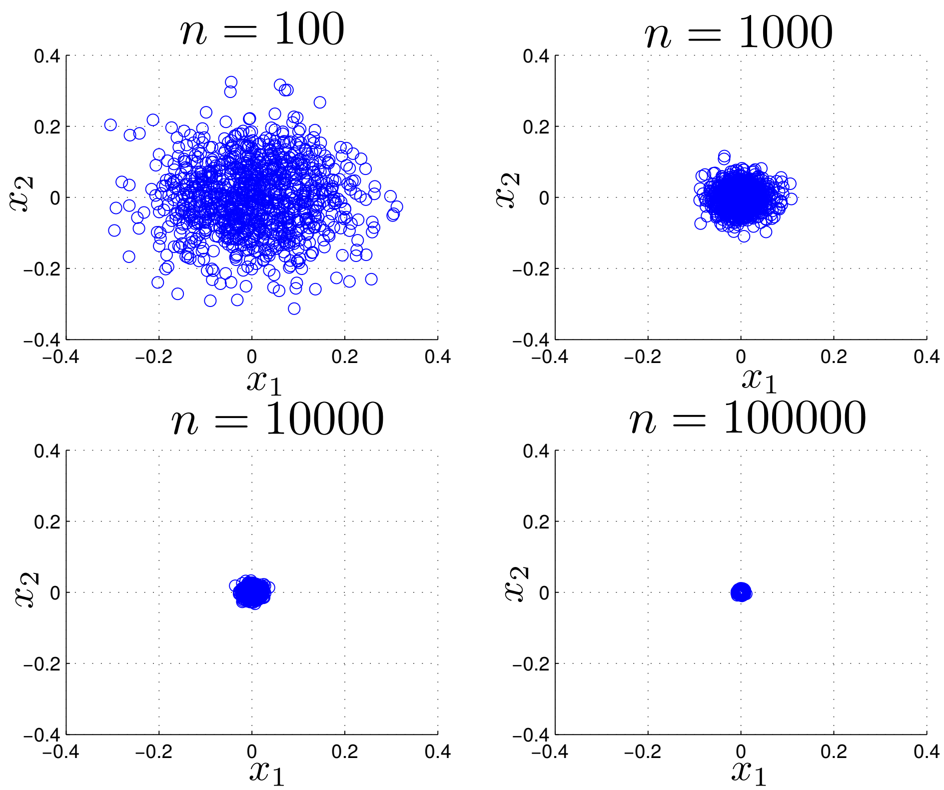
$$\bar{X}_N \xrightarrow{p} \mathbf{E}[\bar{X}_N]$$

if the $X_i$ have common mean $\mu$ then this reduces to $\mathbf{plim}\, \bar{X}_N = \mu$

**strong law of large numbers:** the convergence is instead almost surely

$$\bar{X}_N \xrightarrow{as} \mathbf{E}[\bar{X}_N]$$

## scattergram of $1000$ realizations of the sample mean




- $X_n$ is the sample mean and computed from $n$ samples of $2$-dimensional Gaussian with zero mean
- as $n$ increases, the probability of that $X_n$'s are concentrated at zero is high

# Convergence in Distribution

**Definition:** a random sequence of $X_n$ **converges in distribution** to the continuous random variable $X$, denoted by $X_n \xrightarrow{d} X$ if

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in \mathbf{R}$$

where $F_n$ is CDF of $X_n$ and $F$ is CDF of $X$

- example: $t_{n-1} \xrightarrow{d} \mathcal{N}(0,1)$ ($t$ distribution converges to normal)
- it does not imply that $X_n$ converges at all, e.g.,

$$P(X_n = 1) = 1/2 + 1/(n+1), \quad P(X_n = 2) = 1/2 - 1/(n+1)$$

- convergence in probability implies convergence in distribution

$$X_n \xrightarrow{p} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X$$

# Continuous Mapping Theorem

let $g$ be a continuous function on set $S$ such that $P(X \in S) = 1$

- if $X_n \xrightarrow{p} X$ then $g(X_n) \xrightarrow{p} g(X)$
- if $X_n \xrightarrow{d} X$ then $g(X_n) \xrightarrow{d} g(X)$

- the probability limit can pass through a function if the function is continuous
- useful for determining the asymptotic distribution of test statistics

# Slutsky's Theorem

if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} \alpha$ then

- $X_n + Y_n \xrightarrow{d} X + \alpha$
- $X_n Y_n \xrightarrow{d} \alpha X$
- $X_n / Y_n \xrightarrow{d} X / \alpha$ provided that $P(Y = 0) = 0$

- to find a distribution of the above operations of $(X_n, Y_n)$ we don't need to find a joint distribution of $(X_n, Y_n)$
- also known as rules for limiting distributions

# Product Limit Normal Rule

if $X_N \xrightarrow{d} \mathcal{N}(\mu, A)$ and $H_N \xrightarrow{p} H$ where $H \succ 0$ then

$$H_N X_N \xrightarrow{d} \mathcal{N}(H\mu, HAH^T)$$

example of usage: if we have shown that

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, B)$$

then for any $B_N \succ 0$ that is a consistent estimate for $B$, we have

$$B_N^{-1/2} \cdot \sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, I)$$

# Properties of Estimators

- asymptotic distribution
- unbiased
- consistency (asymptotic properties)
- efficiency (asymptotic properties)

# Asymptotic Distribution of Estimators

suppose that

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \mathcal{N}(0, P)$$

then we say that

- in *large samples* $\hat{\theta}_N$ is $\sqrt{N}$-**asymptotically normally distributed** with

$$\hat{\theta}_N \overset{a}{\sim} \mathcal{N}(\theta, N^{-1}P)$$

- the **asymptotic covariance** of $\hat{\theta}_N$ is $N^{-1}P$, denoted by $\mathbf{Avar}[\hat{\theta}_N]$
- $\widehat{\mathbf{Avar}}[\hat{\theta}_N] = N^{-1}\hat{P}$ denotes the **estimate asymptotic variance matrix** of $\hat{\theta}_N$ where $\hat{P}$ is a consistent estimate of $P$

('in large samples' means $N$ is large enough for $\mathcal{N}(0, P)$ to be a good approximation but not so large that the covariance $N^{-1}P$ goes to zero)

# Unbiased Estimators

an estimator $\hat{\theta}$ of $\theta$ is said to be **unbiased** if

$$\mathbf{E}[\hat{\theta}] = \mathbf{E}[\theta]$$

example: $X_i$'s are i.i.d with mean $\mu$ and variance $\sigma^2$

- the expectation of $\bar{X}$ is carried out by

$$\mathbf{E}[\bar{X}] = \mathbf{E}[(1/N)\sum_{i=1}^{N} X_i] = (1/N)\sum_{i=1}^{N}\mathbf{E}[X_i] = (1/N)\sum_{i=1}^{N}\mu = \mu$$

- one can show that the sample variance satisfies $\mathbf{E}[s^2] = \sigma^2$

hence, the sample mean and the sample variance are unbiased estimators of $\mu$ and $\sigma^2$ respectively

# Consistent Estimators

if a sequence of estimators $\hat{\theta}_N$ of $\theta$, where $N$ is the sample size, satisfies

$$\hat{\theta}_N \xrightarrow{p} \theta \quad \text{for all possible values of } \theta$$

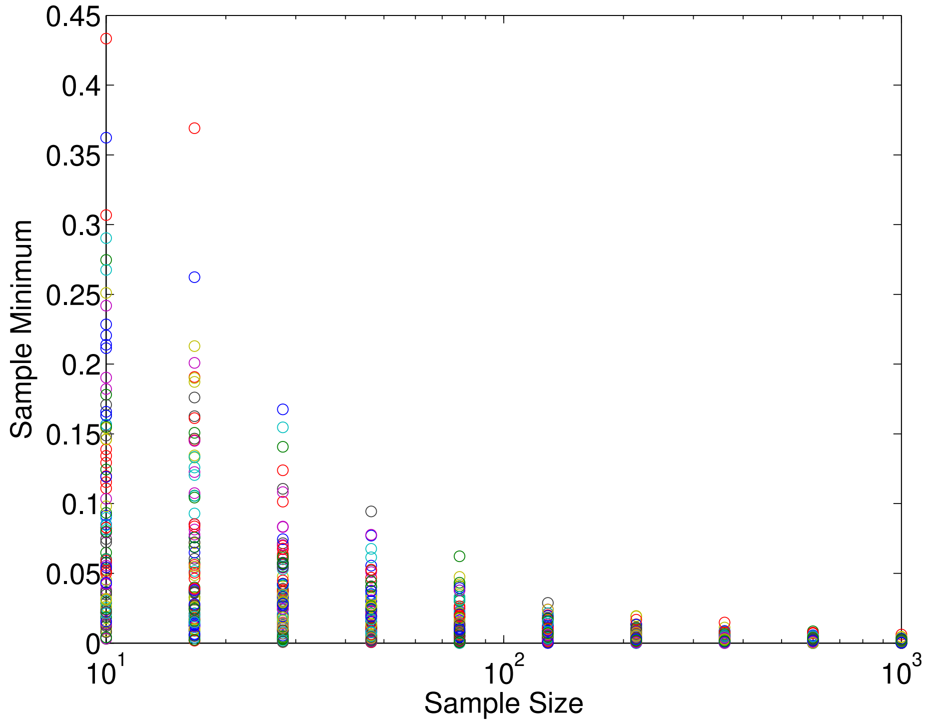
then we say $\hat{\theta}_N$ is a **consistent estimator** of $\theta$

- a consistent estimator converges in probability to the true value
- e.g. the sample mean $\bar{X}_N = (1/N)\sum_i^N X_i$ is a consistent estimator of $\mu$

$$P\left(|\bar{X}_N - \mu)| \geq \epsilon\right) = P\left[\frac{\sqrt{N}|\bar{X}_N - \mu|}{\sigma} \geq \sqrt{N}\epsilon/\sigma\right] = \left(1 - \Phi\left(\frac{\sqrt{N}\epsilon}{\sigma}\right)\right) \to 0$$

as $N \to \infty$ (we have assume $X_i$'s are i.i.d. Gaussian $\mathcal{N}(\mu, \sigma^2)$

**example:** $100$ realizations of the sample minimum of exponential RV




- for each sample size $(N)$, the sample minimum is calculated on $N$ values
- an exponential RV has the minimum at zero
- as $N$ grows, the probability of the sample minimum goes to $0$ is approaching $1$

# Unbiasedness vs Consistency

- unbiasedness needs not imply consistency, *e.g.*, consider i.i.d. sample $X_1, X_2, \ldots, X_n$ of $X$

$$\hat{\theta} \triangleq X_1, \qquad \mathbf{E}[\hat{\theta}] = \mathbf{E}[X_1] = \mathbf{E}[X] \quad \text{(unbiased)}$$

  but $X_1$ never converges to any value (not consistent)

- if the sequence does not converge to a value, then it is not consistent, regardless of whether the estimators in the sequence are biased or not

- consistency needs not imply unbiasedness, *e.g.*, $\hat{\theta}_1 \triangleq \hat{\theta}_N + \frac{1}{N}$
  $\hat{\theta}_1$ is still consistent but not unbiased

# Efficient Estimators

a consistent asympotitcally normal estimator $\hat{\theta}_N$ of $\theta$ is said to be **asympotically efficient** if it has an asymptotic covariance matrix equal to an **efficiency** lower bound

informally, we will have a theorem stating that for any unbiased estimators, it satisfies

$$\mathbf{Avar}[\hat{\theta}_N] \succeq C$$

where $C$ is an important lower bound, derived from the problem statement/assumptions

therefore, if an estimator of interest happens to satisfy

$$\mathbf{Avar}[\hat{\theta}_N] = C$$

then this estimator is **efficient** (since this is the best we can acheive)

# Asymptotic distribution

- definition
- asymptotic efficiency
- the delta method
- asymptotic distribution of nonlinear function

# Asymptotic distribution

**definition:** a distribution that is used to approximate the true finite samples distribution of a random variable

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \mathcal{N}(0, P)$$

we say that in *large samples* $\hat{\theta}_N$ is $\sqrt{N}$-**asymptotically normally distributed** with

$$\hat{\theta}_N \overset{a}{\sim} \mathcal{N}(\theta, N^{-1}P)$$

note that

- the asymptotic distribution is an approximation of the exact distribution
- example: $X_1, \ldots, X_N$ are i.i.d. samples of exponential with parameter $\lambda$
- the exact distribution of $\bar{X}_N$ is $(1/2\lambda N) \cdot \mathcal{X}^2(2N)$
- the asymptotic distribution is $\mathcal{N}(1/\lambda, 1/N\lambda^2)$

# Asymptotic normality and efficiency

suppose that

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \mathcal{N}(0, P)$$

then we say

- $\hat{\theta}_N$ is asymptotically normal
- $\hat{\theta}_N$ is **asymptotically efficient** if the covariance matrix of any other consistent, asymptotically normally distributed estimate exceeds $P/N$ by a non-negative definite matrix

# The delta method

assumptions:

- $\sqrt{N}(x_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$
- $g(x)$ and $g'(x)$ are continuous
- $g'(\mu) \neq 0$ and does not involve $N$

then the delta theorem states that

$$\sqrt{N}(g(x_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, \sigma^2(g'(\mu)^2))$$

*proof.* use the linear Taylor approximation to write

$$\sqrt{N}(g(x_n) - g(\mu)) = g'(\zeta) \cdot \sqrt{N}(x_n - \mu), \quad x_n \leq \zeta \leq \mu$$

and then use the continuous mapping theorem and the product limit normal rule

# Asymptotic distribution of nonlinear function

suppose that

- $\hat{\theta}_N \in \mathbf{R}^n$ is an estimator (vector) such that $\hat{\theta}_N \overset{a}{\sim} \mathcal{N}(\theta, P/N)$
- $f(\theta) : \mathbf{R}^n \to \mathbf{R}^m$ is a *continuous* function with Jacobian $J(\theta) = \partial f(\theta)/\partial\theta$
- $f(\theta)$ is not a function of $N$

then $f(\hat{\theta}_N)$ is also asymptotically normally distributed with

$$f(\hat{\theta}_N) \overset{a}{\sim} \mathcal{N}(f(\theta), (1/N)J(\theta)PJ(\theta)^T)$$

where $\mathbf{Avar}(f(\hat{\theta}_N)) = (1/N)J(\theta)PJ(\theta)^T$ is the asymptotic covariance of $f(\hat{\theta}_N)$

the covariance is quadratically scaled by the Jacobian of $f$

**Proof.**

- by mean-value theorem, $\exists \tilde{\theta} = \alpha\theta + (1-\alpha)\hat{\theta}_N$ with $0 \leq \alpha \leq 1$ we can write

$$\sqrt{N}[f(\hat{\theta}_N) - f(\theta)] = J(\tilde{\theta})\sqrt{N}(\hat{\theta}_N - \theta)$$

- if $\hat{\theta}_N \xrightarrow{p} \theta$ then $\tilde{\theta} \xrightarrow{p} \theta$ because $\tilde{\theta}$ lies between $\hat{\theta}_N$ and $\theta$ and therefore

$$P(\|\hat{\theta}_N - \theta\| \leq \epsilon) \leq P(\|\tilde{\theta} - \theta\| \leq \epsilon) \leq 1$$

- by continuity assumption on $J$ then $J(\tilde{\theta}) \xrightarrow{p} J(\theta)$
- apply the product limit normal rule

# Estimators

- statistics as estimators
- convergence
- properties of estimators
- **sample mean and sample variance**

# Sampling statistics

- useful inequalities
- central limit theorem
- sample mean and sample variance

# Markov and Chebyshev Inequalities

## Markov inequality

let $X$ be a *nonnegative* RV with mean $\mathbf{E}[X]$

$$P(X \geq a) \leq \frac{\mathbf{E}[X]}{a}, \quad a > 0$$

## Chebyshev inequality

let $X$ be an RV with mean $\mu$ and variance $\sigma^2$

$$P\left(|X - \mu| \geq a\right) \leq \frac{\sigma^2}{a^2}$$

# Sample mean

let $X$ be an RV with $\mathbf{E}[X] = \mu$ (unknown)

$X_1, X_2, \ldots, X_N$ denote $N$ independent, repeated measurements of $X$

$X_j$'s are *independent, identically distributed* (i.i.d.) RVs

the **sample mean** of the sequences is used to estimate $\mathbf{E}[X]$:

$$\bar{X} = \frac{1}{N} \sum_{j=1}^{N} X_j$$

two statistical quantities for characterizing the sample mean's properties:

- $\mathbf{E}[\bar{X}]$: we say $\bar{X}$ is unbiased if $\mathbf{E}[\bar{X}] = \mu$
- $\mathbf{var}(\bar{X})$: we examine this value when $N$ is large

the sample mean is an **unbiased estimator** for $\mu$:

$$\mathbf{E}[\bar{X}] = \mathbf{E}\left[\frac{1}{N}\sum_{j=1}^{N} X_j\right] = \frac{1}{N}\sum_{j=1}^{N}\mathbf{E}[X_j] = \mu$$

suppose $\mathbf{var}(X) = \sigma^2$ (true variance)

since $X_j$'s are i.i.d, the variance of $\bar{X}$ is

$$\mathbf{var}(\bar{X}) = \frac{1}{N^2}\sum_{j=1}^{N}\mathbf{var}(X_j) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

hence, the variance of the sample mean approaches zero as the number of samples increases

# Weak Law of Large Numbers

let $X_1, X_2, \ldots, X_N$ be a sequence of i.i.d. RVs with finite mean $\mathbf{E}[X] = \mu$ and variance $\sigma^2$

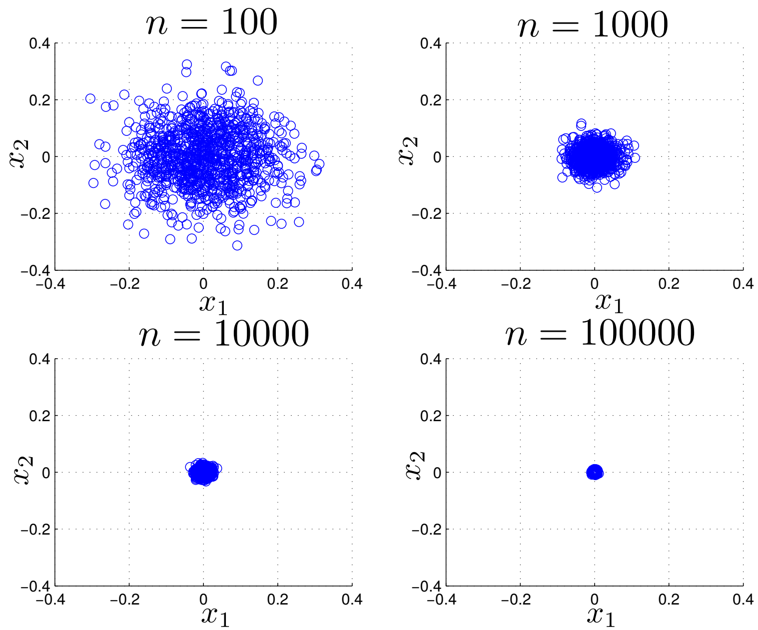
for any $\epsilon > 0$,

$$\lim_{N \to \infty} P[\ |\bar{X} - \mu| < \epsilon\ \ ] = 1$$

- for large enough $N$, the sample mean will be close to the true mean with high probability
- *Proof.* apply Chebyshev inequality:

$$P[|\bar{X} - \mu| \geq \epsilon] \leq \frac{\sigma^2}{N\epsilon^2} \quad \Longrightarrow \quad P[|\bar{X} - \mu| < \epsilon] \geq 1 - \frac{\sigma^2}{N\epsilon^2}$$

## scattergram of $1000$ realizations of the sample mean




- $\bar{X}$'s are computed from $2$-dimensional Gaussian with zero mean
- as $N$ increases, the probability of $\bar{X}$'s are concentrated at zero is high

# Strong Law of Large Numbers

let $X_1, X_2, \ldots, X_N$ be a sequence of iid RVs with finite mean $\mathbf{E}[X] = \mu$ and finite variance, then

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

- $\bar{X}_k$ is the sequence of sample mean computed using $X_1$ through $X_k$
- with probability $1$, every sequence of sample mean calculations will eventually approach and stay close to $\mathbf{E}[X] = \mu$
- the strong law implies the weak law

# Central Limit Theorem (CLT)

let $X_1, X_2, \ldots, X_N$ be a sequence of i.i.d. RVs with

finite mean $\mathbf{E}[X] = \mu$ and finite variance $\sigma^2$

let $S_N$ be the sum of the first $N$ RVs in the sequences:

$$S_N = X_1 + X_2 + \cdots + X_N$$

and define

$$Z_N = \frac{S_N - N\mu}{\sigma\sqrt{N}}$$

then

$$\lim_{N\to\infty} P(Z_N \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

as $N$ becomes large, the CDF of normalized $S_n$ approaches Gaussian distribution

# Proof of Central Limit Theorem

first note that

$$Z_N = \frac{S_N - N\mu}{\sigma\sqrt{N}} = \frac{1}{\sigma\sqrt{N}} \sum_{k=1}^{N} (X_k - \mu)$$

the characteristic function of $Z_N$ is given by

$$\begin{aligned}
\Phi_{Z_N}(\omega) &= \mathbf{E}[e^{\mathrm{i}\omega Z_N}] = \mathbf{E}\left[\exp \frac{\mathrm{i}\omega}{\sigma\sqrt{N}} \sum_{k=1}^{N} (X_k - \mu)\right] \\
&= \mathbf{E}\left[\prod_{k=1}^{N} e^{\mathrm{i}\omega(X_k-\mu)/\sigma\sqrt{N}}\right] \\
&= \left(\mathbf{E}[e^{\mathrm{i}\omega(X-\mu)/\sigma\sqrt{N}}]\right)^N
\end{aligned}$$

(using the fact that $X_k$'s are iid)

expanding the exponential expression gives

$$\begin{aligned}
\mathbf{E}[e^{\mathrm{i}\omega(X-\mu)/\sigma\sqrt{N}}] &= \mathbf{E}\left[1 + \frac{\mathrm{i}\omega}{\sigma\sqrt{N}}(X-\mu) + \frac{(\mathrm{i}\omega)^2}{2!N\sigma^2}(X-\mu)^2 + \ldots\right] \\
&\approx 1 - \frac{\omega^2}{2N}
\end{aligned}$$

(the higher order term can be neglected as $N$ becomes large)

then we obtain

$$\begin{aligned}
\Phi_{Z_N}(\omega) &\to \left(1 - \frac{\omega^2}{2N}\right)^N \\
&\to e^{-\omega^2/2}, \quad \text{as } N \to \infty
\end{aligned}$$

# Multivariate CLT

**Lindeberg-Levy Theorem:** let $X_1, X_2, \ldots, X_N$ be an i.i.d. sequence of random vectors with $\mathbf{E}[X_i] = \mu$ and $\mathbf{cov}(X_i) = \Sigma$ such that the second moment of each component in $X_i$ is finite

define $\bar{X}_N = (1/N)\sum_{i=1}^N X_i$

CLT says that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(X_i - \mathbf{E}[X_i]) = \sqrt{N}(\bar{X}_N - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

more conditions involved if $X_i$'s are NOT i.i.d.

# Multivariate CLT

**Lindeberg-Feller Theorem:** let $X_1, X_2, \ldots, X_N$ be samples of random vectors with $\mathbf{E}[X_i] = \mu_i$ and $\mathbf{cov}(X_i) = C_i$ such that all mixed third moments are finite

moreover, assume that for every $i$

$$\lim_{N\to\infty} \left(\sum_{i=1}^{N} C_i\right)^{-1} C_i = 0, \quad \text{and} \quad C = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} C_i$$

exists and is positive definite

define $\bar{X}_N = (1/N)\sum_{i=1} X_i$ and $\bar{\mu}_N = (1/N)\sum_{i=1} \mu_i$

then CLT says that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(X_i - \mu_i) = \sqrt{N}(\bar{X}_N - \bar{\mu}) \xrightarrow{d} \mathcal{N}(0, C)$$

# Distribution of $\bar{X}$

let $X_1, \ldots, X_N$ be a sample from a population with mean $\mu$ and variance $\sigma^2$

let $\bar{X} = (1/N)\sum_{i=1}^N X_i$ be the sample mean

- if $X_i$ is **normal**, then $\bar{X}$ is also **normal** with mean $\mu$ and variance $\sigma^2/N$
- from the central limit theorem, the sample mean is **approximately normal** when $N$ is large where
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$
has *approximately* a **standard normal** distribution

# Sample Variance

let $X_1, \ldots, X_N$ be a sample from a population with mean $\mu$ and variance $\sigma^2$

the statistic $s^2$, defined by

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$$

is called the **sample variance**

- $s = \sqrt{s^2}$ is called the sample standard deviation
- using $(N-1)s^2 = \sum_{i=1}^N X_i^2 - N\bar{X}^2$, we have

$$\mathbf{E}[s^2] = \sigma^2 \quad \text{(equal to population variance)}$$

# Joint distribution of sample mean and variance

let $X_1, \ldots, X_N$ be a sample from a **normal** popolution, we obtain the identity

$$\sum_{i=1}^{N} \left( \frac{X_i - \mu}{\sigma} \right)^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{\sigma^2} + \left[ \frac{\sqrt{N}(\bar{X} - \mu)}{\sigma} \right]^2$$

- LHS is a chi-square of $N$ degrees of freedom
- the second term on RHS is a chi-square with $1$ degree of freedom
- the sum of two independent chi-squares with $N$ and $M$ DFs is a chi-square with $N + M$
- it would seem that the first term on RHS is a chi-square with $N - 1$ degree of freedoms

**Theorem:** if $X_1, \ldots, X_N$ is a sample from a **normal** population with mean $\mu$ and variance $\sigma^2$

- $\bar{X}$ is normal with mean $\mu$ and variance $\sigma^2/N$
- $(N-1)s^2/\sigma^2$ is chi-square with $N-1$ degrees of freedom
- $\bar{X}$ and $s^2$ are **independent**

**Corollary:** let $s$ be the sample standard deviation

$$\frac{\sqrt{N}(\bar{X}-\mu)}{s} \sim t_{N-1}$$

followed from

$$\frac{\sqrt{N}(\bar{X}-\mu)}{s} = \frac{\sqrt{N}(\bar{X}-\mu)/\sigma}{\sqrt{s^2/\sigma^2}} \triangleq \frac{\text{standard normal}}{\text{chi-square with } N-1 \text{ DF}}$$

# References

Chapter 3 in

J.M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, the MIT press, 2010

Appendix in

W.H. Greene, *Econometric Analysis*, 7th edition, Pearson, 2012

Chapter 6 in

R.A. DeFusco, D.W. McLeavey, J.E.Pinto and D.E. Runkle, *Quantitative Investment Analysis*, 2nd edition, Wiley, 2004

Chapter 6 in

S.M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 4th edition, Academic press, 2009