10: The Normal Distribution

Lisa Yan

October 14, 2019

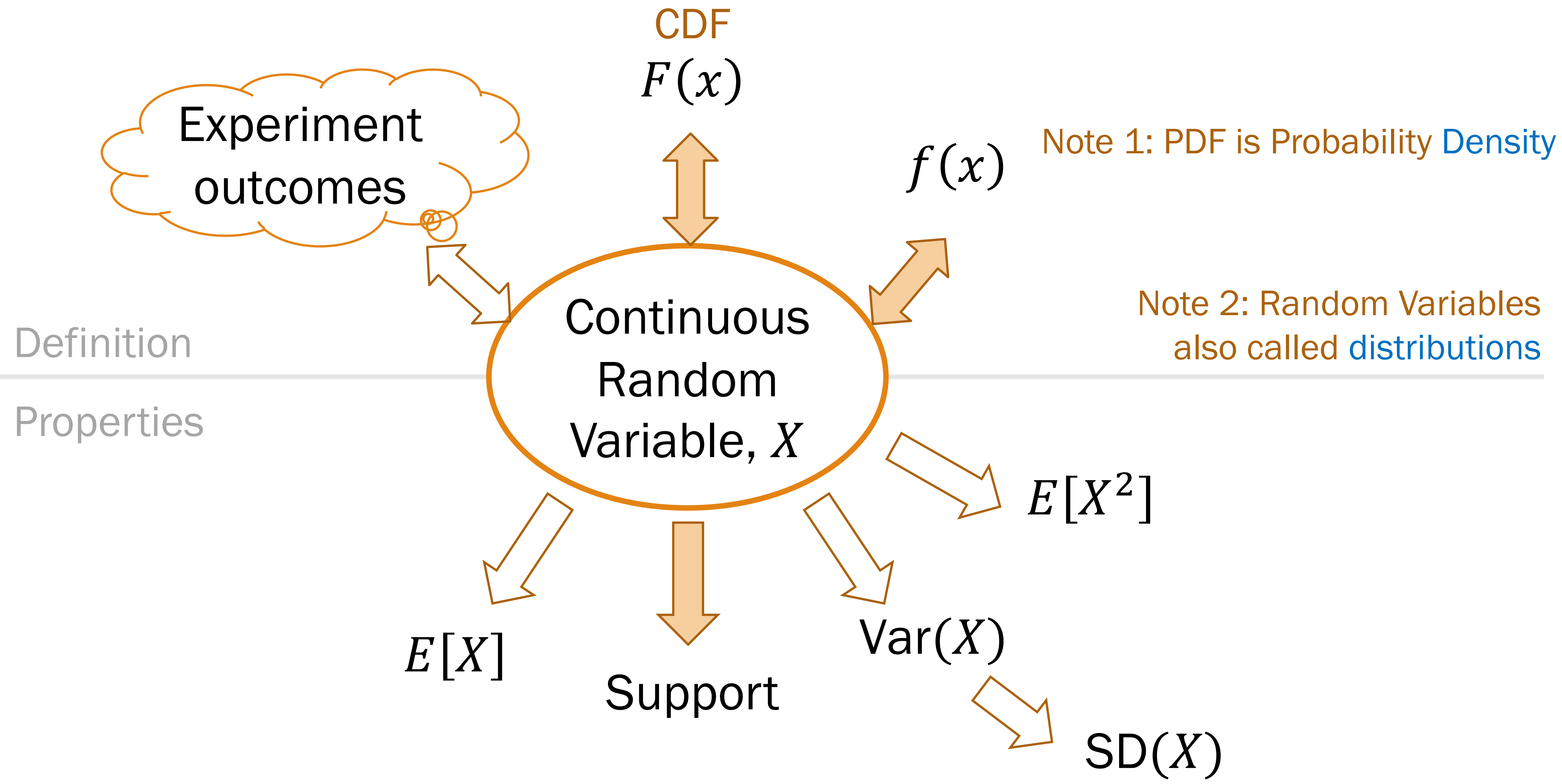
In today's class

Each question (from a unique person) today as a $p = 0.3$ probability of winning a pomegranate.

Let X be the number of questions until we run out of fruit.

$$X \sim \text{NegBin}(r = 5, p = 0.3)$$

Continuous random variables

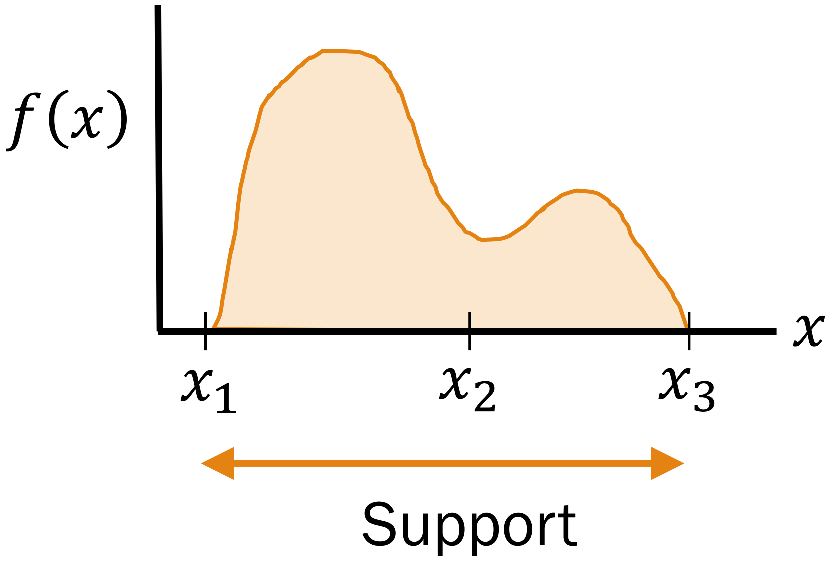


Probability from a PDF

For a continuous RV X with PDF f,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$



Loving, not scary

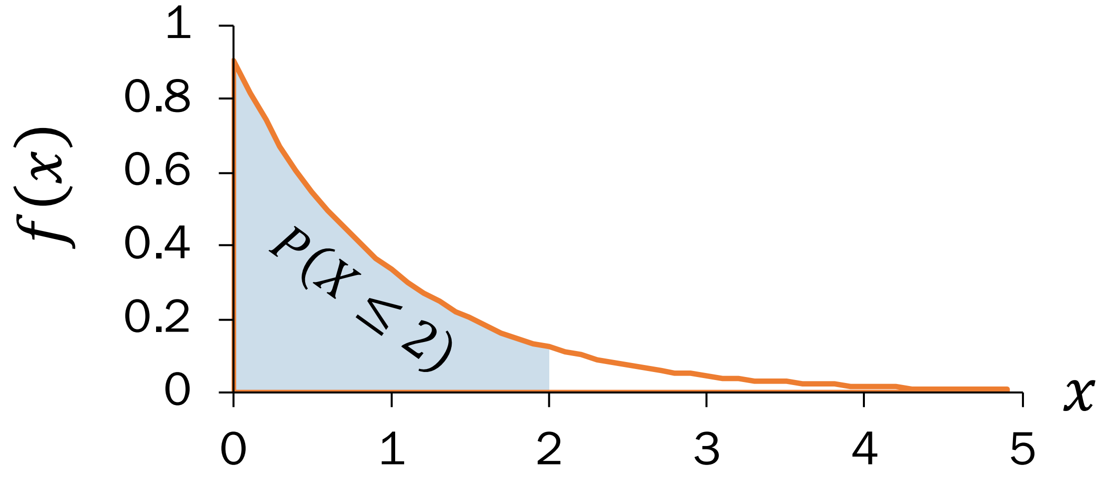
CDF

For a continuous random variable X with PDF $f(x)$, the CDF of X is

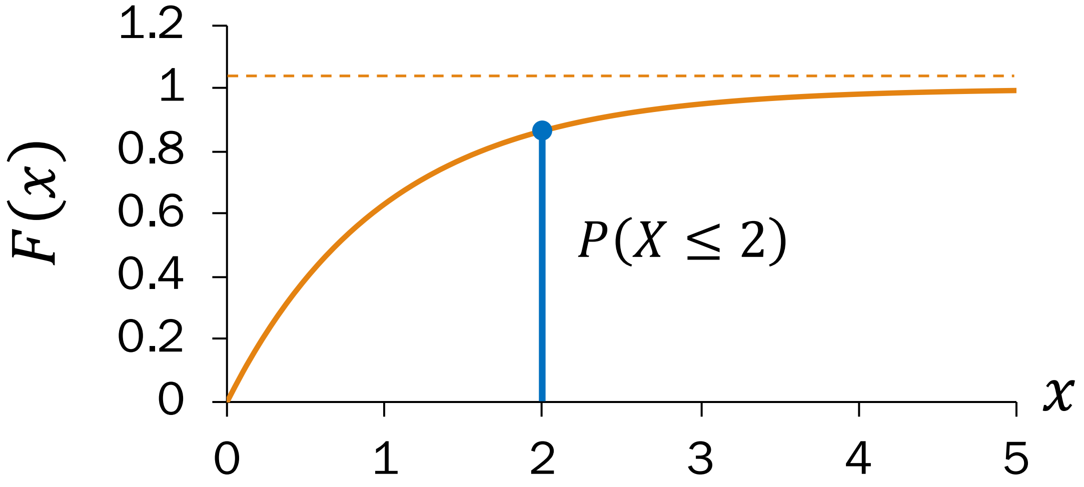
$$P(X \leq a) = F(a) = \int_{-\infty}^{a} f(x)dx$$

Exponential RV, X

PDF $f(x) = \lambda e^{-\lambda x}$
(not a probability)



CDF $F(x) = 1 - e^{-\lambda x}$
(a probability)



CDF of a continuous RV

(where we left off)

For a continuous random variable X with PDF $f(x)$, the CDF of X is

$$F(a) = \int_{-\infty}^{a} f(x)dx$$

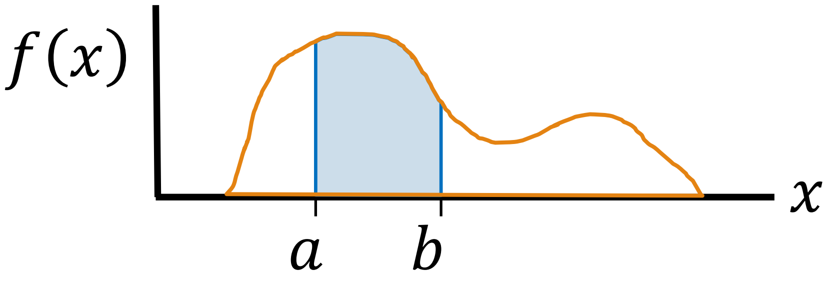
Important property: $P(a \leq X \leq b) = F(b) - F(a)$

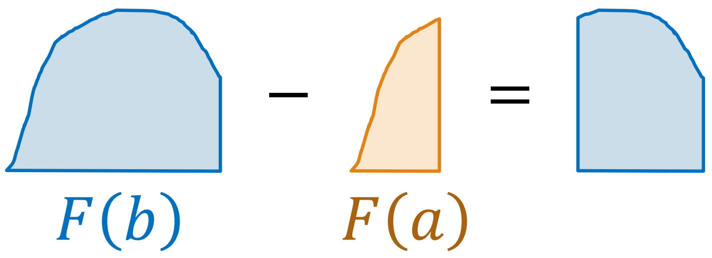
Proof:

$$F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$

$$= \left(\int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx\right) - \int_{-\infty}^{a} f(x)dx$$

$$= \int_{a}^{b} f(x)dx$$





Earthquakes

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?
3. What is the probability of zero major earthquakes next year?

We know:

$$500 \frac{\text{years}}{\text{earthquake}}$$

$$0.002 \frac{\text{earthquakes}}{\text{year}}$$

$$1 \frac{\text{earthquakes}}{\text{500 years}}$$

Strategy:

A. Bayes' Theorem
B. Total Probability
C. Uniform RV
D. Poisson RV
E. Exponential RV

*In California, according to historical data form USGS, 2015

Earthquakes

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?
3. What is the probability of zero major earthquakes next year?

We know:

$$500 \ \frac{\text{years}}{\text{earthquake}}$$

$$0.002 \ \frac{\text{earthquakes}}{\text{year}}$$

$$1 \ \frac{\text{earthquakes}}{500 \text{ years}}$$

Strategy:

A. Bayes' Theorem
B. Total Probability
C. Uniform RV
D. Poisson RV
E. Exponential RV

*In California, according to historical data form USGS, 2015

Earthquakes

$X \sim \text{Poi}(\lambda) \quad p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?
3. What is the probability of zero major earthquakes next year?

Strategy: D. Poisson RV

Strategy: E. Exponential RV

Define events/RVs & state goal

X: # earthquakes next year

$X \sim \text{Poi}(\lambda = 0.002)$

$\lambda: \frac{\text{earthquakes}}{\text{year}}$

Want: $P(X = 0)$

Solve

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} \approx 0.998$$

*In California, according to historical data form USGS, 2015

Earthquakes

$X \sim \text{Exp}(\lambda) \quad F(x) = 1 - e^{-\lambda x}$

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?
3. What is the probability of zero major earthquakes next year?

Strategy: D. Poisson RV

Define events/RVs & state goal

X: # earthquakes next year

$X \sim \text{Poi}(\lambda = 0.002)$

$\lambda: \frac{\text{earthquakes}}{\text{year}}$

Want: $P(X = 0)$

Solve

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} \approx 0.998$$

Strategy: E. Exponential RV

Define events/RVs & state goal

X: when first earthquake happens

$X \sim \text{Exp}(\lambda = 0.002)$

Want: $P(X > 1) = 1 - F(1)$

Solve

$$P(X > 1) = 1 - \left(1 - e^{-\lambda \cdot 1}\right) = e^{-\lambda}$$

*In California, according to historical data form USGS, 2015

Today's plan

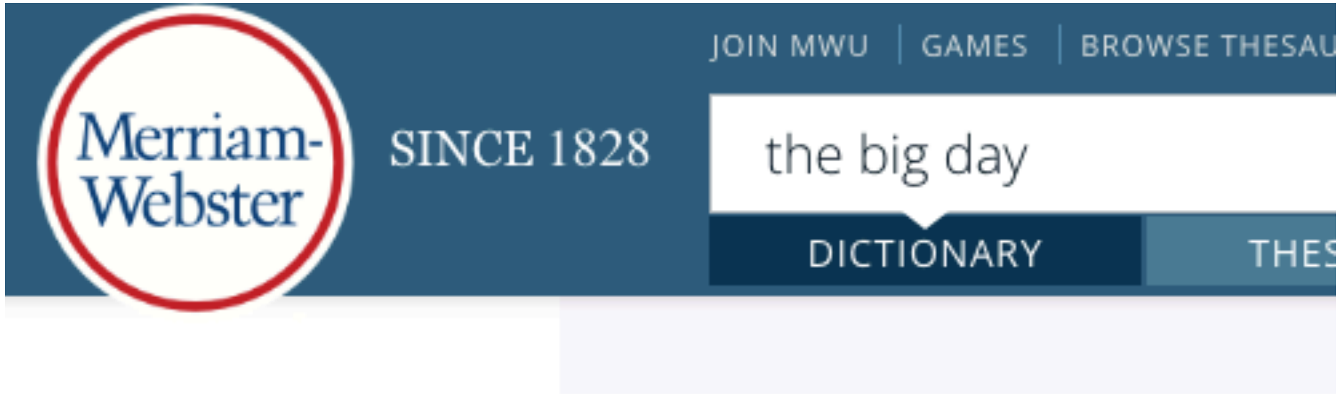
Normal (Gaussian) RV

The Standard Normal, Z

Sampling with the Normal

Normal approximation for Binomial (if time)

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: $(-\infty, \infty)$

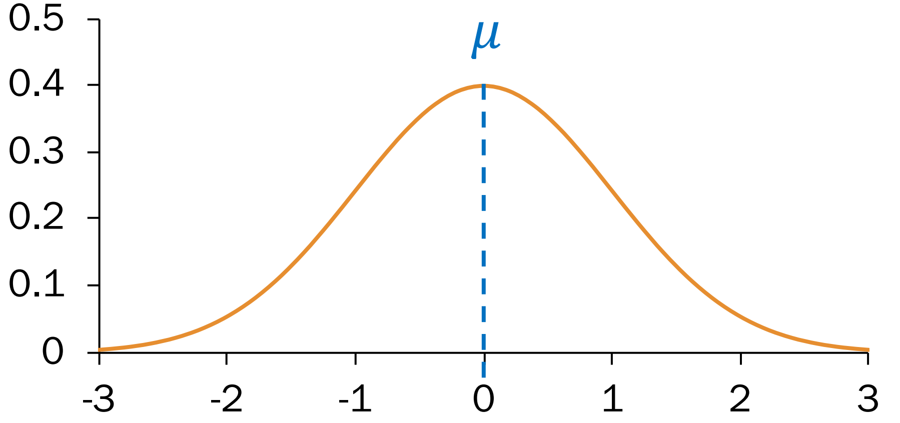
PDF	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$
Expectation	$E[X] = \mu$
Variance	$\text{Var}(X) = \sigma^2$

Other names: **Gaussian** random variable

mean

variance

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



Normal Random Variable

mean variance

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

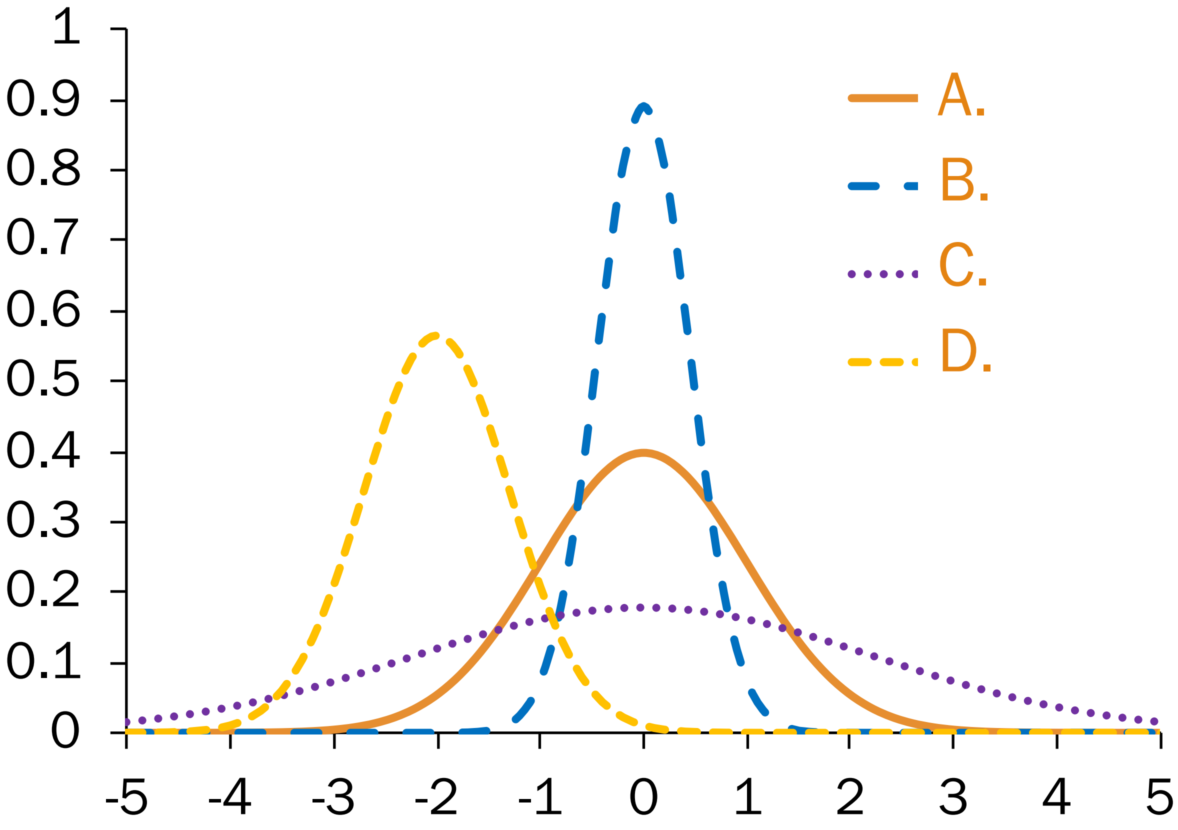
Match PDF to distribution:

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

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

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

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



Normal Random Variable

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

mean: μ — variance: σ^2

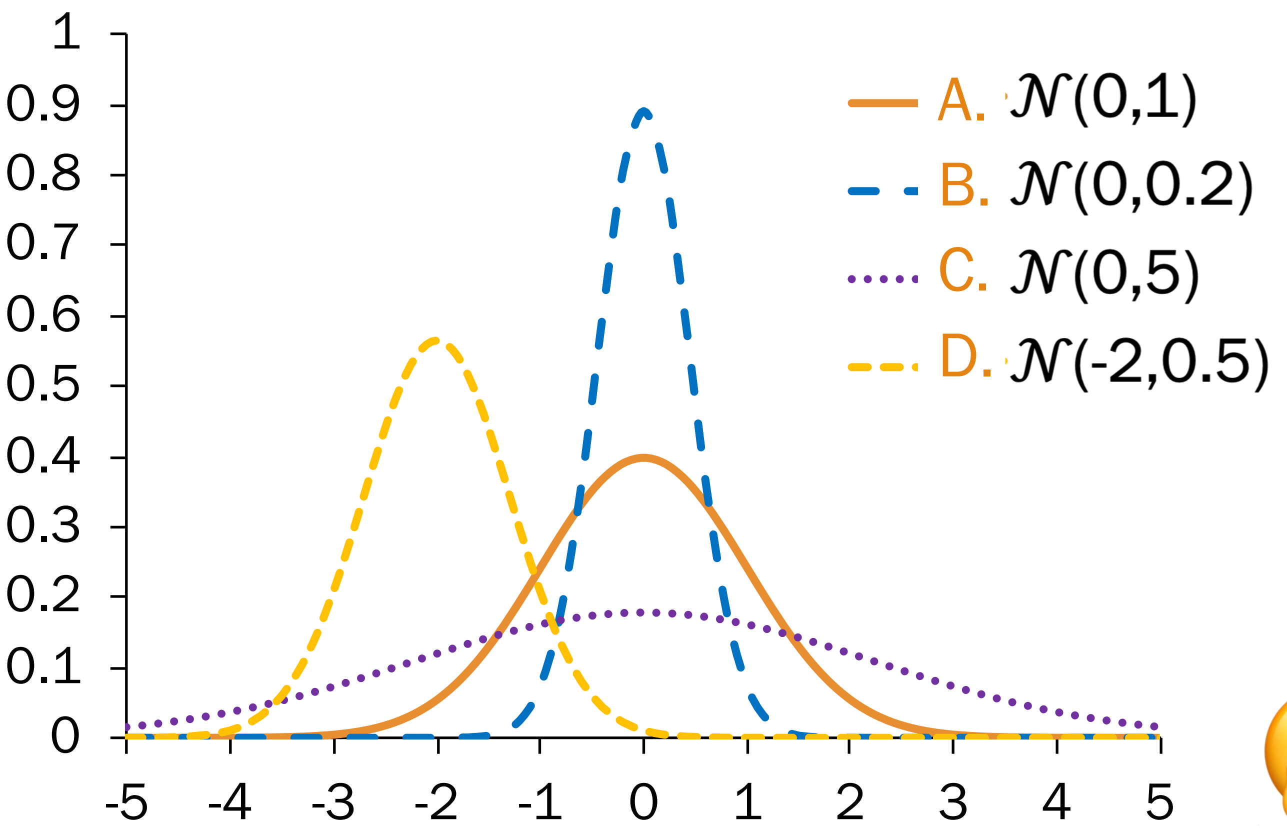
Match PDF to distribution:

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

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

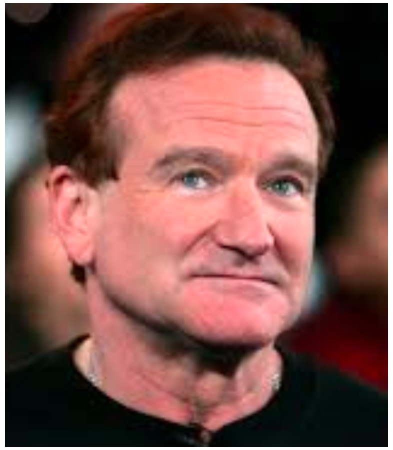
$\mathcal{N}(0, 5)$ (C)

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



Carl Friedrich Gauss

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



Johann Carl Friedrich Gauss (/gaʊs/; German: *Gauß* [gaʊs] (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)

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

Okay, so why the Normal?

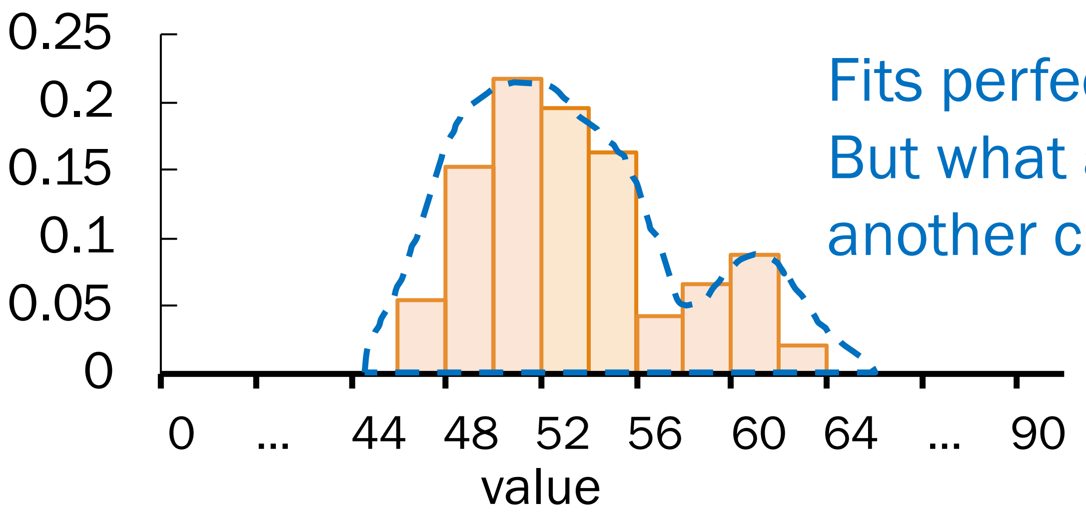
Part of CS109 learning goals:

- 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?

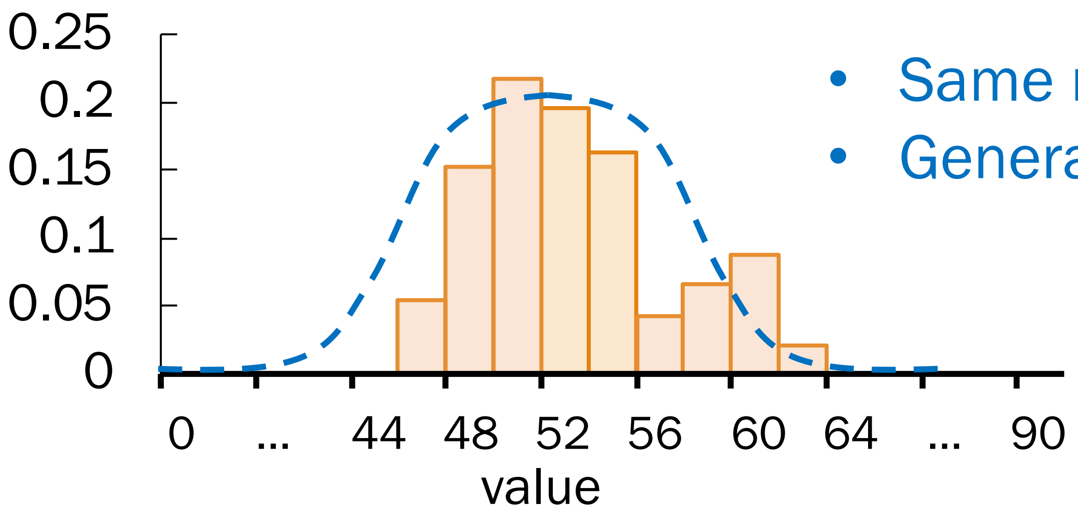
Part of CS109 learning goals:

- 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

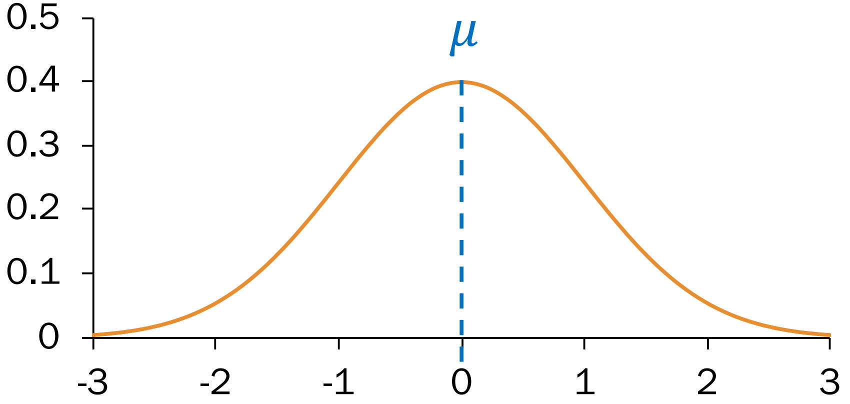
Occam's Razor:

"Non sunt multiplicanda entia sine necessitate."

Entities should not be multiplied without necessity.

A Gaussian maximizes entropy for a given mean and variance.

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

(we'll spend the next few slides getting here)

Linear transformations of Normal RVs

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. If $Y = aX + b$, then
Y is also Normal, where $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Proof:

- Y is also Normal

Proof in Ross, 10th ed (Section 5.4)

- $$E[Y] = E[aX + b] = aE[X] + b = a\mu + b$$

Linearity of Expectation

- $$\text{Var}(Y) = \text{Var}(aX + b) = a^2\text{Var}(X) = a^2\sigma^2$$

$\text{Var}(aX + b) = a^2\text{Var}(X)$

Today's plan

Normal (Gaussian) RV

The Standard Normal, Z

Sampling with the Normal

Normal approximation for Binomial (if time)

Standard Normal RV, Z

The **Standard Normal** random variable Z is defined as follows:

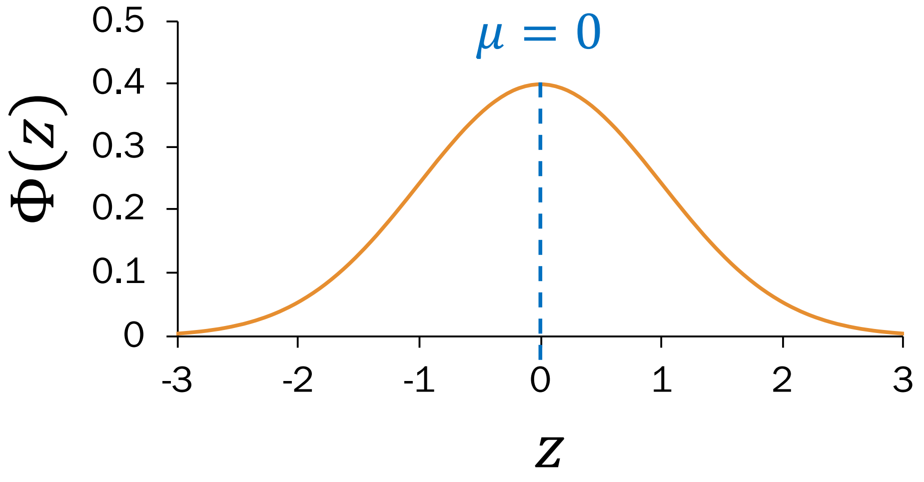
$$Z \sim \mathcal{N}(0, 1)$$

Expectation $E[Z] = \mu = 0$

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

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

Other names: **Unit Normal**



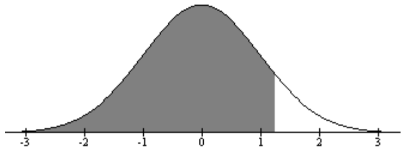
CDF of Z defined as:

$$P(Z \leq z) = \Phi(z)$$

Φ has been numerically computed

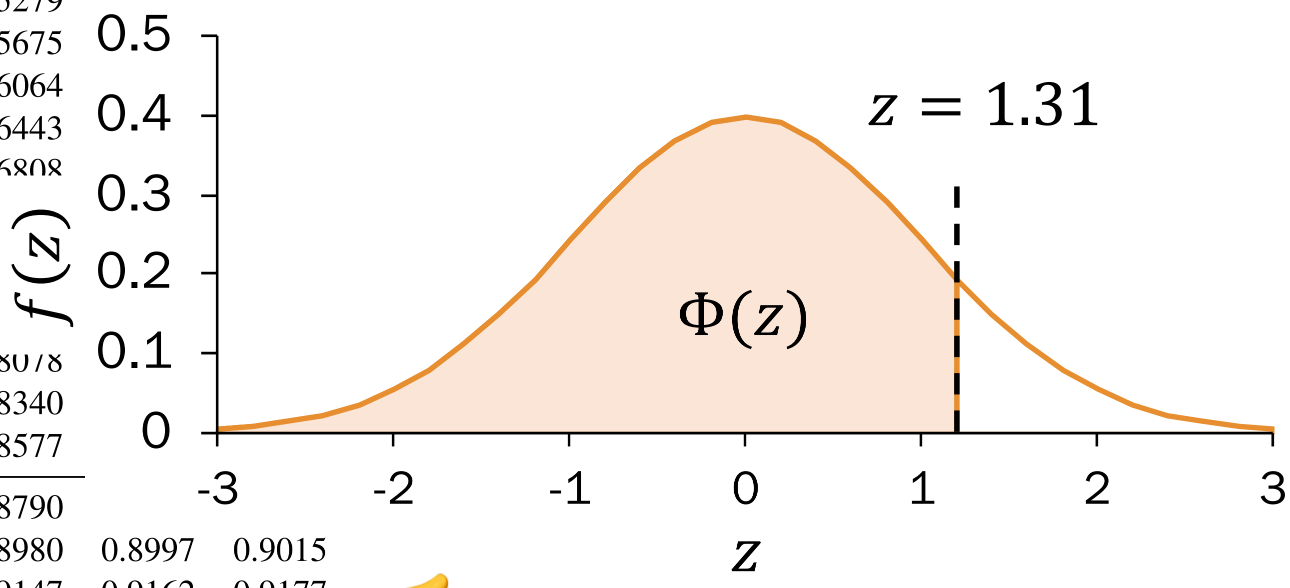
Standard Normal Table

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



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	[illegible]	[illegible]
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	[illegible]	[illegible]
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	[illegible]	[illegible]
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	[illegible]	[illegible]
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	[illegible]	[illegible]
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	[illegible]	[illegible]	[illegible]
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	[illegible]	[illegible]	[illegible]
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	[illegible]	[illegible]	[illegible]
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	[illegible]	[illegible]
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	[illegible]	[illegible]
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	[illegible]	[illegible]
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	[illegible]	[illegible]
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

$$P(Z \leq 1.31) = \Phi(1.31)$$
$$= 0.9049$$

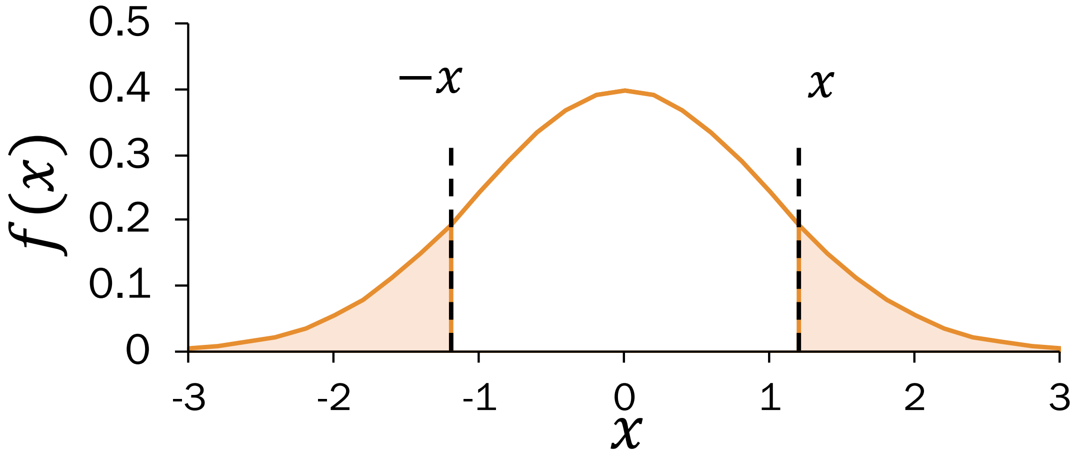


Standard Normal Table has probabilities $\Phi(z)$ for $z \geq 0$.

Using symmetry of the Normal RV

Recall that a Normal RV has a **symmetric** PDF.

$Z \sim \mathcal{N}(0,1)$ has a numeric lookup table for $\Phi(x)$, where $\boldsymbol{x \geq 0}$.



How do we compute the following probabilities in terms of $\Phi(x)$?

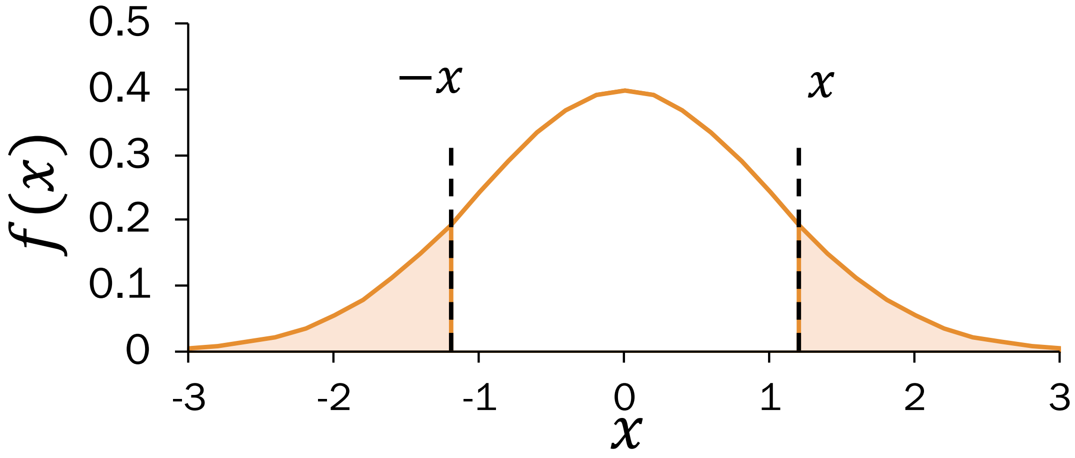
1. $P(Z \leq x)$ ——— A
2. $P(Z < x)$
3. $P(Z \geq x)$
4. $P(Z \leq -x)$
5. $P(Z \geq -x)$
6. $P(a < Z < b)$

A. $\Phi(x)$
B. $1 - \Phi(x)$
C. $\Phi(b) - \Phi(a)$

Using symmetry of the Normal RV

Recall that a Normal RV has a **symmetric** PDF.

$Z \sim \mathcal{N}(0,1)$ has a numeric lookup table for $\Phi(x)$, where $\boldsymbol{x \geq 0}$.



How do we compute the following probabilities in terms of $\Phi(x)$?

1. $P(Z \leq x) \quad = \Phi(x)$
2. $P(Z < x) \quad = \Phi(x)$
3. $P(Z \geq x) \quad = 1 - \Phi(x)$
4. $P(Z \leq -x) \quad = 1 - \Phi(x)$
5. $P(Z \geq -x) \quad = \Phi(x)$
6. $P(a < Z < b) \quad = \Phi(b) - \Phi(a)$

A. $\Phi(x)$
B. $1 - \Phi(x)$
C. $\Phi(b) - \Phi(a)$

Use symmetry to compute probabilities $\Phi(z)$ for $z < 0$.

Probabilities for a general Normal RV

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

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF F. Then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, where Φ is the Standard Normal $Z \sim \mathcal{N}(0, 1)$.

Proof:

$$F(x) = P(X \le x)$$ Definition of CDF

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

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

- Let $Z = \frac{X-\mu}{\sigma} = \frac{X}{\sigma} - \frac{\mu}{\sigma}$, a linear transform of X.
- Then Z is normal, where $Z \sim \mathcal{N}(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2)$.
- Then $Z \sim \mathcal{N}(0,1)$ with CDF Φ.

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

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?

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

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

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

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

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

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

Break for jokes/ announcements

Announcements

Concept checks

Due date: Tuesdays 1:00pm

Selected anonymous answers (with consent)

Problem Set 3

Out: today

Due: Wednesday 10/23

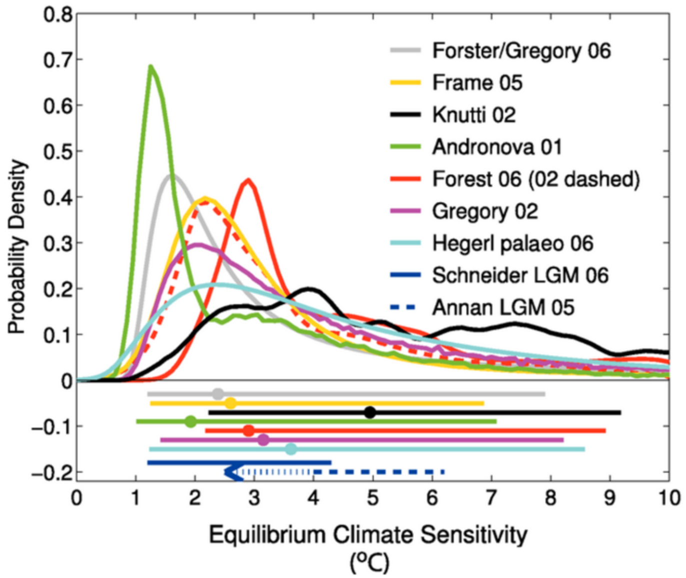
Covers: through this Wednesday

Late days

Free: 2 free class days

No late days after last day of quarter (Fri 12/7)

(note PS#6 due Wed 12/5)



Model the tail of climate change on problem set 3!

Which is random?

Sequence 1

TTHHTHTTHTTTHTTTHTTTHTTHTHHTHHTHTHHTTTHHTHTHTTHTHHT
THTHHTHTTTHHTTHHTTHHHTHHTHTTHTHTTHHTHHHTTHTHTTTHHTT
HTHTHTHTHTTHTHTHHHTTHTHTHHTHHHTHTHTTHTTHHTHTHTHTTHH
TTHTHTTHHHTHTHTHTTHTTHHTTHTHHTHHHTTHHTHTTHTHTHTHTHT
HTHHHTHTHTHTTHTHHTHTHTTHTTTHHTHTTTHTHHTHHHHTTTHHTHT
HTHTHHHTTHHTHTTTHTHHTHTHTHHTHTTHTTHTHHTHTHTTT

Sequence 2

HTHHHTHTTHHTTTTTTTTHHHTTTHHTTTTHHTTHHHTTHTHTTTTTTHT
HTTTTHHHHTHTHTTHTTTHTTHTTTTHTHHTHHHHTTTTTHHHHTHHHTT
TTHTHTTHHHHTHHHHHHHHTTHHTHHTHHHHHHHTTHTHTTTHHTTTTHT
HHTTHTTHTHTHTTHHHHHTTHTTTHTHTHHTTTTHTTTTTHHTHTHHHHT
TTTHTHHHHHHTHTHTHTHHHTHTTHHHTHHHHHHTHHHTHTTTHHHTTTH
HTHTTHHTHHHTHTTHTTHTTTHHTHTHTTTTHTHTHTTHTHTHT

Find out on problem set 3!

New-school lookup tables

Standard Normal Table

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

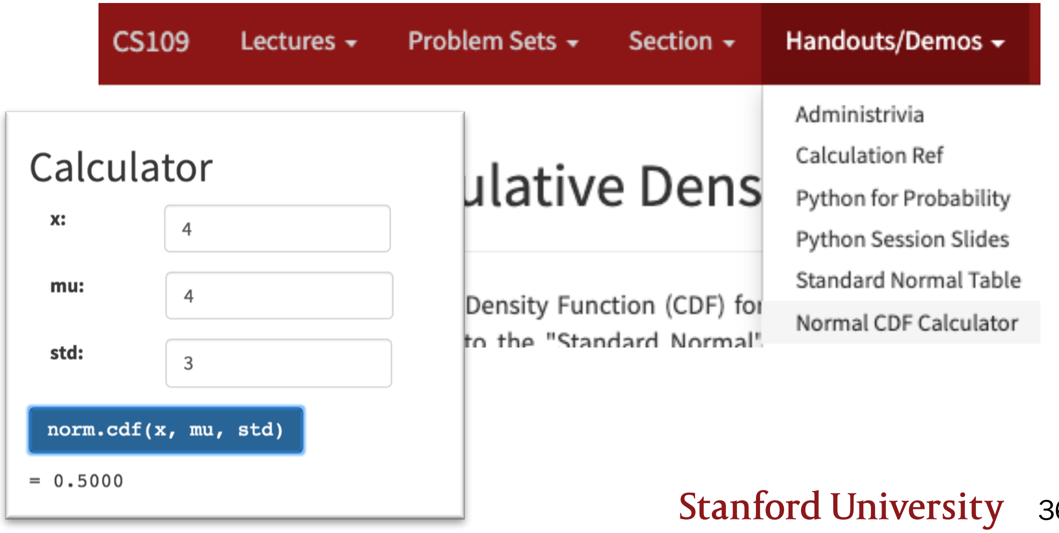
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	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

(You should know how to use a lookup table for the exam)

Python 3:

```
scipy.stats.norm(mean, std).cdf(x)
```

CS109 website:



Get your Gaussian On

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

"Get ur freak on"
Missy Elliott, 2001

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

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

Look up $\Phi(z)$ in table

$$P(X > 0) = 1 - F(0)$$

$$= 1 - \Phi\left(\frac{0-\mu}{\sigma}\right)$$

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

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

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

$$P(X > 0) = 1 - F(0)$$

$$= 1 - \Phi\left(\frac{0-\mu}{\sigma}\right)$$

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

Look up $\Phi(z)$ in table

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

$$= \Phi\left(\frac{3}{4}\right)$$

$$\approx 0.7734$$

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

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

$$P(2 < X < 5) = F(5) - F(2)$$

$$= \Phi\left(\frac{5-3}{4}\right) - \Phi\left(\frac{2-3}{4}\right)$$

$$= \Phi\left(\frac{2}{4}\right) - \Phi\left(\frac{-1}{4}\right)$$

Get your Gaussian On

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

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

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

$$P(2 < X < 5) = F(5) - F(2)$$

$$= \Phi\left(\frac{5-3}{4}\right) - \Phi\left(\frac{2-3}{4}\right)$$

$$= \Phi\left(\frac{2}{4}\right) - \Phi\left(\frac{-1}{4}\right)$$

Look up $\Phi(z)$ in table

$$= \Phi\left(\frac{2}{4}\right) - \left(1 - \Phi\left(\frac{1}{4}\right)\right)$$

$$\approx 0.6915 - (1 - 0.5987)$$

$$= 0.2902$$

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

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

B. $P(X < 9) - P(X > -3)$

C. $\Phi\left(\frac{|X-3|}{4}\right) > \Phi\left(\frac{6-3}{4}\right)$

D. $1 - \Phi\left(\frac{6-3}{4}\right)$

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

Look up $\Phi(z)$ in table

A. $P(X < -3) + P(X > 9)$
B. $P(X < 9) - P(X > -3)$
C. $\Phi\left(\frac{|X-3|}{4}\right) > \Phi\left(\frac{6-3}{4}\right)$
D. $1 - \Phi\left(\frac{6-3}{4}\right)$

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

Look up $\Phi(z)$ in table

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

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

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

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

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

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

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

Look up $\Phi(z)$ in table

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

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

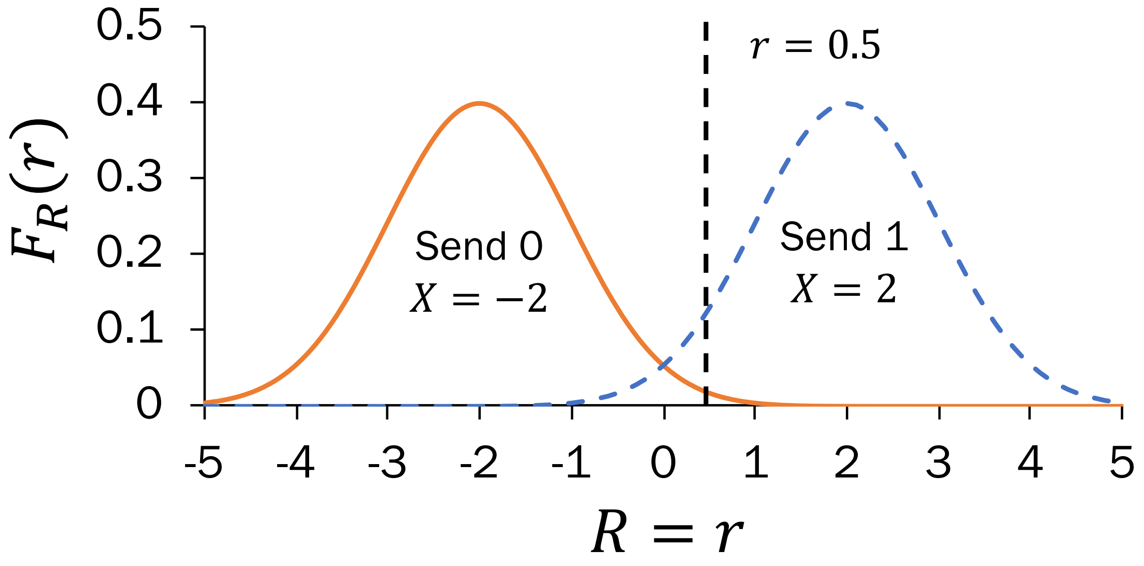
$$\approx 0.1337$$

Noisy Wires

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

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

Decode: 1 if $R \geq 0.5$
0 otherwise.



1. What is P(decoding error | original bit is 1)?

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

$$= \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$ Y is Standard Normal

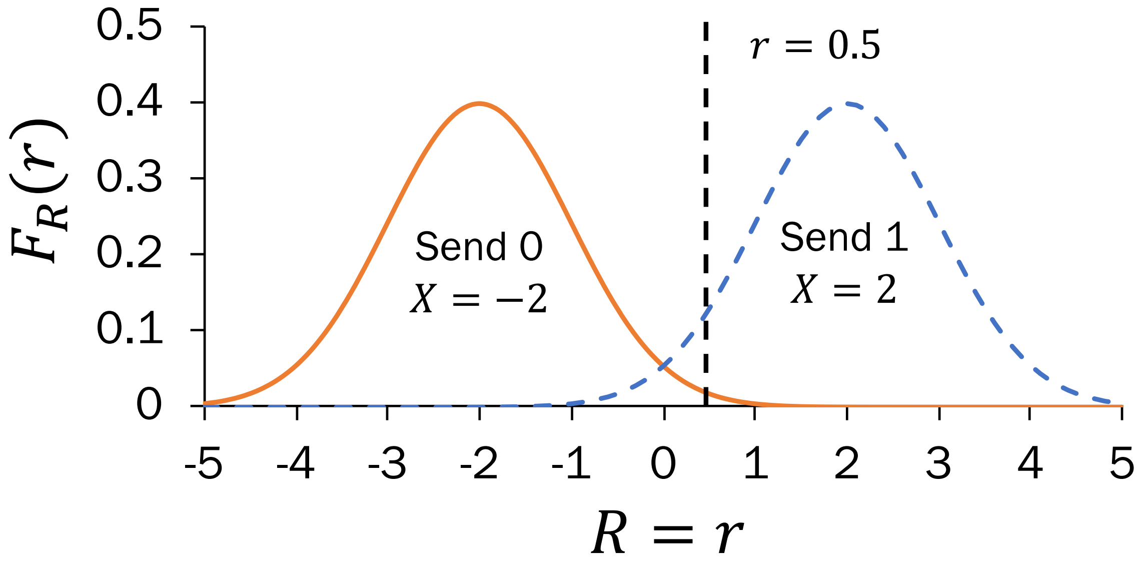
2. What is P(decoding error | original bit is 0)?

Noisy Wires

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

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

Decode: 1 if $R \geq 0.5$
0 otherwise.



1. What is P(decoding error | original bit is 1)?

$$P(R < 0.5 | X = 2) = P(2 + Y < 0.5) = P(Y < -1.5)$$
$$= \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$ Y is Standard Normal

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

$$P(R \geq 0.5 | X = -2) = P(-2 + Y \geq 0.5) = 1 - \Phi(2.5) \approx 0.0062$$

Today's plan

Normal (Gaussian) RV

The Standard Normal, Z

Sampling with the Normal

Normal approximation for Binomial (if time)

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.

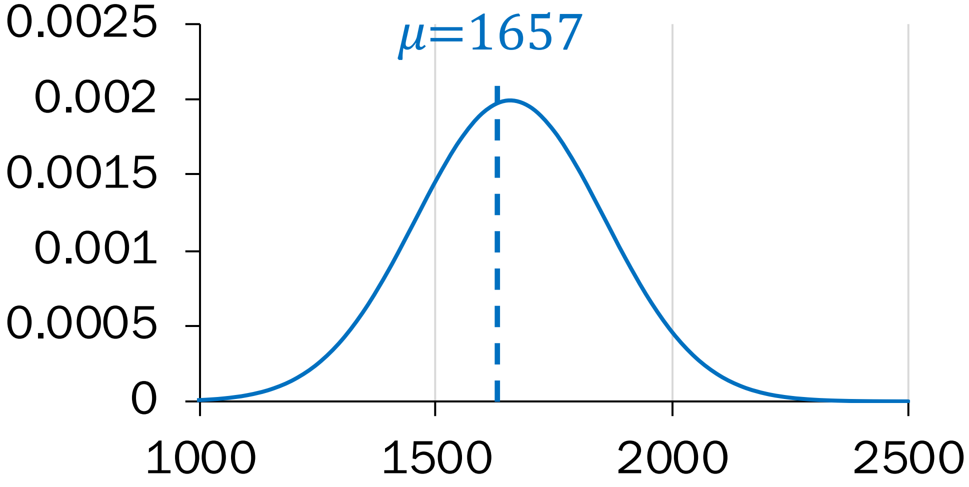
- Each game, a team has ability $A \sim \mathcal{N}(S, 200^2)$.
- The team with the higher sampled ability wins.

Arpad Elo

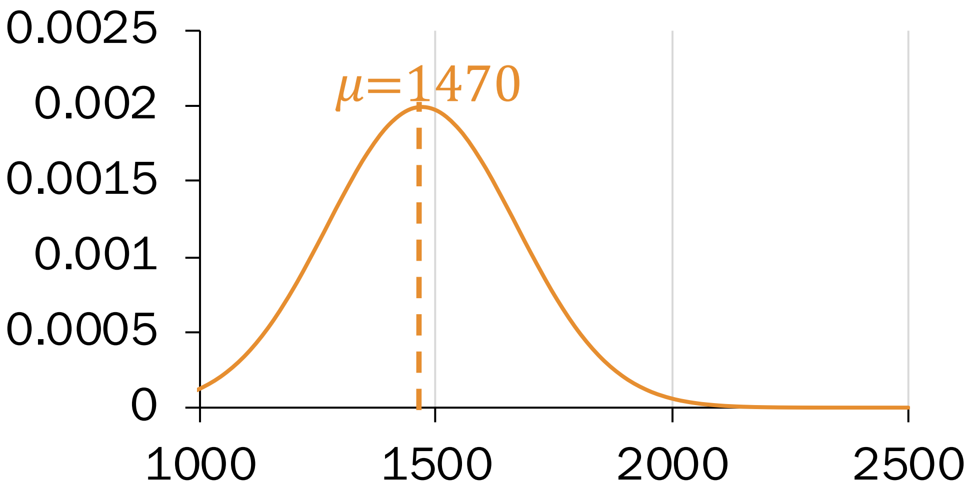
What is the probability that Warriors win this game?

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

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



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



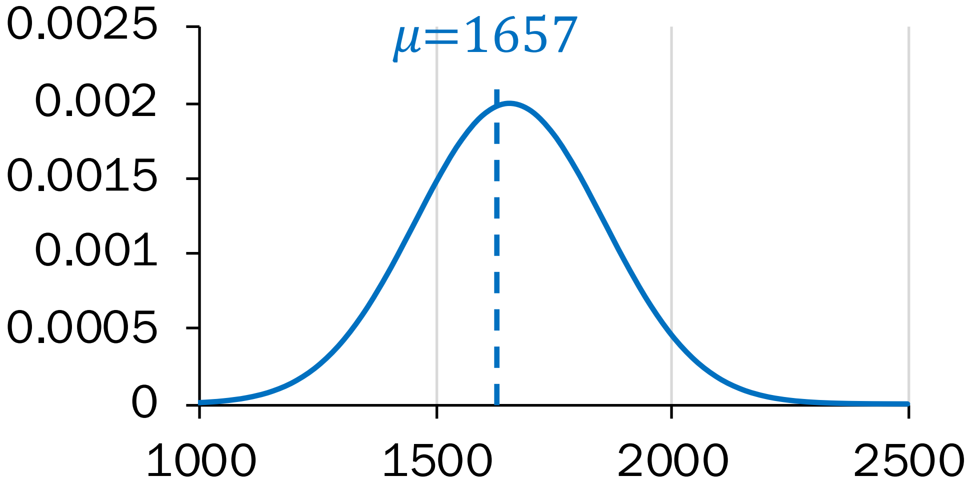
ELO ratings

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

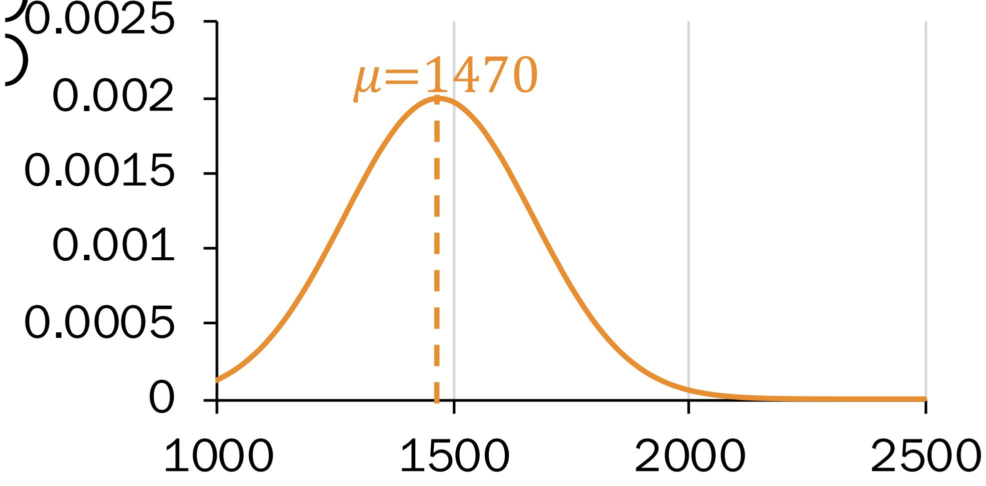
```python
from scipy import stats
WARRIORS_ELO = 1657
OPPONENT_ELO = 1470
STDEV = 200
NTRIALS = 10000

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



Today's plan

Normal (Gaussian) RV

The Standard Normal, Z

Sampling with the Normal

Normal approximation for Binomial (if time)

Website testing

- 100 people are given a new website design.
- $X = \#$ people whose time on site increases
- CEO will endorse the new design if $X \geq 65$.
- The design actually has no effect, so P(time on site increases) = 0.5 independently.

What is P(CEO endorses change)? *Give a numerical approximation*.

Strategy:

A. Poisson
B. Bayes' Theorem
C. Binomial
D. Normal (Gaussian)
E. Uniform

Website testing

- 100 people are given a new website design.
- X = # people whose time on site increases
- CEO will endorse the new design if $X \geq 65$.
- The design actually has no effect, so P(time on site increases) = 0.5 independently.

What is P(CEO endorses change)? *Give a numerical approximation*.

Strategy:

A. Poisson
B. Bayes' Theorem
C. Binomial
D. Normal (Gaussian) — Yes, actually!
E. Uniform