Week 10 Lecture II - Central Limit Theorem

Yilin

DigiPen

BELL CURVES

Introduction

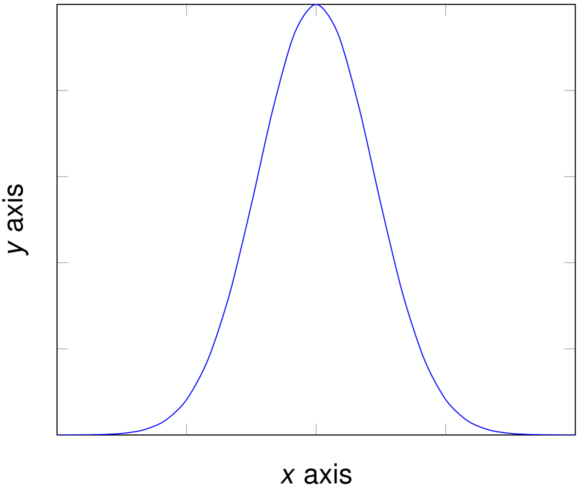
- The **normal distribution** is one of the most important probability distributions.
- It has applications everywhere in probability and statistics, and beyond.
- The study of this distribution starts with the study of **bell curves**.

Question

Why do bell curves appear so frequently when modelling distributions of seemingly unrelated phenomena?

Bell Curves

Ideally, a bell curve is a curve with a “bell” shape:



Example 1: 16 Flips

- Flip a fair coin is 16 times.
- Let X be the number of heads.

Recall, X is a binomial r.v. with $n = 16$ and $p = 0.5$, for $k = 0, 1, \ldots, 16$. So

$$\Pr(k \text{ heads}) = \Pr(X = k) = \binom{16}{k}(0.5)^k(0.5)^{16-k} = \binom{16}{k}(0.5)^{16}.$$

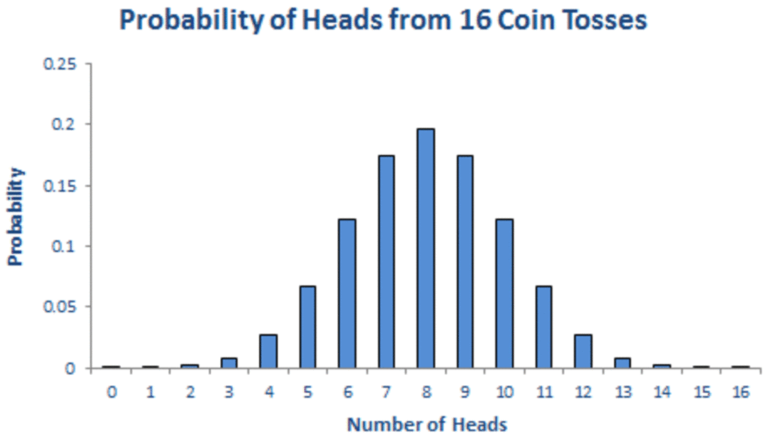
Note: This distribution is symmetric about its mean, $E(X) = np = 8$ since

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}.$$

e.g.

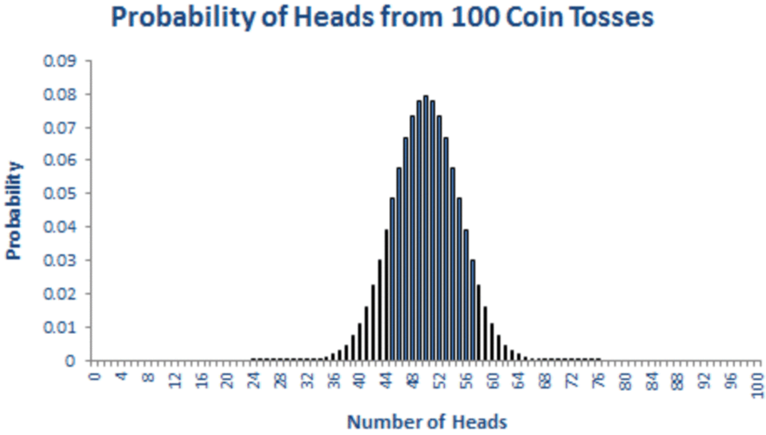
$$\binom{16}{0} = \binom{16}{16}, \quad \binom{16}{1} = \binom{16}{15}, \ldots$$

Example 1: Histogram for 16 Flips



Example 2: Histogram for 100 Flips

Suppose we flip the coin 100 times, the histogram for the distribution is:



Example 3: Shoes

Here is a histogram for quantity of shoes, by size, sold in the US in 1998:

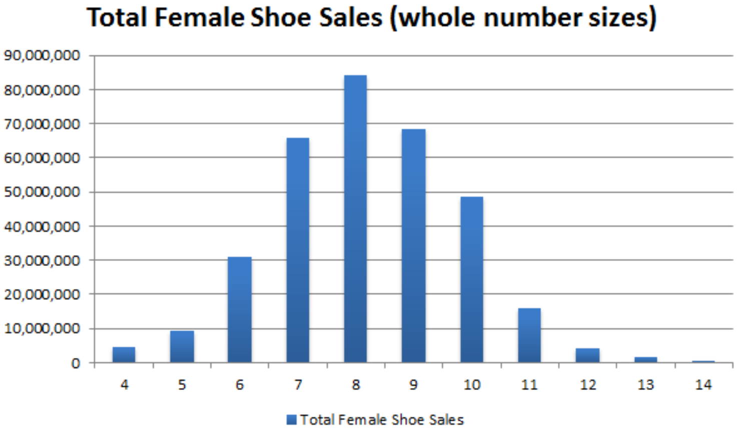


Figure: `https://thoughtburner.org/tag/normal-distribution/`

Example 4: SAT Scores

Here is a histogram for SAT scores in 2010.

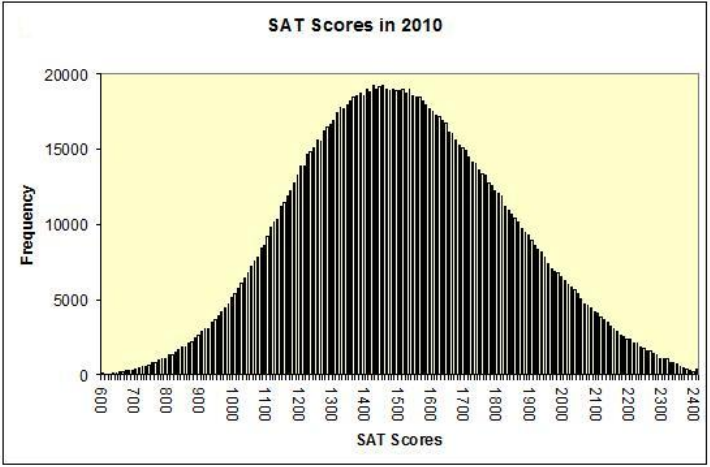


Figure: `http://ptrow.com/articles/Galton_June_07.htm`

Question

Question

Why is this bell-shape so common?

We will answer this shortly..

NORMAL DISTRIBUTION

Standard Normal

Definition (Standard Normal)

A continuous r.v. X has a **standard normal distribution** if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

We call this the **normal distribution with mean 0 and variance 1** or just $X \sim N(0, 1)$.

We see that $E(X) = 0$ and $\text{var}(X) = 1$ since

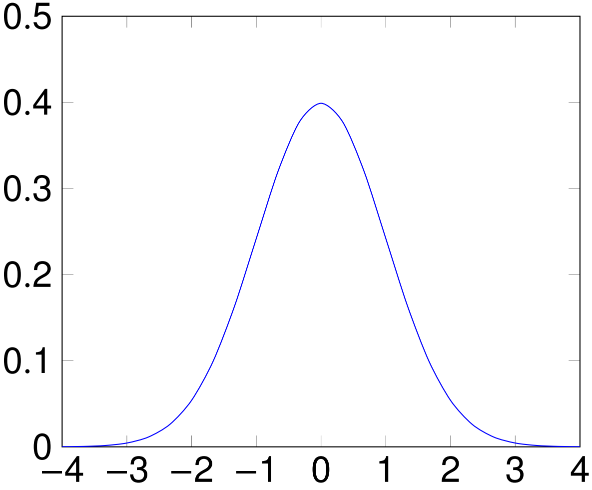
$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} \, dx = 0,$$

and

$$\text{var}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} \, dx - (E(X))^2 = 1.$$

Graph of pdf for $N(0,1)$

Graph of $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$:



CDF for Standard Normal

- The family of functions e^{-ax^2} have no elementary anti-derivative
- I.e., we cannot directly compute the cdf in terms of standard functions.
- We can still define the cdf to be the integral of it and use numerical methods to approximate it.

Definition

The cdf of the standard normal distribution is denoted by a capital Phi, Φ. If X has a standard normal distribution, then we define

$$\Phi(x) := \Pr(X \leq x) = \int_{-\infty}^{x} f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t.$$

Example 5: A normally distributed r.v.

Suppose X is $N(0,1)$ distributed.

a. What is the probability that X is negative?
b. What is the probability that X lies between -1 and 1?

Solution:

a. We can use an online normal distribution calculator or symmetry to compute

$$\Pr(X < 0) = \int_{-\infty}^{0} f(x)\,dx = \Phi(0) = \frac{1}{2}.$$

b. Using an online normal distribution calculator,

$$\Pr(-1 \leq X \leq 1) = \int_{-1}^{1} f(x)\,dx = \Phi(1) - \Phi(-1) = 0.6827.$$

Normal Distribution with arbitrary Mean and Variance

Definition

We say a r.v. X has a **normal distribution** with mean μ, 'mu', and variance, σ^2, if it distributed via the pdf

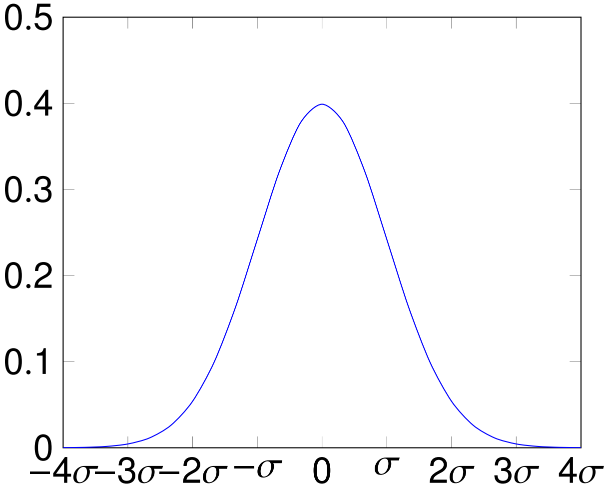
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}.$$

Via calculus one can show that the CDF is a transformation of the CDF from the standard normal,

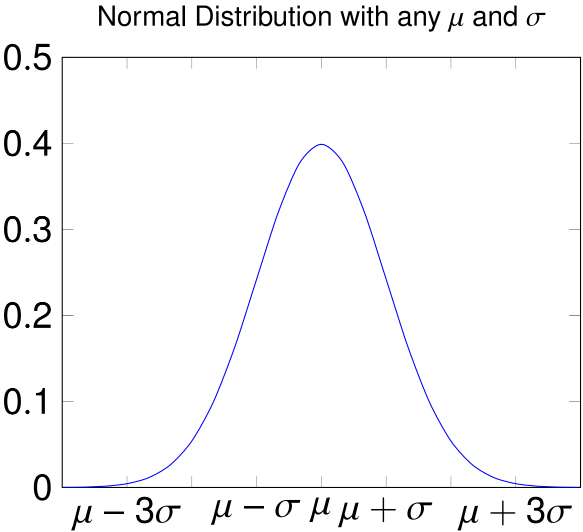
$$F(x) = \Pr(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

We say $X \sim N(\mu, \sigma^2)$.

Graph of Normal Distribution with $\mu = 0$, any σ



Graph of Normal Distribution with any μ, σ



LAW OF LARGE NUMBERS

Samples

Consider any real-life experiment, e.g.

- scores on a test
- heights of a generation of plants
- price of a stock over a time period

In real-life scenarios such as these we usually **do not** know the underlying probability distribution.

- At the core of statistics is trying to form hypotheses off of **samples**.
- That is, we run an experiment and have a certain number, n, of **observable** outcomes.
- From this, we can compute certain quantaties to better understand the data we are given.

One such quantity is the sample average.

Sample Average

Definition (Sample Average)

Suppose $X_1, X_2, \ldots, X_n$ are randomly chosen samples from a the same experiment. Then each is an r.v. and the **sample average** of these is defined to be their arithmetic mean:

$$S_n := \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Example 6: Sample Average of Body Weight

Suppose we measure the body weight of 6 people.

- Let $X_1, X_2, \ldots, X_6$ be r.v.s representing the (random) weights of 6 people respectively.
- Suppose we record that they weigh 175, 102, 194, 152, 120, and 133 lbs respectively.

That is, we observe the events

$$(X_1 = 175),\ (X_2 = 102),\ (X_3 = 194),\ (X_4 = 152),\ (X_5 = 120),\ (X_6 = 13$$

From this, we observe the event

$$\left(S_n = \frac{175 + 102 + 194 + 152 + 120 + 133}{6} = 146\right).$$

Note: S_n is the sample average, which is an r.v. The number 146 is just one outcome from the distribution of S_n.

Independent and Identically Distributed

If all of these samples come from the same experiment, we usually say they are **identically distributed**.

Definition (Identically Distributed)

We call r.v.s $X_1, X_2, \ldots, X_n$ **identically distributed** if each has the same underlying probability distribution.

If they are also independently selected, we say..

Definition (Independent and Identically Distributed)

We call r.v.s $X_1, X_2, \ldots, X_n$ **independent and identically distributed** if each is independent of each other and they are all identically distributed. The shorthand for this is **i.i.d.**.

Example 7: Mouse Diet

Suppose we run an experiment trying to measure how much food a mouse eats in one day.

- We have 100 mice as test subjects.
- Let $X_1, X_2, \ldots, X_{100}$ be r.v.s representing the amount of food mouse 1 thru 100 eats.
- If each mouse is separately caged from one another, then X_1 thru X_{100} are independent.
- If each is given the same amount of food (assuming all else is the same), then X_1 thru X_{100} come from the same probability distribution.

Then $X_1, \ldots, X_{100}$ are **i.i.d.**.

Theorem: Law of Large Numbers

If we have a large number of i.i.d. samples, the following theorem says we can approximate the **sample average** S_n by the **true average** μ.

Theorem (Law of Large Numbers)

Suppose $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d. r.v.s, and the underlying distribution has finite mean and variance, i.e. $E(X_1) = \mu$ is finite and var$(X) = \sigma^2 < \infty$. Then the sample average converges to μ with probability 1, that is

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu.$$

Interpretation

Interpretation:

- For very large n, the sample average closely approximates the true average.
- In other words, if we run an experiment(with finite average outcome and variance) , then if we have “enough” samples (say, 10,000), then we can expect the sample average S_n to be very close to the true average μ.
- If we do not know μ (usually don’t) then the sample average serves as a good estimator for μ, depending on σ.
- Since we may not know the rate of this convergence, we use other statistical methods to measure “confidence” that our measured average is as close as we want to the actual mean.

A Note on the Sample Average

Note that if X_1 thru X_n are identically distributed,

$$\begin{aligned} E(S_n) &= E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\ &= \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} \\ &= \frac{n\mu}{n} = \mu \end{aligned}$$

- Obviously the average sample average will be the same as the true average.
- The law of large numbers says the sample average **ceases to be random** as the number of samples tends to infinity.
- This is becuase the sample average tends exactly to the true average.

CENTRAL LIMIT THEOREM

Central Limit Theorem

The Central Limit Theorem serves to tell us how sample averages are distributed for large samples.

Theorem (Central Limit Theorem)

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. r.v.s where $E(X_1) = \mu$ is finite and $var(X_1) = \sigma^2 < \infty$. Then there is a normal r.v. S, where $S \sim N(\mu, \sigma^2/n)$, so that

$$S_n \approx S, \text{ for } n \text{ large.}$$

In other words, for **large** *n,*

$$\Pr(-\infty \le S_n \le x) \approx \Pr(-\infty \le S \le x). = \sqrt{\frac{n}{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-n(t-\mu)^2/2\sigma^2}\, dt$$

$$= \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right)$$

Example 7: CLT applied to Mice

Suppose we are modelling the amount of food a mouse eats in a day. Assume

- We know beforehand (a priori) that the average mouse eats 3 grams of food in a day,
- The standard deviation from this is 2 grams.

a. If we have 36 independent mice, approximate the probability that the sample average of food eaten is between 2.5 and 3.5 grams.
b. If we have 36 independent mice, approximate the probability that the sample average of food eaten is between 3 and 4 grams.

Exmaple 7: Solutions

Let $S_{36} = \dfrac{X_1 + \cdots + X_{36}}{36}$ be the sample average, where each X_i is the mass of the food that mouse i eats. Then $E(X_i) = 3$, and $\sigma_{X_i} = 2$.

a. The probability S_{36} is between 2.5 and 3.5 is

$$\begin{aligned}\Pr(2.5 \le S_{36} \le 3.5) &\approx \Phi\left(\frac{\sqrt{36}(3.5-3)}{2}\right) - \Phi\left(\frac{\sqrt{36}(2.5-3)}{2}\right)\\ &= \Phi(1.5) - \Phi(-1.5)\\ &\approx 0.8664\end{aligned}$$

b. The probability S_{36} is between 3 and 4 is

$$\begin{aligned}\Pr(3 \le S_{36} \le 4) &\approx \Phi\left(\frac{\sqrt{36}(4-3)}{2}\right) - \Phi\left(\frac{\sqrt{36}(3-3)}{2}\right)\\ &= \Phi(5) - \Phi(0)\\ &\approx 0.4987.\end{aligned}$$

Corollary to the CLT: Sums

Theorem

With the same conditions on $X_1, X_2, \ldots,$ we have that the r.v. $Z_n = X_1 + \cdots + X_n$, is approximately distributed $N(n\mu, n\sigma^2)$ for large n.

Example: If we have 40 test scores where the mean score is 70 and the standard deviation is 15, let X_1 thru X_{40} be each score. Then

$$Z_{40} = X_1 + \cdots + X_{40} \sim N(40(70), 40(15^2)) = N(2800, 9000).$$

E.g., the probability that the sum of scores is less than 2700 is

$$\Pr(Z_{40} \leq 2700) = \Phi\left(\frac{2700 - 2800}{\sqrt{9000}}\right) \approx \Phi(-1.05) \approx 0.1469.$$