

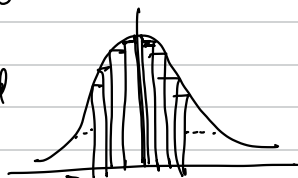
LECTURE 4: Gaussians, Sum of random variables, Central Limit Theorem

References:

- Intro to Probability Theory by Feller
- Ryan's Lecture 4.
- Terence Tao's blog on Central Limit Theorem.

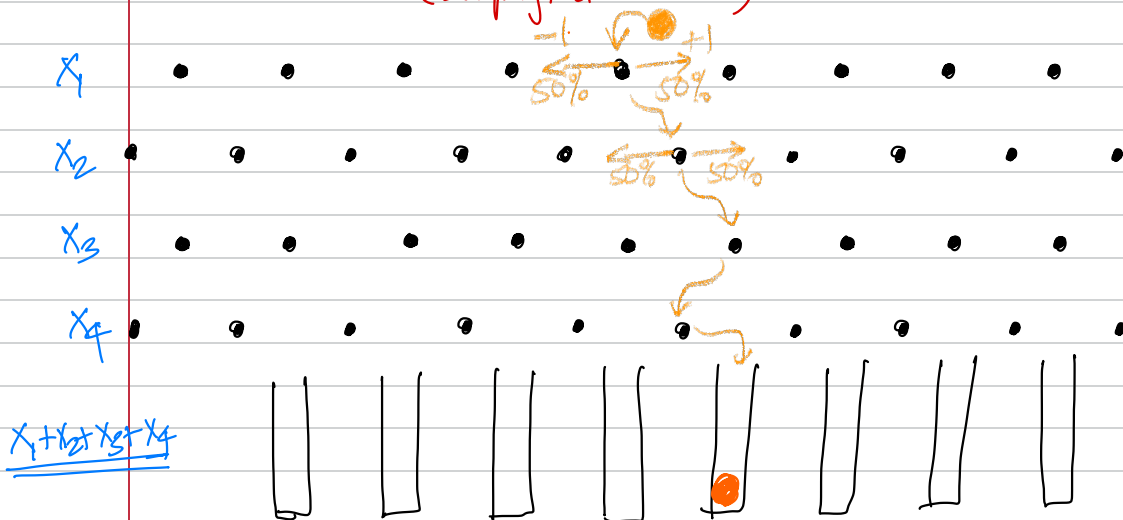
Examples

- ① Take large # of people sitting in "similar demographic" and plot their heights, these heights tend to follow a "normal distribution"
- ② Take large # of natural numbers n and ask # distinct prime factors of n then this answer closely looks like certain normal distributions (Erdős-Kac Theorem)



Normal Distribution shows up very commonly!

GALTON BOARD (Simplified model)



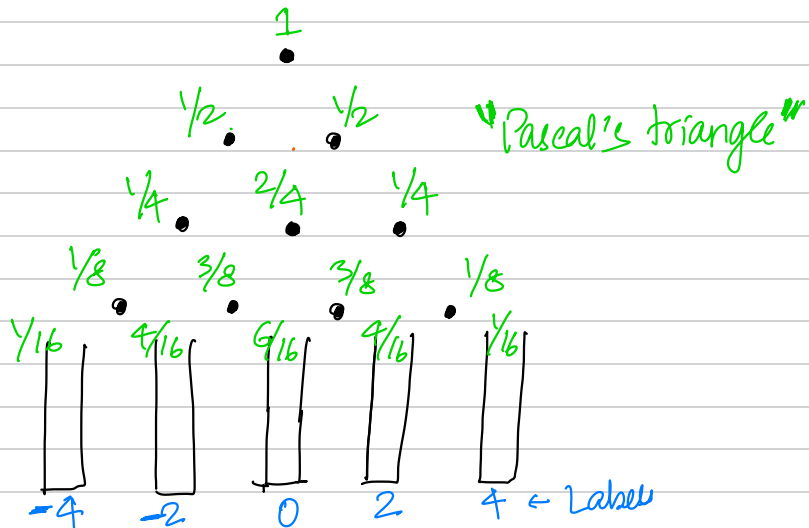
Some (unrealistic) assumptions

- The ball hits middle of these pins on each stage of it's descent.
- Each time it has 50% chance of going left (in which case you add (-1)) & 50% chance of going right (in which case you add $(+1)$)

$$\text{R.V. } X_i = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$$

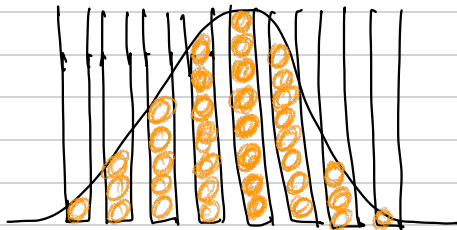
rows here = n (in above drawing it's 4)
then the sum $X_1 + X_2 + X_3 + X_4$ will be the label of the different buckets

- Make another assumption that you do this for many balls falling down → assuming they don't influence each other.



Evaluating probabilities of falling into a bucket

Informally $\rightarrow$ as you increase # rows $= n$ in the above Galton board ($n \rightarrow \infty$), then the distribution of where the ball falls "looks like" a bell curve!



This is what "Central Limit Theorem" captures.

Setup

$X_1, X_2, \dots, X_n$ be i.i.d random variables

$\downarrow$
independent and identically distributed

$\Pr\{X_i=1\}=p, \Pr\{X_i=0\}=q=1-p. \quad (X_i \sim \text{Bernoulli}(p))$

Let $S_n = X_1 + X_2 + \dots + X_n \rightarrow$ want to understand this!

$$\mathbb{E}[S_n] = np \quad \text{Var}[S_n] = n \cdot \text{Var}[X_i] = npq \quad \hookrightarrow \text{why?}$$

Variance: $\text{Var}[X] = \mathbb{E}[(X-\mu)^2]$, where $\mu = \mathbb{E}[X]$ $\leftarrow$
(measure of how much the R.V. differs from it's mean)

Prove it! $\rightarrow \text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ if X & Y independent

$$\text{Var}[X + \text{constant}] = \text{Var}[X]$$

$$\text{Var}[cX] = c^2 \text{Var}[X]$$

Standard deviation $= \sigma := \sqrt{\text{Var}[X]}$ or $\text{Var}[X] = \sigma^2$

Some "standardizing" you do to a random variable.

standardizing

① Centering: $S_n \rightarrow$ has mean np .
How to get mean 0? Take $(S_n - \mu) = (S_n - np)$

does not change variance!

② Get variance 1: $\frac{(S_n - \mu)}{\sigma} = \frac{(S_n - np)}{\sqrt{npq}}$

Enough to study $Z_n = \frac{(S_n - \mu)}{\sigma}$ instead of S_n .

Come back to Galton board:

$p = q = 1/2$ $S_n = X_1 + X_2 + \dots + X_n$
prob of going left or right. $\uparrow$ index of bucket! $\uparrow$ throws in the Galton board

$$\mu = E[S_n] = \frac{n}{2} \quad \text{Var}[S_n] = \frac{n}{4} \Rightarrow \sigma = \frac{\sqrt{n}}{2}$$

$$Z_n = \frac{S_n - n/2}{\sqrt{n}/2} = \frac{1}{\sqrt{n}} ((2X_1 - 1) + (2X_2 - 1) + \dots + (2X_n - 1))$$

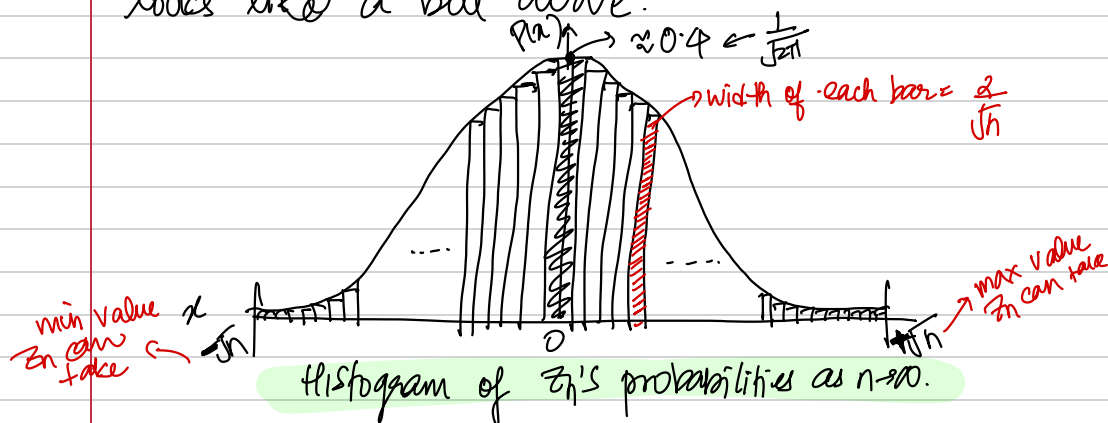
$\downarrow \quad \downarrow$
independent ± 1 with prob $1/2$ each
"Rademachers"

Now, for n even

$$\Pr[S_n = \frac{n}{2}] = \Pr[Z_n = 0] \sim \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right)$$

(Recall from last time, using Stirling)

Now, plotting Z_n for fixed $0 < p < 1$, as $n \rightarrow \infty$, it looks like a bell curve.



This bell-curve captures a Gaussian prob. distribution!

GAUSSIAN DISTRIBUTION

$X \sim N(\mu, \sigma^2)$: X is a Gaussian random variable has probability density function (pdf):
 means the r.v. Z has the distribution

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$Z \sim N(0, 1)$ is called standard Gaussian or Normal random variable.

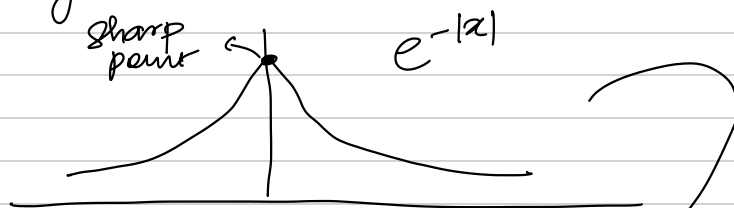
To understand this, why this pdf, let's look at
 (you can refer to: [3Blue1Brown](https://youtu.be/zeJD6dqJ5lo) <https://youtu.be/zeJD6dqJ5lo>)

Let's look at

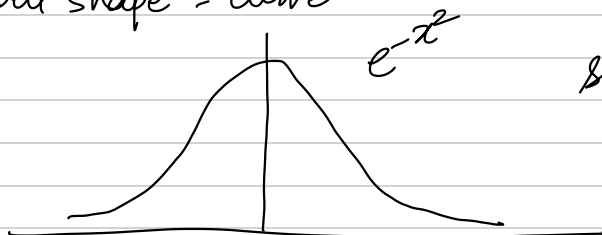


e^{-x} (decays exponentially)

To get decay in both directions



More "bell shape" - curve



(Why specifically this bell curve & not another function also has a great reasoning
 ↳ see Blue/Brown on Herschel-Maxwell derivation for Gaussian

(<https://youtu.be/cy8r7WSuT1I>)

Having constant c & taking e^{-cx^2} allows you to vary the width of these bell curves.



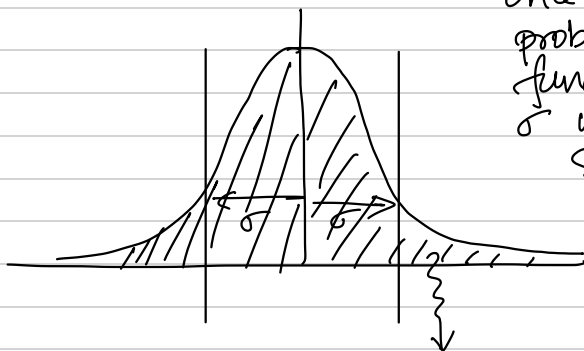
But $e^{-cx^2} =$

But, we use e , as we will see now, for a particular meaningful reason!

$(e^c)^{-x^2}$

↳ varying c varies this constant!
 Nothing special about e could have been another number instead like π^c → still gives bell shape curve!

$$e^{-1/2(x/\sigma)^2}$$



Once we get to a probability density function from this, σ will capture the standard deviation.

But to interpret this as a probability density function, we need area under the curve = 1.

Just taking e^{-x^2} , $\int_{-\infty}^{\infty} e^{-x^2} = \sqrt{\pi}$

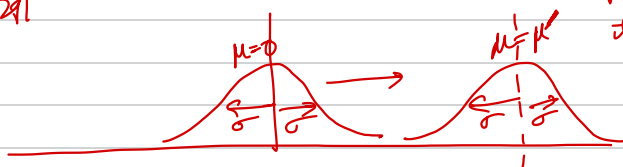
I will show you this
But, if this is true then we take $\frac{1}{\sqrt{\pi}} e^{-x^2}$ to get area 1!

Now, take normalization factor for $e^{-1/2(x/\sigma)^2}$ to get

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2(x/\sigma)^2} \rightarrow \text{valid probability density function.}$$

Captures width of the bell curve = standard deviation.

Now, to be able shift the curve (consider $\mu \neq 0$) take $\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2(x-\mu/\sigma)^2}$



helps slide this bell curve left & right!

This gives you the Gaussian distribution!

To Prove: $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$

Proof:

Can't directly integrate $\int_{-\infty}^{\infty} e^{-x^2} dx$. $\rightarrow$ no anti derivative!

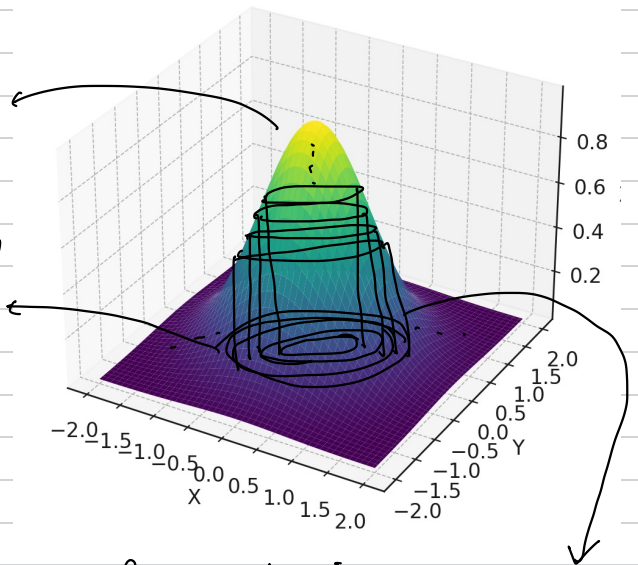
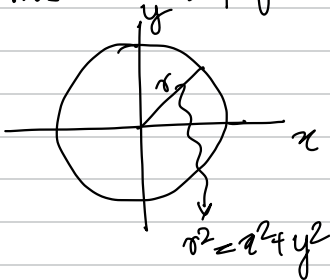
But take instead

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy$$

$$z = e^{-(x^2+y^2)}$$

want the volume under this!

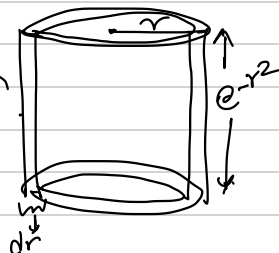
this circular projection



Now the volume can be figured by integrating thin cylinders like these (underneath the surface)

just take small thickness (dr), which makes volume $= 2\pi r \cdot e^{-r^2} \cdot dr$

Area = circumference $\times$ Height
 $= 2\pi r \cdot e^{-r^2}$



Adding all gives

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy = \int_0^{\infty} 2\pi r \cdot e^{-r^2} dr = \pi \int_0^{\infty} 2r \cdot e^{-r^2} dr$$

$$= \pi \rightarrow \textcircled{1}$$

But we want to evaluate $\int_{-\infty}^{\infty} e^{-x^2} dx \rightarrow$ denote by C

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx dy = \int_{-\infty}^{\infty} C e^{-y^2} dy = C \cdot C = C^2 \rightarrow \textcircled{2}$$

$$\textcircled{1} \& \textcircled{2} \Rightarrow C = \sqrt{\pi} \Rightarrow \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Properties of Gaussian

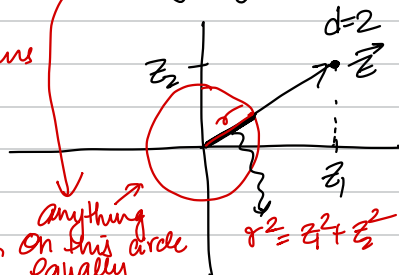
$\textcircled{1}$ Let $\vec{Z} = (Z_1, \dots, Z_d) \in \mathbb{R}^d$, where $Z_1, \dots, Z_d \sim N(0, 1)$ (iid)

Then $\vec{Z}$'s distribution is 'rotationally symmetric'

pdf of $\vec{Z}$ at $\vec{z} = (z_1, \dots, z_d)$

$$= \phi(z_1) \phi(z_2) \dots \phi(z_d)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-\underbrace{(z_1^2 + z_2^2 + \dots + z_d^2)}_{\|\vec{z}\|^2}} / 2$$



what this means is that the pdf of $\vec{Z}$ at $\vec{z}$ is only a function of its length!

(This is used in the proof of the integral $\int_{-\infty}^{\infty} e^{-x^2} dx$ above!)

II "Sum of independent Gaussians is Gaussian"

$Y_1 \sim N(\mu_1, \sigma_1^2)$ & $Y_2 \sim N(\mu_2, \sigma_2^2)$ be independent
then $Y = a_1 Y_1 + a_2 Y_2 \sim N(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$

Exercise → prove this. (Hint: use property (I))

CENTRAL LIMIT THEOREM → see Ryan's book
(Berry-Essen Theorem) in references on Pg 1.
(Ch 5)

Let $X_1, \dots, X_n$ be independent, with $\mathbb{E}[X_i] = 0 \neq 1$,
and $\sum_{i=1}^n \sigma_i^2 = 1$, where $\sigma_i^2 = \text{Var}[X_i]$

Let $S = X_1 + X_2 + \dots + X_n$ ($\therefore \mathbb{E}[S] = 0, \text{Var}[S] = 1$)
and let $Z \sim N(0, 1)$ be a normal R.V.

Then

$\forall u \in \mathbb{R}, |\Pr[S \leq u] - \Pr[Z \leq u]| \leq C\gamma$, where

$\gamma = \sum_{i=1}^n \mathbb{E}[|X_i|^3]$ & C is a universal constant
($C \approx 0.56 \rightarrow$ Shevtsova, 2010)

Is this small? → in "most" cases, yes!

Remark: If all the X_i 's satisfy: $|X_i| \leq \epsilon$ with prob 1
then $\gamma = \sum_{i=1}^n \mathbb{E}[|X_i|^3] \leq \epsilon \cdot \sum_{i=1}^n \mathbb{E}[|X_i|^2] = \epsilon \cdot \sum_{i=1}^n \sigma_i^2 = \epsilon$
i.e. γ is small!

Example: fair coin flips: $X_i = \begin{cases} +1/\sqrt{n} & \text{with prob. } 1/2 \\ -1/\sqrt{n} & \text{with prob. } 1/2 \end{cases}$

Then $\sigma_f^2 = E[X^2] = \frac{1}{n}$ & $\mu = 0$

$E[|X|^3] = \frac{1}{n^{3/2}} \Rightarrow \gamma = \frac{1}{\sqrt{n}}$

$\therefore |\Pr[S \leq u] - \Pr[Z \leq u]| \leq \frac{0.56}{\sqrt{n}} \rightarrow$ gives a more precise estimate than CLT

Example 2: Flip n fair coins, count #Heads!

Then

(Why?) $S_n = \frac{\#Heads - \#Tails}{\sqrt{n}}$

check if you are unsure!

$= \frac{H - (n - H)}{\sqrt{n}}, H = \#Heads$

(interpret heads as success $\rightarrow +1$ & tails as failure $\rightarrow -1$)

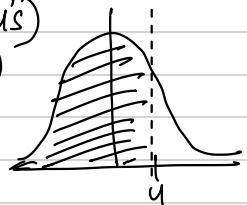
$= \frac{2H}{\sqrt{n}} - \sqrt{n}$

$S_n \leq u \Leftrightarrow H \leq \frac{n}{2} + u \frac{\sqrt{n}}{2}$

$\therefore \Pr[H \leq \frac{n}{2} + u \frac{\sqrt{n}}{2}] \approx \Pr[Z \leq u] = \int_{-\infty}^u \phi(z) dz$

(Can find & estimate for specific u 's)
What is γ here?

($Z \sim N(0, 1)$)



Assignment question $\rightarrow$ how to modify Berry Esseen for $\mu \neq 0$ & $\sigma^2 \neq 1$