ECON3206/5206: Review of CLT & LLN

Dr. Rachida Ouysse

School of Economics
UNSW

- LLN and CLT – pillars of all statistics
  - Let $\{Z_1, Z_2, \dots, Z_T\}$ be a set of *independent* RVs with common mean μ and variance σ^2.
  - **Law of large numbers:** the probability that $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t$ differs from μ converges to zero as T goes to infinity.
  - **Central limit theorem:** the distributions of $\frac{\bar{Z}-\mu}{\sqrt{\mathrm{Var}(\bar{Z})}}$ converges to $N(0,1)$ as T goes to infinity.
  - Note that $\mathrm{Var}(\bar{Z}) = \sigma^2/T$ (see Rule 8).

    What happens if $\{Z_1, Z_2, \dots, Z_T\}$ are correlated?

Distribution of Sample Mean

Consider N random variables $X_1, \cdots X_N$.

- Let's consider $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$.
- $\bar{X}$ is called the "sample mean" or the "empirical mean".
- $\bar{X}$ is a random variable.

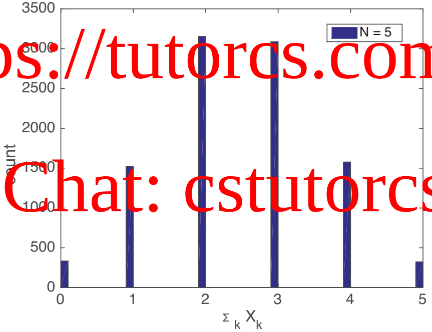
Suppose we observe values for $X_1, \cdots X_N$ and calculate the empirical mean of the observed values. That gives us one value for $\bar{X}$. But the value of $\bar{X}$ changes depending on the observed values.

- Suppose we toss a fair coin $N = 5$ times and get H, H, H, T, T. Let $X_k = 1$ when come up heads. Then $\frac{1}{N}\sum_{k=1}^{N} X_k = \frac{3}{5}$
- Suppose we toss the coin another $N = 5$ times and get T, T, H, T, H. Now $\frac{1}{N}\sum_{k=1}^{N} X_k = \frac{2}{5}$

Distribution of Sample Mean

Toss fair coin $N = 5$ times and calculate $\sum_{k=1}^{N} X_k$. Repeat, and plot histogram of values. Its binomial $Bin(N, \frac{1}{2})$.

- X=[];for i=1:10000,X=[X,sum((rand(1,5)<0.5))]; end; hist(X,50)



Distribution of Sample Mean

Random variable $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$.

- Suppose the X_k are **independent and identically distributed** (i.i.d)
- Each X_k has mean $E(X_k) = \mu$ and variance $Var(X_k) = \sigma^2$.

Then we can calculate the mean of $\bar{X}$ as:

$$E(\bar{X}) = E(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N}\sum_{k=1}^{N} E(X_k) = \mu$$

NB: recall linearity of expectation: $E(X+Y) = E(X) + E(Y)$ and $E(aX) = aE[X]$

- We say $\bar{X}$ is an **unbiased estimator** of μ since $E[\bar{X}] = x$

Distribution of Sample Mean

We can calculate the variance of $\bar{X}$ as:

$$var(\bar{X}) = var(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N^2} var(\sum_{k=1}^{N} X_k)$$

$$= \frac{1}{N^2}\sum_{k=1}^{N} var(X_k) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

NB: recall $Var(aX) = a^2 Var(X)$ and $Var(X+Y) = Var(X) + Var(Y)$ when X, Y independent.

- As N increases, the variance of $\bar{X}$ falls.
- $Var(NX) = N^2 Var(X)$ for random variable X.
- But when add together **independent** random variables $X_1 + X_2 + \cdots$ the variance is only $NVar(X)$ rather than $N^2 Var(X)$
- This is due to **statistical multiplexing**. Small and large values of X_i tend to cancel out for large N.

Weak Law of Large Numbers[1]

Consider N independent identically distributed (i.i.d) random variables $X_1, \cdots X_N$ each with mean μ and variance σ^2. Let $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$. For any $\epsilon > 0$:

$$P(|\bar{X} - \mu| \geq \epsilon) \to 0 \text{ as } N \to \infty$$

That is, $\bar{X}$ **concentrates** around the mean μ as N increases.
Proof:

- $E(\bar{X}) = E(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N}\sum_{k=1}^{N} E(X_k) = \mu$
- $var(\bar{X}) = var(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$
- By Chebyshev's inequality: $P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}$

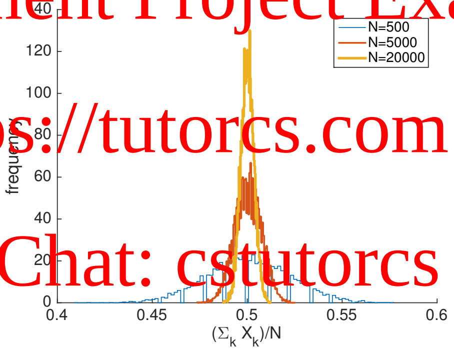
[1] There is also a **strong law of large numbers**, but we won't deal with that here.

Who cares ?

- Suppose we have an event E
- Define indicator random variable X_i equal to 1 when event E is observed in trial i and 0 otherwise
- Recall $E[X_i] = P(E)$ is the probability that event E occurs.
- $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$ is then the relative frequency with which event E is observed over N experiments.
- And ...
  $P(|\bar{X} - \mu| \geq \epsilon) \rightarrow 0$ as $N \rightarrow \infty$
  tells us that this observed relative frequency $\bar{X}$ converges to the probability $P(E)$ of event E as N grows large.
- So the law of large numbers formalises the intuition of probability as frequency when an experiment can be repeated many times. But probability still makes sense even if cannot repeat an experiment many times – all our analysis still holds.

Central Limit Theorem (CLT)

Histogram of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as N increases, but now we normalise to keep the area under the curve fixed:

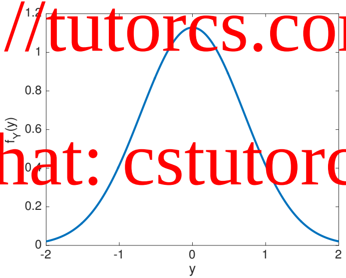


- See that (i) curve narrows as n increases, it concentrates as we already know from weak law of large numbers.
- Curve becomes more "bell-shaped" as N increases – this is the CLT

The Normal (or Gaussian) Distribution

Define the following function

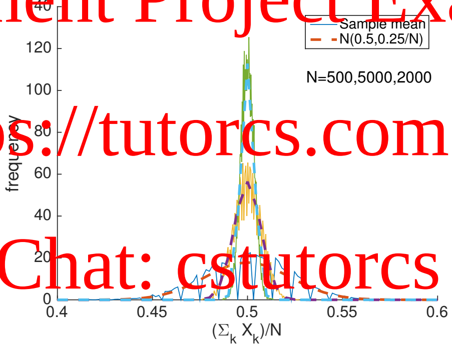
$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$



$\mu = 0, \, \sigma = 1$

Central Limit Theorem (CLT)

Overlay the Normal distribution, with parameter μ equal to the mean and σ^2 equal to the variance of each of the measured histograms:



- CLT says that as N increases the distribution of $\bar{X}$ converges to a Normal (or Gaussian) distribution.
- Variance $\to 0$ as $N \to \infty$, i.e. distribution concentrates around its mean as N increases.