10: The Normal (Gaussian) Distribution

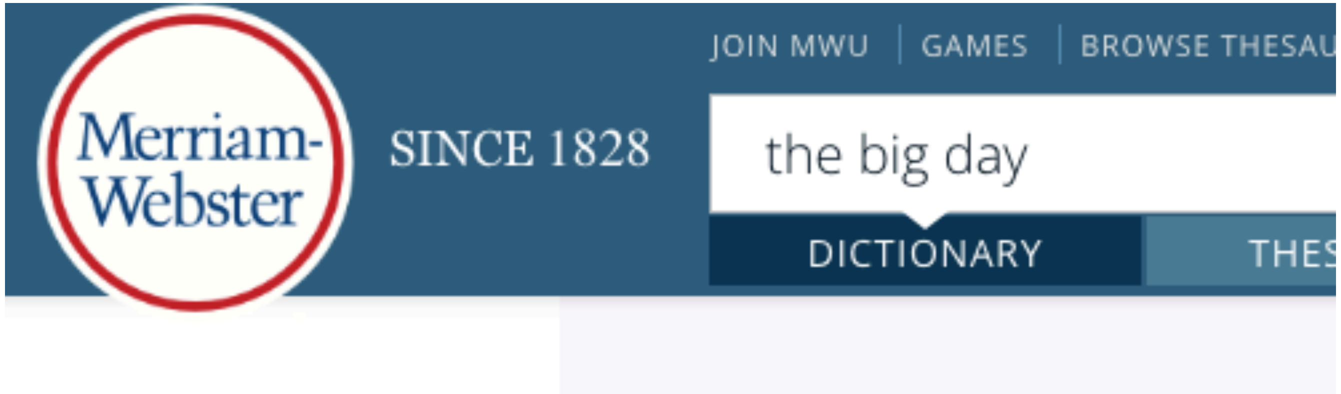
Lisa Yan

April 27, 2020

Quick slide reference

Normal RV

Today's the Big Day



the big day noun phrase

Definition of *the big day*

: the day that something important happens

// Today *is the big day*.

also : the day someone is to be married

// So, when's *the big day*?

Normal Random Variable

def An **Normal** random variable X is defined as follows:

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

Support: $(-\infty, \infty)$

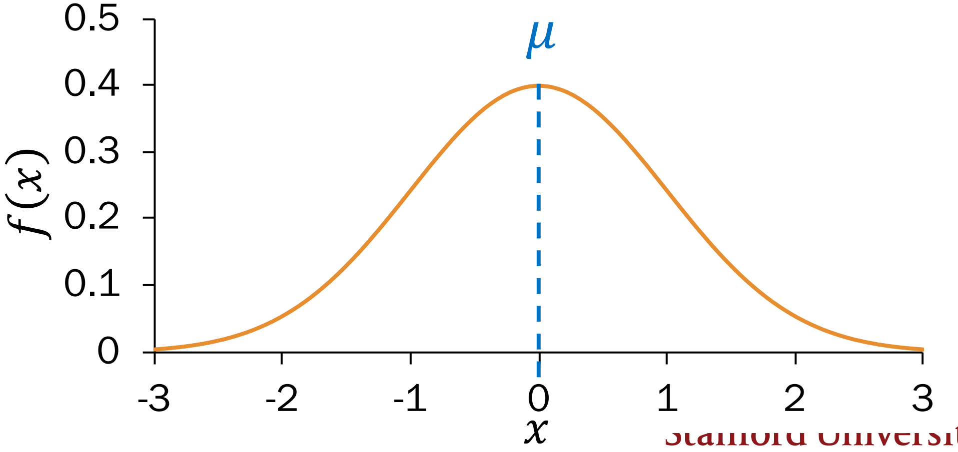
PDF	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
Expectation	$E[X] = \mu$
Variance	$\text{Var}(X) = \sigma^2$

Other names: **Gaussian** random variable

mean

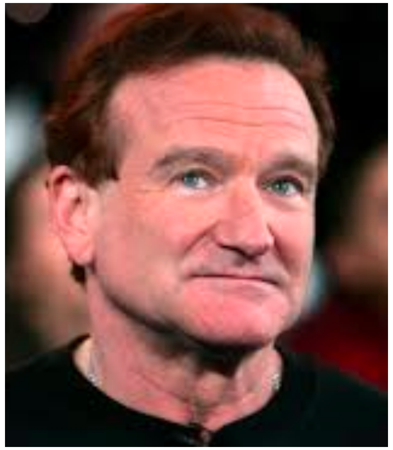
variance

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



Carl Friedrich Gauss

Carl Friedrich Gauss (1777-1855) was a remarkably influential German mathematician.



Johann Carl Friedrich Gauss (/ɡaʊs/; German: *Gauß* [ɡaʊs] (listen); Latin: *Carolus Fridericus Gauss*; 30 April 1777 – 23 February 1855) was a German mathematician and physicist who made significant contributions to many fields, including algebra, analysis, astronomy, differential geometry, electrostatics, geodesy, geophysics, magnetic fields, matrix theory, mechanics, number theory, optics and statistics.

Sometimes referred to as the *Princeps mathematicorum*[1] (Latin for "the foremost of mathematicians") and "the greatest mathematician since antiquity", Gauss had an exceptional influence in many fields of mathematics and science, and is ranked among history's most influential mathematicians.[2]

Did not invent Normal distribution but rather popularized it

Why the Normal?

- Common for natural phenomena: height, weight, etc.
- Most noise in the world is Normal
- Often results from the sum of many random variables
- Sample means are distributed normally

That's what they want you to believe…

Why the Normal?

- Common for natural phenomena: height, weight, etc.

  Actually log-normal

- Most noise in the world is Normal

  Just an assumption

- Often results from the sum of many random variables

  Only if equally weighted

- Sample means are distributed normally

  (okay this one is true, we'll see this in 3 weeks)

Okay, so why the Normal?

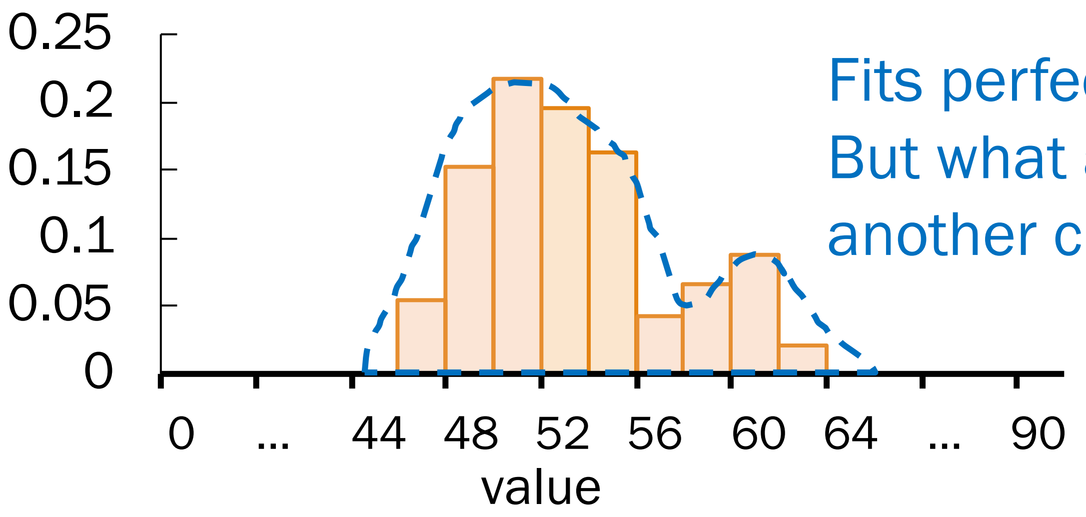
Part of CS109 learning goals:

- Translate a problem statement into a random variable

In other words: model real life situations with probability distributions

How do you model student heights?

- Suppose you have data from one classroom.



Fits perfectly!
But what about in another classroom?

Okay, so why the Normal?

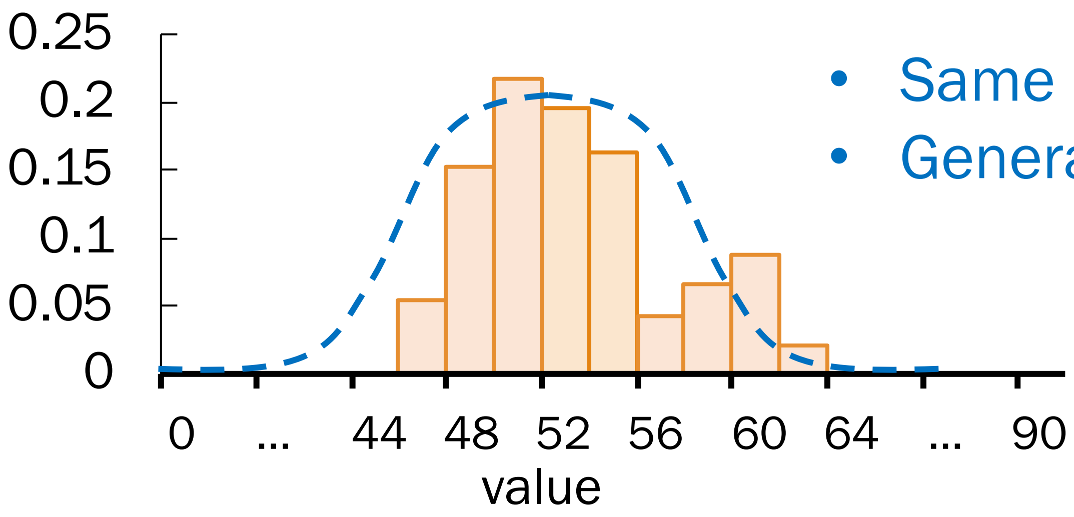
Part of CS109 learning goals:

- Translate a problem statement into a random variable

In other words: model real life situations with probability distributions

How do you model student heights?

- Suppose you have data from one classroom.



- Same mean/var
- Generalizes well

Occam's Razor:

"Non sunt multiplicanda entia sine necessitate."

Entities should not be multiplied without necessity.

A Gaussian maximizes entropy for a given mean and variance.

Why the Normal?

- Common for natural phenomena: height, weight, etc.

  Actually log-nor l

- Most noise in the world is Normal

  mption

- Often results from random va

  Only if equally weighted

- Sample means are distributed normally

  (okay this one is true, we'll see this in 3 weeks)

because it's easy to use

I encourage you to stay critical of how to model real-world phenomena.

Anatomy of a beautiful equation

Let $X \sim \mathcal{N}(\mu, \sigma^2)$.

The PDF of X is defined as:

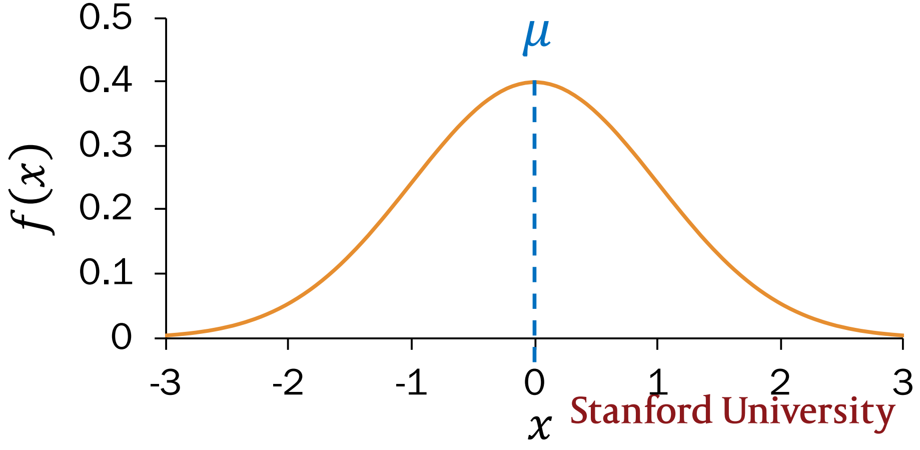
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

symmetric around μ

variance σ^2 manages spread

exponential tail

normalizing constant



Campus bikes

You spend some minutes, X, traveling between classes.

- Average time spent: $\mu = 4$ minutes
- Variance of time spent: $\sigma^2 = 2$ minutes^2

Suppose X is normally distributed. What is the probability you spend ≥ 6 minutes traveling?

$$X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2)$$

$$P(X \geq 6) = \int_6^\infty f(x)dx = \int_6^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

(call me if you analytically solve this)

Loving, not scary

...except this time

Computing probabilities with Normal RVs

For a Normal RV $X \sim \mathcal{N}(\mu, \sigma^2)$, its CDF has no closed form.

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

Cannot be solved analytically

However, we can solve for probabilities numerically using a function Φ:

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

CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$

A function that has been solved for numerically

To get here, we'll first need to know some properties of Normal RVs.

Normal RV: Properties

Properties of Normal RVs

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF $P(X \leq x) = F(x)$.

1. Linear transformations of Normal RVs are also Normal RVs.

   If $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

2. The PDF of a Normal RV is symmetric about the mean μ.

$$F(\mu - x) = 1 - F(\mu + x)$$

1. Linear transformations of Normal RVs

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF $P(X \leq x) = F(x)$.

Linear transformations of X are also Normal.

If $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Proof:

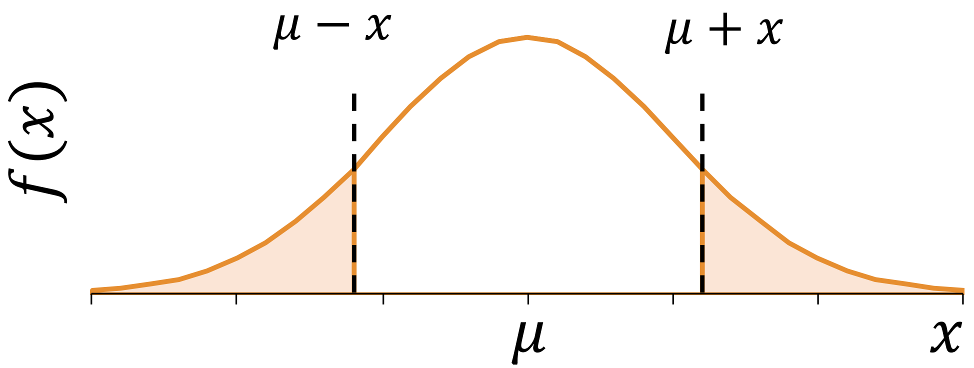
- $E[Y] = E[aX + b] = aE[X] + b = a\mu + b$ — Linearity of Expectation
- $\text{Var}(Y) = \text{Var}(aX + b) = a^2\text{Var}(X) = a^2\sigma^2$ — $\text{Var}(aX + b) = a^2\text{Var}(X)$
- Y is also Normal — Proof in Ross, 10th ed (Section 5.4)

2. Symmetry of Normal RVs

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF $P(X \le x) = F(x)$.

The PDF of a Normal RV is symmetric about the mean μ.

$$F(\mu - x) = 1 - F(\mu + x)$$



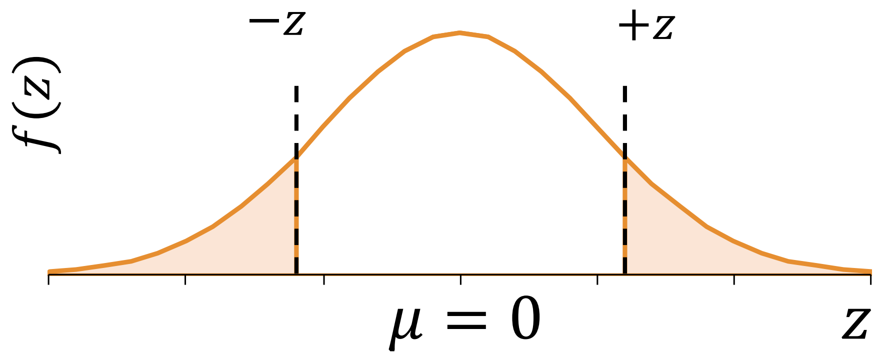
Using symmetry of the Normal RV

$F(\mu - x) = 1 - F(\mu + x)$

Let $Z \sim \mathcal{N}(0,1)$ with CDF $P(Z \leq z) = F(z)$.

Suppose we only knew numeric values for $F(z)$ and $F(y)$, for some $z, y \geq 0$.

How do we compute the following probabilities?



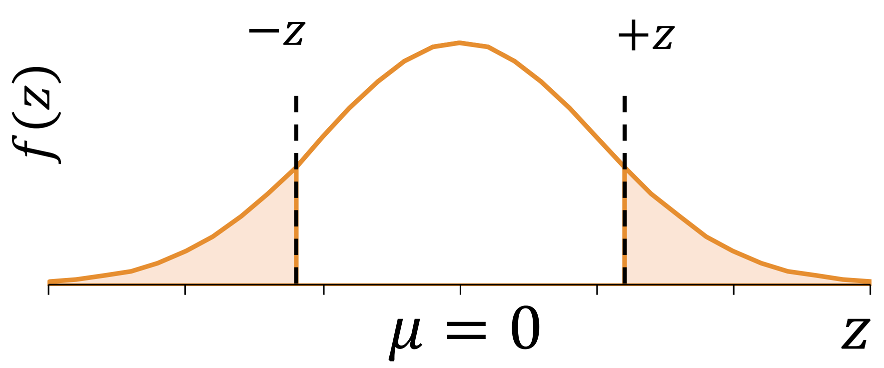
1. $P(Z \leq z)$ $\quad = F(z)$
2. $P(Z < z)$
3. $P(Z \geq z)$
4. $P(Z \leq -z)$
5. $P(Z \geq -z)$
6. $P(y < Z < z)$

A. $F(z)$
B. $1 - F(z)$
C. $F(z) - F(y)$

Using symmetry of the Normal RV

$F(\mu - x) = 1 - F(\mu + x)$

Let $Z \sim \mathcal{N}(0,1)$ with CDF $P(Z \le z) = F(z)$.



Suppose we only knew numeric values for $F(z)$ and $F(y)$, for some $z, y \ge 0$.

How do we compute the following probabilities?

1. $P(Z \le z) \quad = F(z)$
2. $P(Z < z) \quad = F(z)$
3. $P(Z \ge z) \quad = 1 - F(z)$
4. $P(Z \le -z) \quad = 1 - F(z)$
5. $P(Z \ge -z) \quad = F(z)$
6. $P(y < Z < z) \quad = F(z) - F(y)$

A. $F(z)$

B. $1 - F(z)$

C. $F(z) - F(y)$

Symmetry is particularly useful when computing probabilities of zero-mean Normal RVs.

Normal RV: Computing probability

Computing probabilities with Normal RVs

Let $X \sim \mathcal{N}(\mu, \sigma^2)$.

To compute the CDF, $P(X \le x) = F(x)$:

- We cannot analytically solve the integral (it has no closed form)
- ...but we *can* solve numerically using a function Φ:

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

CDF of the
Standard Normal, Z

Standard Normal RV, Z

The Standard Normal random variable Z is defined as follows:

$$Z \sim \mathcal{N}(0, 1)$$

Expectation $E[Z] = \mu = 0$ "zero mean"

Variance $\text{Var}(Z) = \sigma^2 = 1$ "unit variance"

Note: not a new distribution; just a special case of the Normal

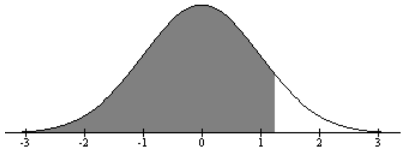
Other names: Unit Normal

CDF of Z defined as: $P(Z \leq z) = \Phi(z)$

Φ has been numerically computed

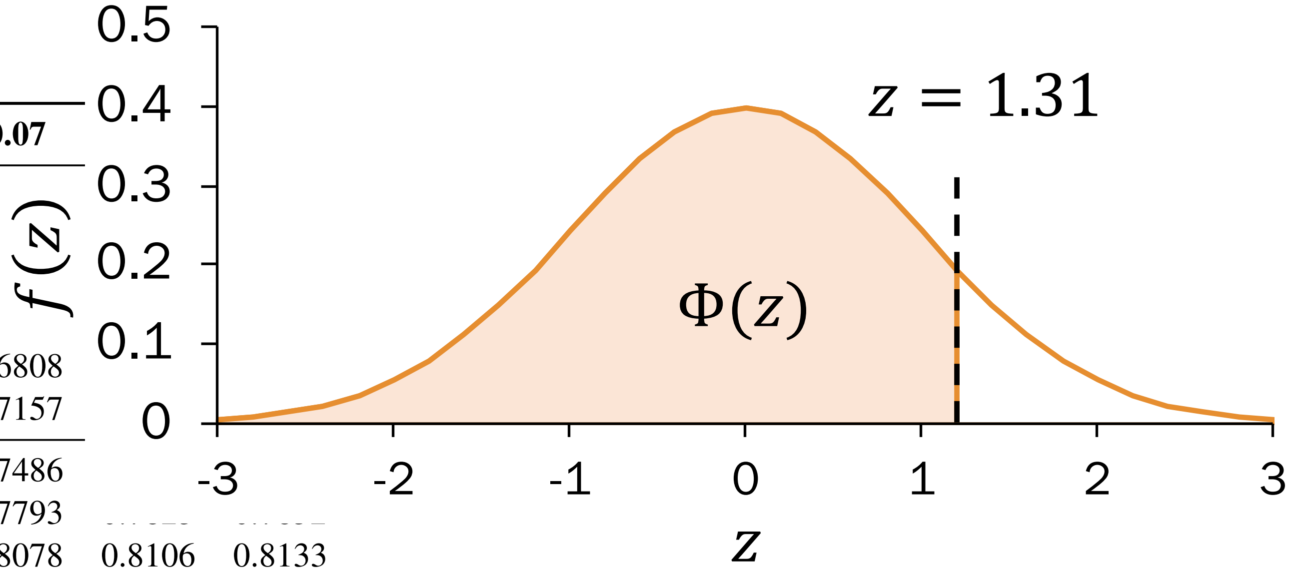
Standard Normal Table

An entry in the table is the area under the curve to the left of z, $P(Z \le z) = \Phi(z)$.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07		
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0		
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0		
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0		
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0		
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808		
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157		
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486		
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793		
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

$$P(Z \le 1.31) = \Phi(1.31)$$



Standard Normal Table only has probabilities $\Phi(z)$ for $z \ge 0$.

History fact: Standard Normal Table

TABLES

SERVANT

AU CALCUL DES RÉFRACTIONS

APPROCHANTES DE L'HORIZON.

TABLE PREMIÈRE.

Intégrales de $e^{-tt}\,dt$, depuis une valeur quelconque de t jusqu'à t infinie.

$\int_a^\infty e^{-t^2}\,dt$

t	Intégrale.	Diff. prem.	Diff. II.	Diff. III.
0,00	0,88622692	999968	201	199
0,01	0,87622724	999767	400	199
0,02	0,86622057	999367	599	200
0,03	$\boxed{0,85623590}$	998768	799	199
0,04	0,84624822	997969	998	197
0,05	0,83626853	996971	1195	199
0,06	0,82629882	995776	1394	196

The first Standard Normal Table was computed by Christian Kramp, French astronomer (1760–1826), in *Analyse des Réfractions Astronomiques et Terrestres*, 1799

Used a Taylor series expansion to the third power

integral from x = 0.03 to infinity of e^{-x^2}

Extended Keyboard   Upload

Definite integral:

$$\int_{0.03}^{\infty} e^{-x^2}\,dx = 0.856236$$

Probabilities for a general Normal RV

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. To compute the CDF $P(X \le x) = F(x)$, we use Φ, the CDF for the Standard Normal $Z \sim \mathcal{N}(0, 1)$:

$$\boxed{F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)}$$

Proof:

$$\begin{aligned} F(x) &= P(X \le x) && \text{Definition of CDF} \\ &= P(X - \mu \le x - \mu) = P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right) && \text{Algebra} + \sigma > 0 \\ &= P\left(Z \le \frac{x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

- $\frac{X-\mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma}$ is a linear transform of X.
- This is distributed as $\mathcal{N}\left(\frac{1}{\sigma}\mu - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2\right) = \mathcal{N}(0,1)$
- In other words, $\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)$ with CDF Φ.

Probabilities for a general Normal RV

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. To compute the CDF $P(X \le x) = F(x)$, we use Φ, the CDF for the Standard Normal $Z \sim \mathcal{N}(0, 1)$:

$$\boxed{F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

Proof:

$$F(x) = P(X \le x) \qquad \text{Definition of CDF}$$

$$= P(X - \mu \le x - \mu) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) \qquad \text{Algebra} + \sigma > 0$$

$$= P\left(Z \le \frac{x-\mu}{\sigma}\right)$$

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

- $\frac{X-\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$ is a linear transform of X.

1. Compute $z = (x - \mu)/\sigma$.
2. Look up $\Phi(z)$ in Standard Normal table.

Campus bikes

You spend some minutes, X, traveling between classes.

- Average time spent: $\mu = 4$ minutes
- Variance of time spent: $\sigma^2 = 2$ minutes^2

Suppose X is normally distributed. What is the probability you spend ≥ 6 minutes traveling?

$$X \sim \mathcal{N}(\mu = 4, \sigma^2 = 2)$$

$\sigma = \sqrt{2}$

✗ $P(X \geq 6) = \int_6^\infty f(x)dx$ (no analytic solution)

1. Compute $z = \frac{(x-\mu)}{\sigma}$

Φ x

$$P(X \geq 6) = 1 - F_X(6)$$
$$= 1 - \Phi\left(\frac{6-4}{\sqrt{2}}\right)$$
$$\approx 1 - \Phi(1.41)$$

2. Look up $\Phi(z)$ in table

$$1 - \Phi(1.41)$$
$$\approx 1 - 0.9207$$
$$= 0.0793$$

Is there an easier way? (yes)

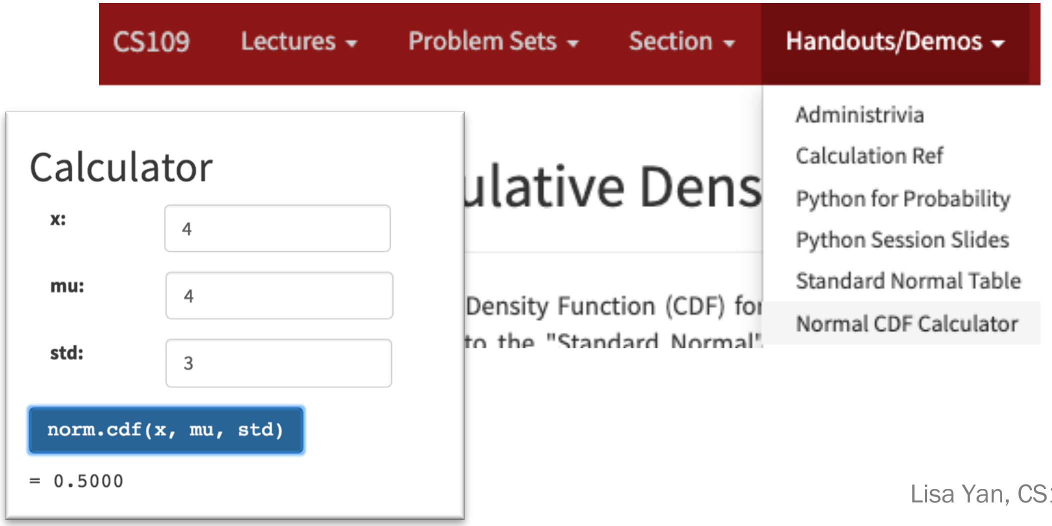
Let $X \sim \mathcal{N}(\mu, \sigma^2)$. What is $P(X \leq x) = F(x)$?

- Use Python

```
from scipy import stats
X = stats.norm(mu, std)
X.cdf(x)
```

SciPy reference:
https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html

- Use website tool



Website tool:
https://web.stanford.edu/class/cs109/handouts/normalCDF.html

(live) 10: The Normal (Gaussian) Distribution

Lisa Yan

April 27, 2020

The Normal (Gaussian) Random Variable

Let $X \sim \mathcal{N}(\mu, \sigma^2)$.

The PDF of X is defined as:

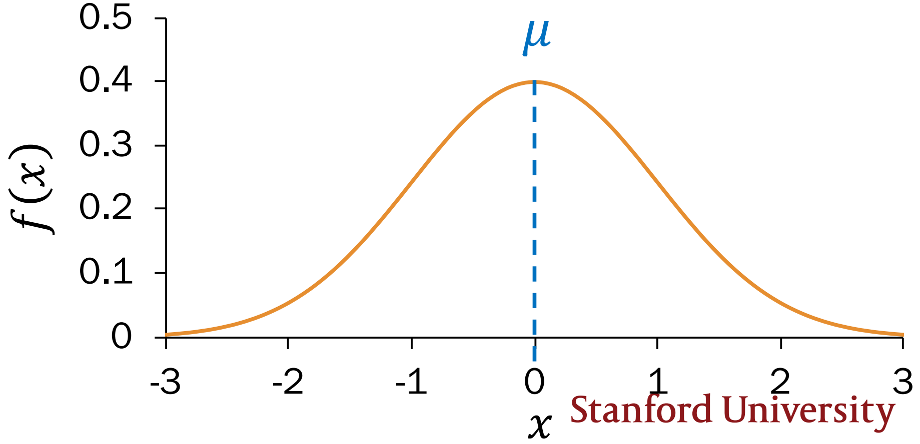
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

symmetric around μ

variance σ^2 manages spread

normalizing constant

exponential tail



Think

Slide 34 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/46499

Think by yourself: 2 min

(by yourself)

Normal Random Variable

mean variance

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

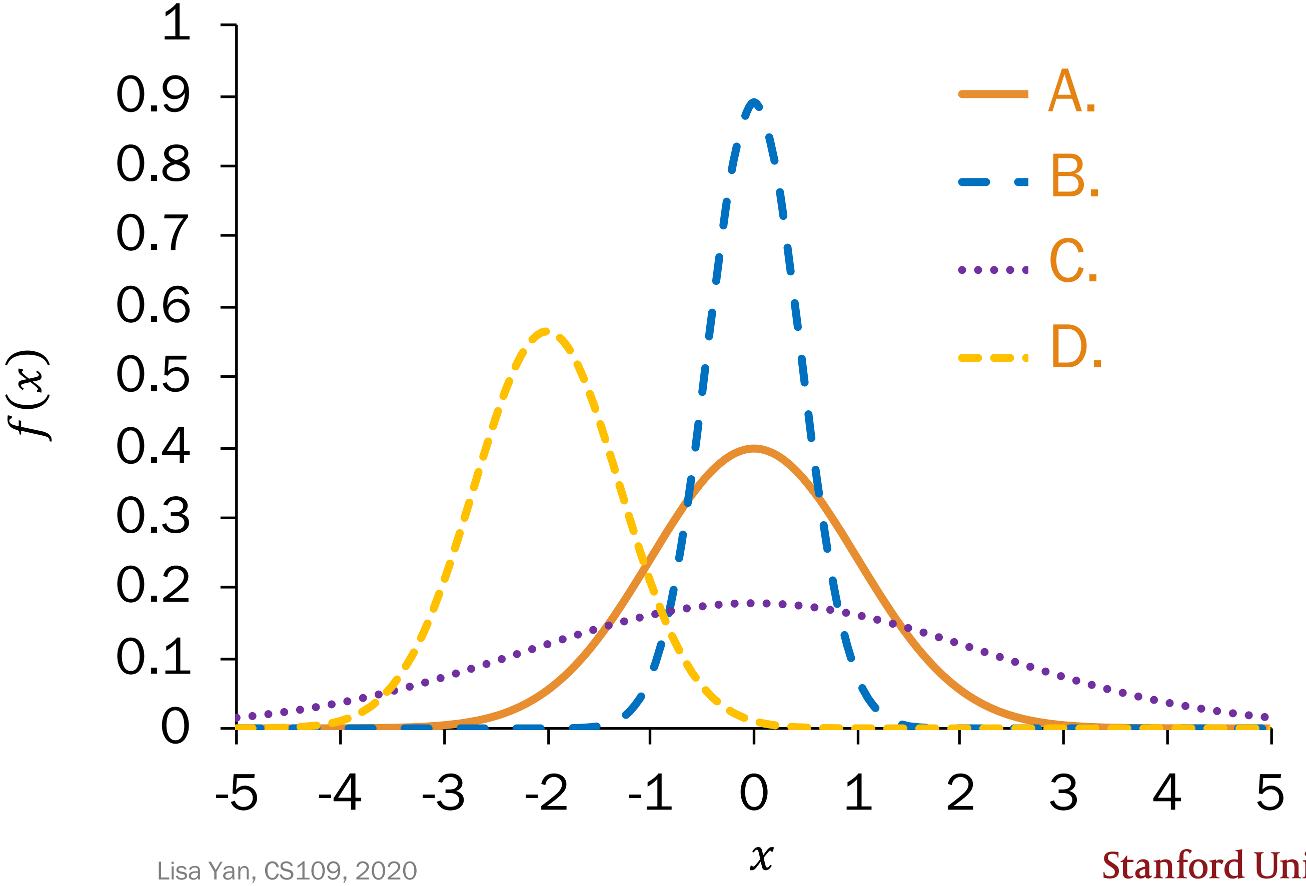
Match PDF to distribution:

$\mathcal{N}(0, 1)$

$\mathcal{N}(-2, 0.5)$

$\mathcal{N}(0, 5)$

$\mathcal{N}(0, 0.2)$



(by yourself)

Normal Random Variable

mean variance

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

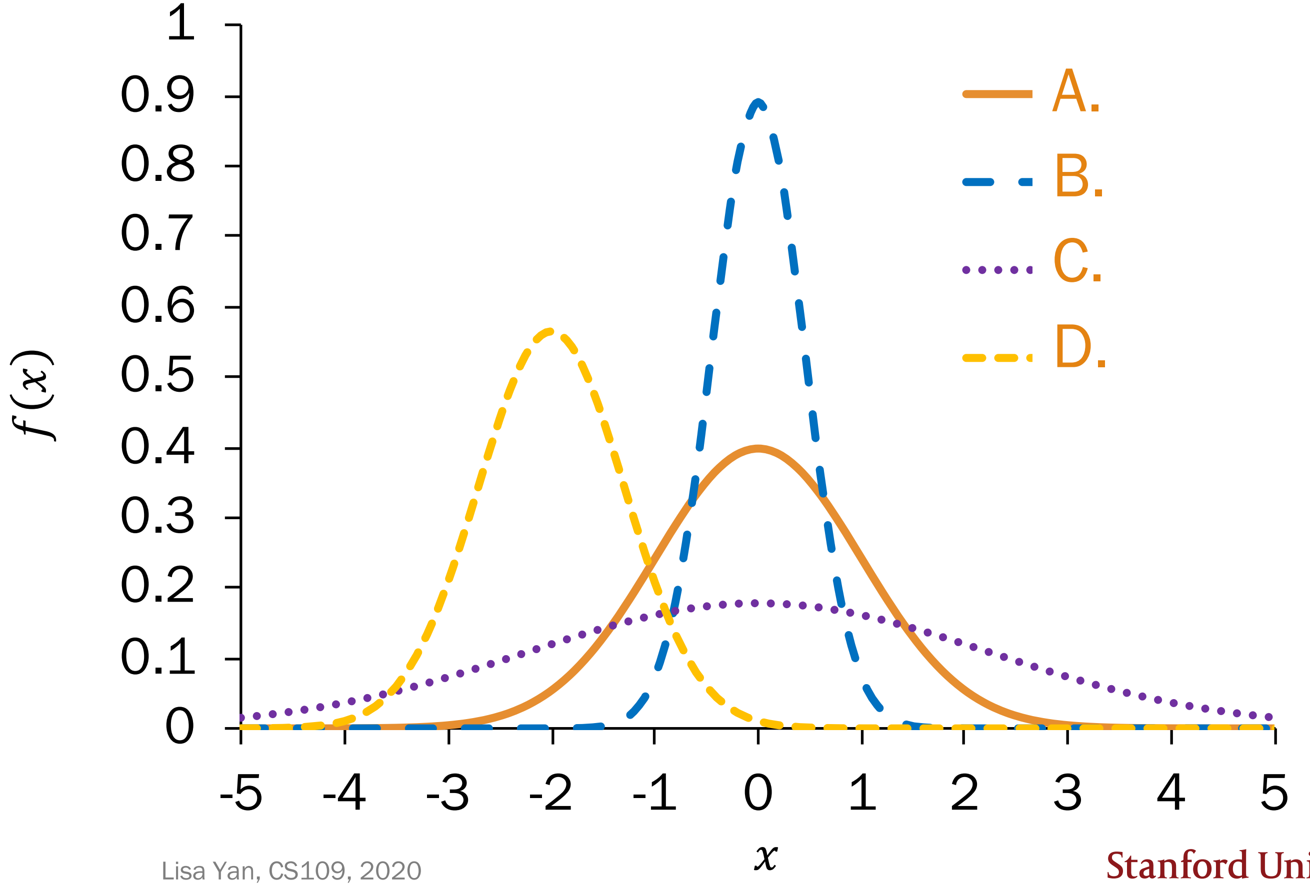
Match PDF to distribution:

$\mathcal{N}(0, 1)$

$\mathcal{N}(-2, 0.5)$

$\mathcal{N}(0, 5)$

$\mathcal{N}(0, 0.2)$



Computing probabilities with Normal RVs: Old school

Standard Normal Table *

Note: An entry in the table is the area under the curve to the left of z, $P(Z \le z) = \Phi(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.5714	0.5753
0.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6103	0.6141
0.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

$\Phi(z)$ for non-negative z

*particularly useful when we have closed book exams with no calculator**

**we have open book exams with calculators this quarter

Knowing how to use a Standard Normal Table will still be useful in our understanding of Normal RVs.

Computing probabilities with Normal RVs

Review

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. What is $P(X \leq x) = F(x)$?

1. Rewrite in terms of standard normal CDF Φ by computing $z = \frac{(x-\mu)}{\sigma}$.

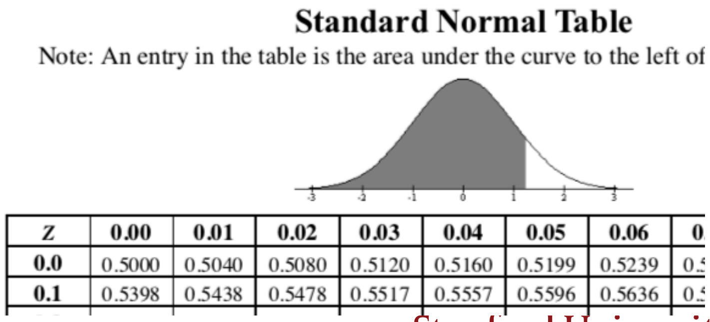
   Linear transforms of Normals are Normal:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad Z = \frac{(X-\mu)}{\sigma}, \text{ where } Z \sim \mathcal{N}(0,1)$$

2. Then, look up in a Standard Normal Table, where $z \geq 0$.

   Normal PDFs are symmetric about their mean:

$$\Phi(-z) = 1 - \Phi(z)$$


Standard Normal Table

Note: An entry in the table is the area under the curve to the left of

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-z) = 1 - \Phi(z)$

Breakout Rooms

Slide 39 has two questions to go over in groups.

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/46499

Breakout rooms: 5 mins

🤔

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$.
Note standard deviation $\sigma = 4$.

How would you write each of the below probabilities as a function of the standard normal CDF, Φ?

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-z) = 1 - \Phi(z)$

1. $P(X > 0)$ (we just did this)
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-z) = 1 - \Phi(z)$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-x) = 1 - \Phi(x)$

Compute $z = \frac{(x-\mu)}{\sigma}$

Look up $\Phi(z)$ in table

$$P(X < -3) + P(X > 9)$$

$$= F(-3) + \big(1 - F(9)\big)$$

$$= \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-x) = 1 - \Phi(x)$

Compute $z = \frac{(x-\mu)}{\sigma}$

$$P(X < -3) + P(X > 9)$$

$$= F(-3) + \left(1 - F(9)\right)$$

$$= \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$

Look up $\Phi(z)$ in table

$$= \Phi\left(-\frac{3}{2}\right) + \left(1 - \Phi\left(\frac{3}{2}\right)\right)$$

$$= 2\left(1 - \Phi\left(\frac{3}{2}\right)\right)$$

$$\approx 0.1337$$

Interlude for jokes/announcements

Announcements

Quiz #1

Time frame: Thursday 4/30 12:00am-11:59pm PT
Covers: Up to end of Week 3 (including Lecture 9)
Review session (Tim): Tuesday 4/28 12-2pm PT
https://stanford.zoom.us/j/92275547392
Info and practice: https://web.stanford.edu/class/cs109/exams/quizzes.html

Section this week

Optional (not graded)

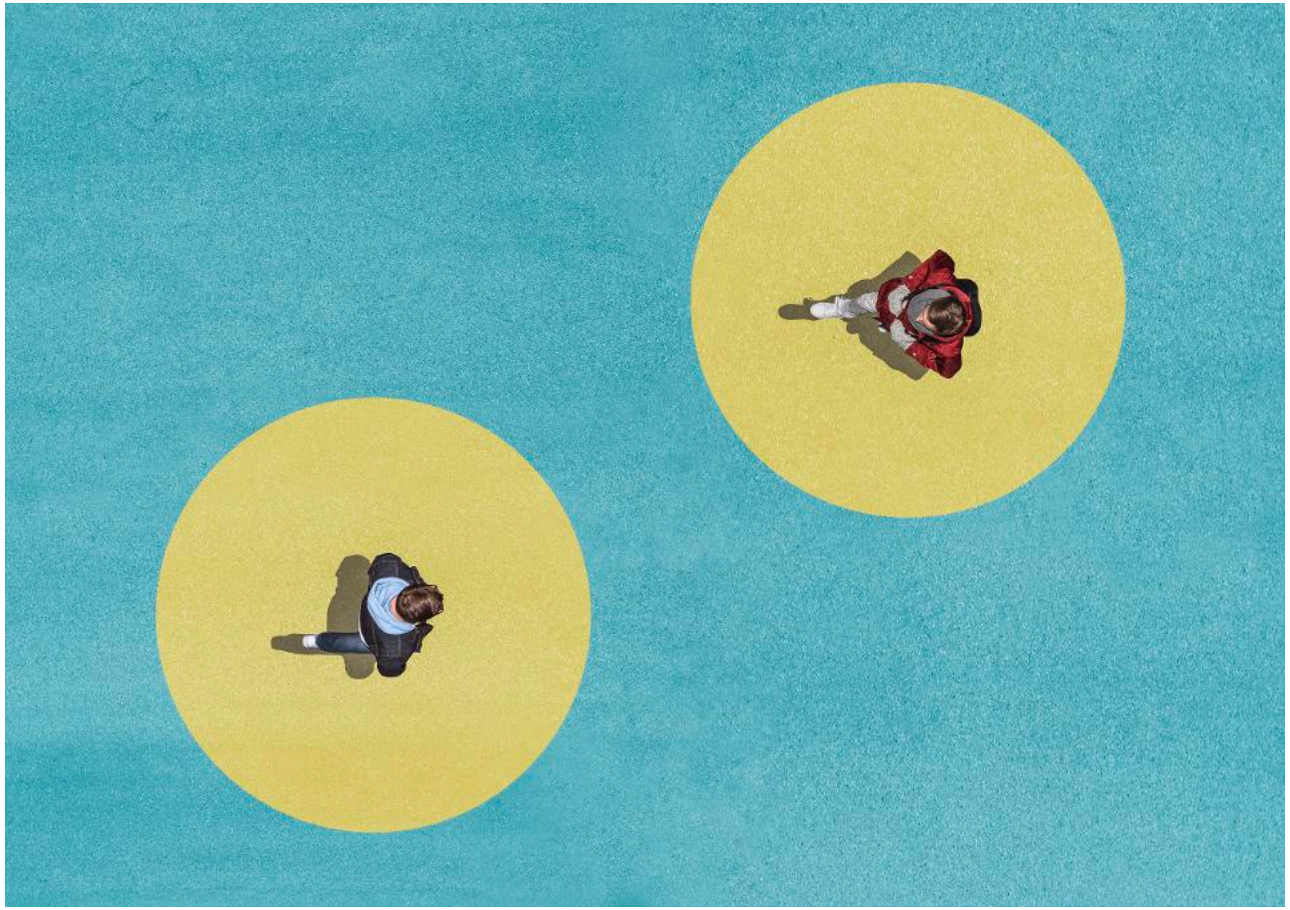
You can “makeup” and attend any section

Friday’s concept check (#12)

Extra credit

Interesting probability news

On The Probabilities Of Social Distancing As Gleaned From AI Self-Driving Cars



Submitted 10 times over PS1/PS2!

https://www.forbes.com/sites/lanceeliot/2020/04/12/on-the-probabilities-of-social-distancing-as-gleaned-from-ai-self-driving-cars/#218da4489472

Breakout Rooms

Slide 47 has two questions to go over in groups.

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/46499

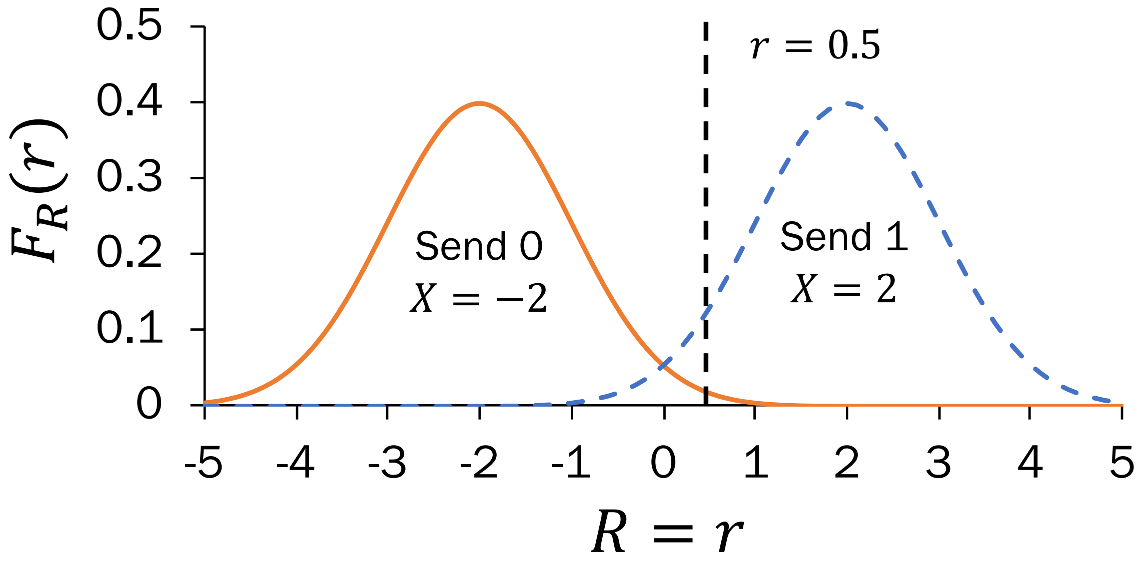
Breakout rooms: 5 mins

Noisy Wires

Send a voltage of 2 V or -2 V on wire (to denote 1 and 0, respectively).

- X = voltage sent (2 or -2)
- Y = noise, $Y \sim \mathcal{N}(0, 1)$
- $R = X + Y$ voltage received.

Decode: $\begin{cases} 1 & \text{if } R \geq 0.5 \\ 0 & \text{otherwise.} \end{cases}$



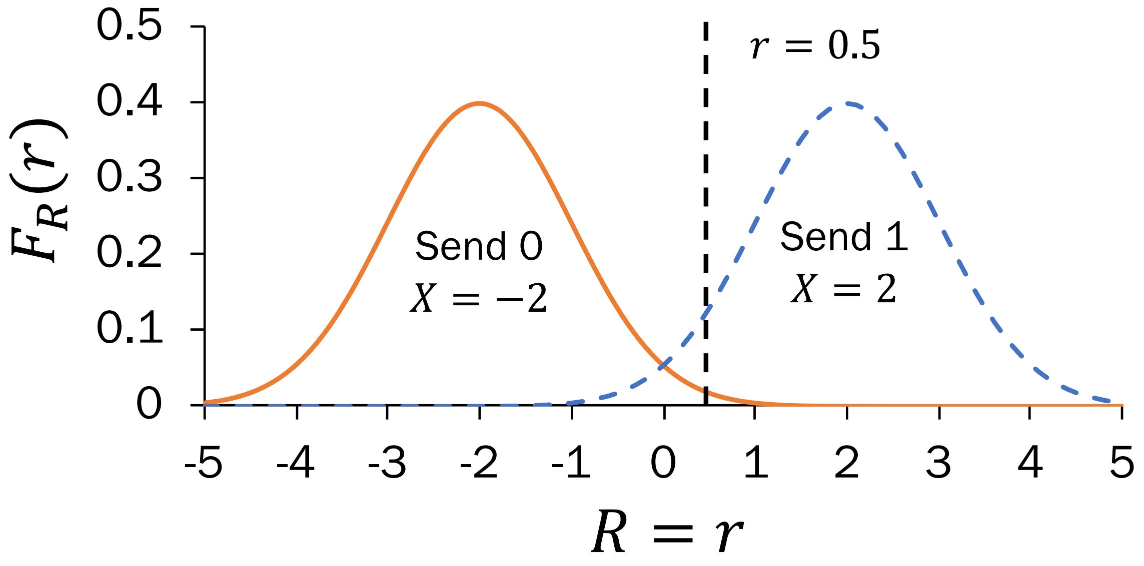
1. What is P(decoding error | original bit is 1)?
   i.e., we sent 1, but we decoded as 0?
2. What is P(decoding error | original bit is 0)?

These probabilities are unequal. Why might this be useful?

Noisy Wires

Send a voltage of 2 V or −2 V on wire (to denote 1 and 0, respectively).

- X = voltage sent (2 or −2)
- Y = noise, $Y \sim \mathcal{N}(0, 1)$
- $R = X + Y$ voltage received.



Decode: $\begin{cases} 1 & \text{if } R \geq 0.5 \\ 0 & \text{otherwise.} \end{cases}$

1. What is P(decoding error | original bit is 1)? i.e., we sent 1, but we decoded as 0?

$$P(R < 0.5 | X = 2) = P(2 + Y < 0.5) = P(Y < -1.5)$$

Y is Standard Normal

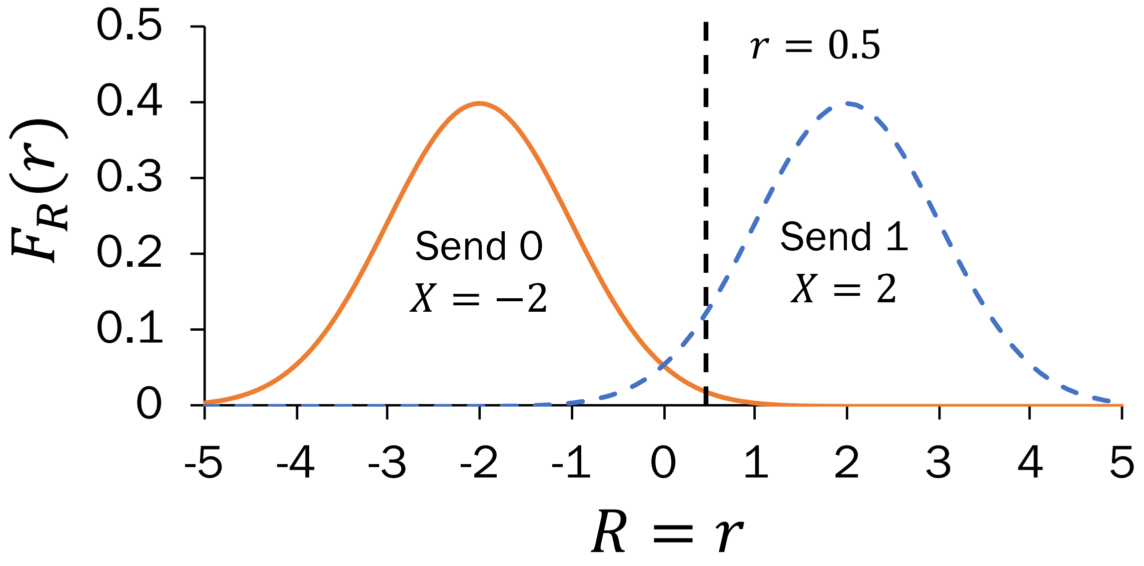
$$= \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$

Noisy Wires

Send a voltage of 2 V or -2 V on wire (to denote 1 and 0, respectively).

- X = voltage sent (2 or -2)
- Y = noise, $Y \sim \mathcal{N}(0, 1)$
- $R = X + Y$ voltage received.

Decode: $\begin{cases} 1 & \text{if } R \geq 0.5 \\ 0 & \text{otherwise.} \end{cases}$



1. What is P(decoding error | original bit is 1)? i.e., we sent 1, but we decoded as 0?

   0.0668

2. What is P(decoding error | original bit is 0)?

$$P(R \geq 0.5 | X = -2) = P(-2 + Y \geq 0.5) = P(Y \geq 2.5) \approx 0.0062$$

Asymetric decoding probability: We would like to avoid mistaking a 0 for 1. Errors the other way are tolerable.

Challenge: Sampling with the Normal RV

ELO ratings

What is the probability that the Warriors win?
How do you model zero-sum games?

ELO ratings

Each team has an ELO score S, calculated based on their past performance.

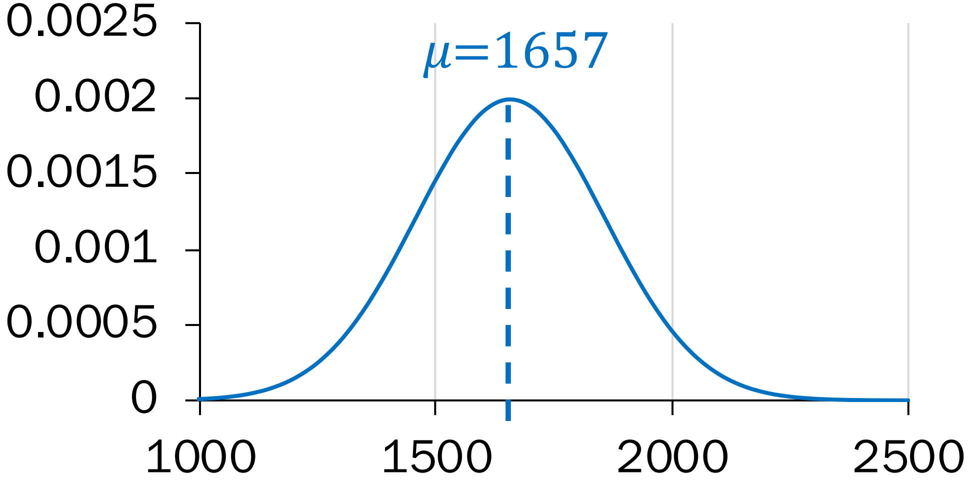
- Each game, a team has ability $A \sim \mathcal{N}(S, 200^2)$.
- The team with the higher sampled ability wins.

Arpad Elo

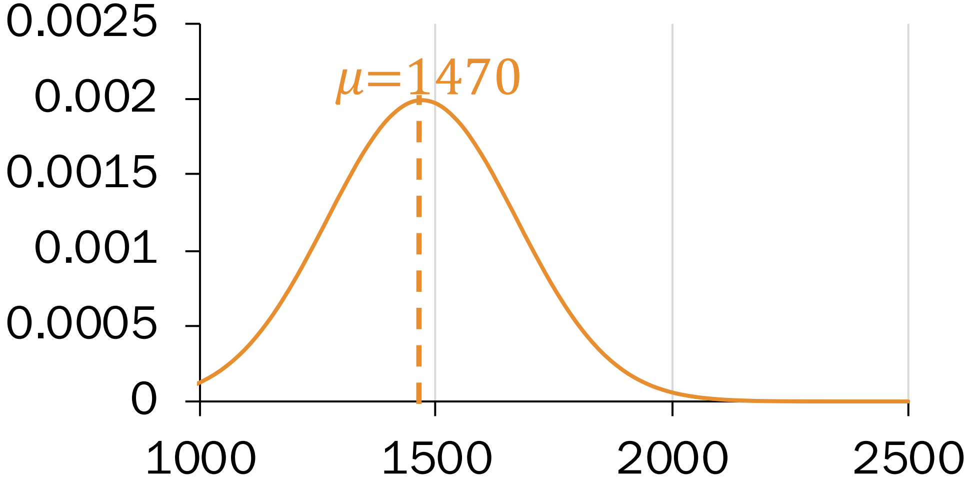
What is the probability that Warriors win this game?

Want: $P(\text{Warriors win}) = P(A_W > A_B)$

Warriors $A_W \sim \mathcal{N}(S = 1657, 200^2)$



Opponents $A_B \sim \mathcal{N}(S = 1470, 200^2)$



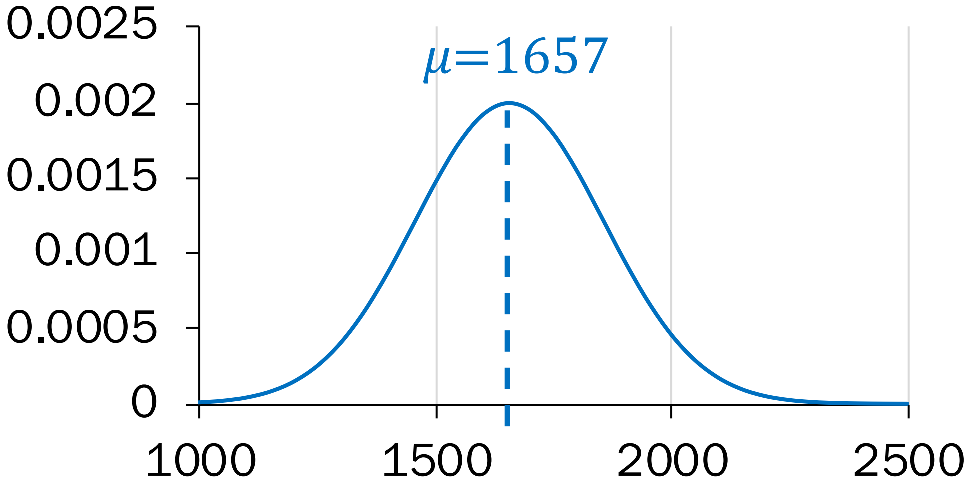
ELO ratings

Want: $P(\text{Warriors win}) = P(A_W > A_B)$

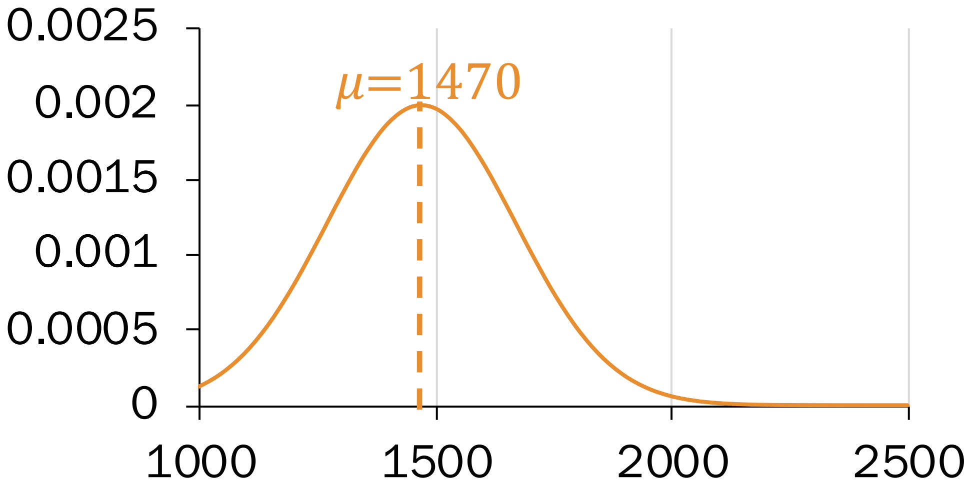
```
from scipy import stats
WARRIORS_ELO = 1657
OPPONENT_ELO = 1470
STDEV = 200
NTRIALS = 10000

nSuccess = 0
for i in range(NTRIALS):
  w = stats.norm.rvs(WARRIORS_ELO, STDEV)
  b = stats.norm.rvs(OPPONENT_ELO, STDEV)
  if w > b:
    nSuccess += 1
print("Warriors sampled win fraction",
      float(nSuccess) / NTRIALS)
```

≈ 0.7488, calculated by sampling

Warriors $A_W \sim \mathcal{N}(S = 1657, 200^2)$



Opponents $A_B \sim \mathcal{N}(S = 1470, 200^2)$



Is there a better way?

$$P(A_W > A_B)$$

- This is a probability of an event involving ***two*** random variables!
- We'll solve this problem analytically in two weeks' time.

Big goal for next time: Events involving two ***discrete*** random variables.
Stay tuned!