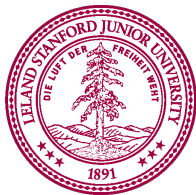
Normal Distribution

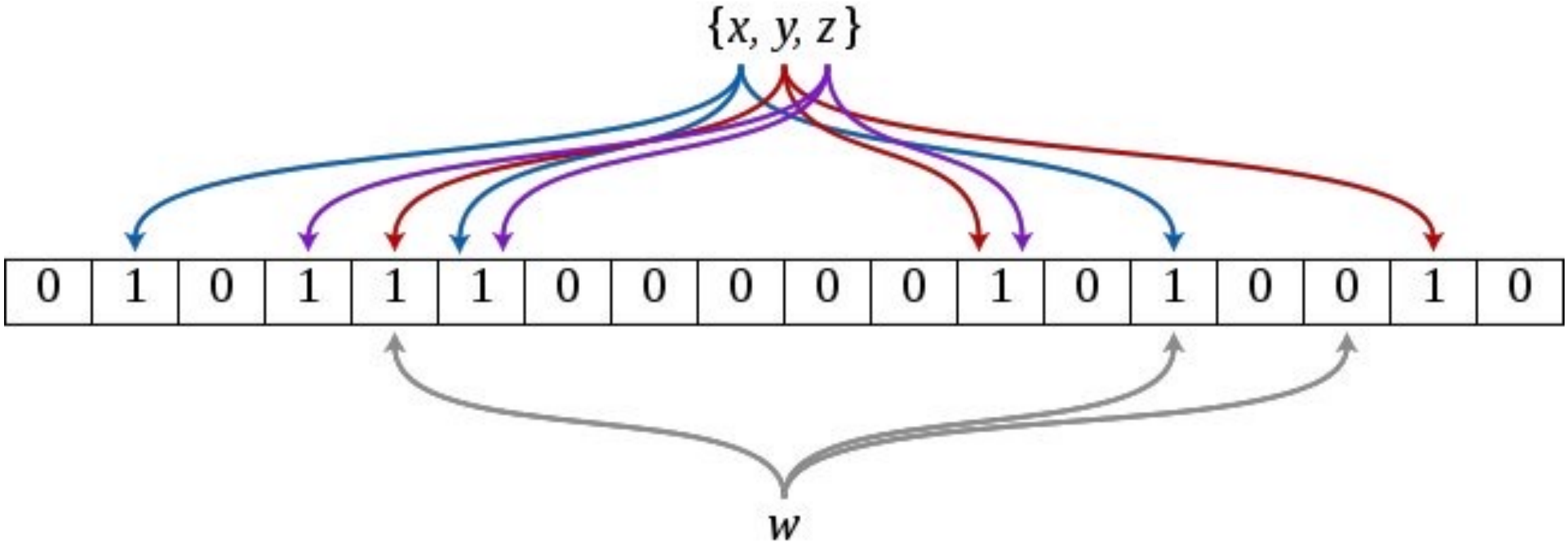
Chris Piech
CS109, Stanford University

Announcements

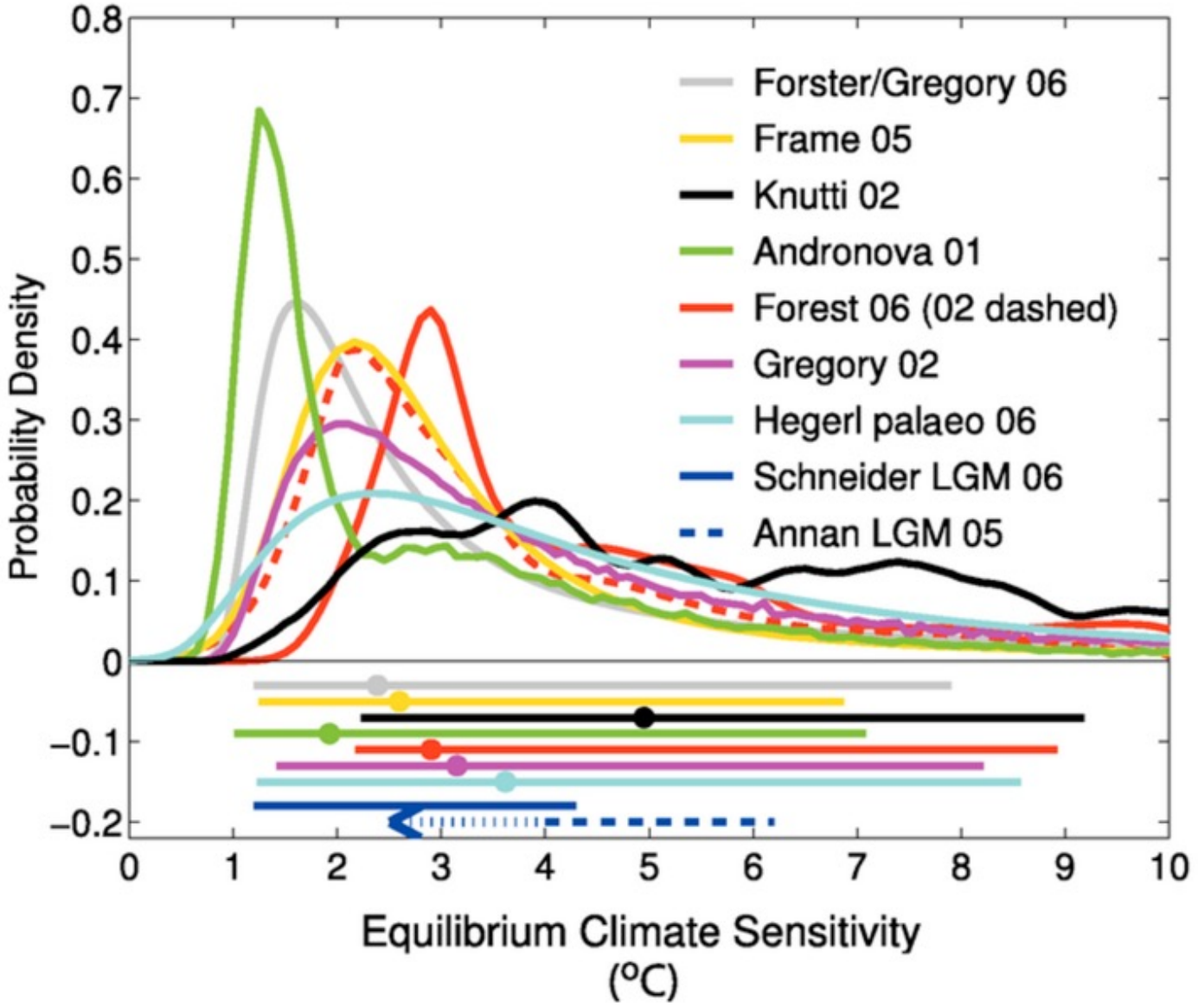
- Pset #3 is Out
- Happy Monday!



Bloom Filter



Climate Sensitivity



Which is Random?

Sequence 1:

TTHHTHTTHTTTHTTTHTTTHTTHTHHTHHTHTHHTTTHHTHTHTTHTHHTTHTHHTHTTTH
HTTHHTTHHHTHHTHTTHTHTTHHTHHHTTHTHTTTHHTTHTHTHTHTHTTHTHTHHHTTHT
HTHHTHHHTHTHTTHTTHHTHTHTHTTHHTTHTHTTHHHTHTHTHTTHTTHHTTHTHHTHHH
TTHHTHTTHTHTHTHTHTHTHHHTHTHTHTTHTHHTHTHTTHTTTHHTHTTTHTHHTHHHHT
TTHHTHTHTHTHHHTTHHTHTTTHTHHTHTHTHHTHTTHTTHTHHTHTHTHTTT

Sequence 2:

HTHHHTHTTHHTTTTTTTTHHHTTTHHTTTTHHTTHHHTTHTHTTTTTTTHTHTTTTHHHHTH
THTTHTTTHTTHTTTTHTHHTHHHHTTTTTHHHHTHHHTTTTHTHTTHHHHTHHHHHHHHTT
HHTHHTHHHHHHHTTHTHTTTHHTTTTHTHHTTHTTHTHTHTTHHHHHTTHTTTHTHTHHTT
TTHTTTTTHHTHTHHHHTTTTHTHHHHHHTHTHTHHHTHTTHHHTHHHHHHTHHHTHTTT
HHHTTTHHTHTTHHTHHHTHTTHTTHTTTHHTHTHTTTTHTHTHTTHTHTHT

Using Code to Learn

a. $X \sim \text{Ber}(p = 0.4)$
1 or 0 to indicate whether or not an underlying event was "successful."

```
from numpy.random import rand

def simulate_bernoulli(p=0.4):
    if rand() < p:
        return 1
    return 0
```

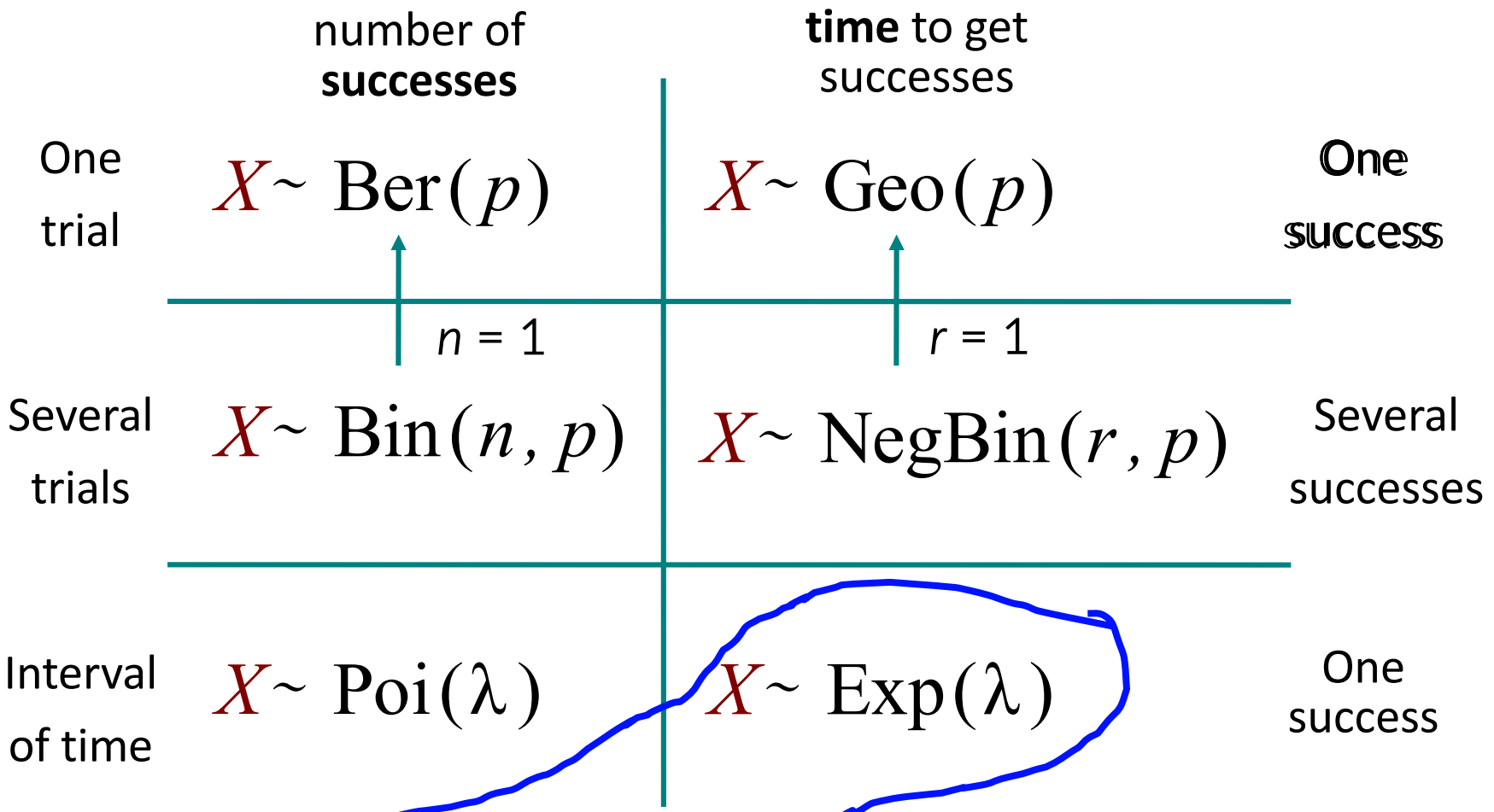
b. $X \sim \text{Bin}(n = 20, p = 0.4)$
The number of successes after 20 independent experiments.

c. $X \sim \text{Geo}(p = 0.03)$
The number of trials until the first success.

d. $X \sim \text{NegBin}(r = 5, p = 0.03)$
The number of trials until 5 successes.

Review

Review



$X \sim \text{Uni}(\alpha, \beta)$ All values are equally likely

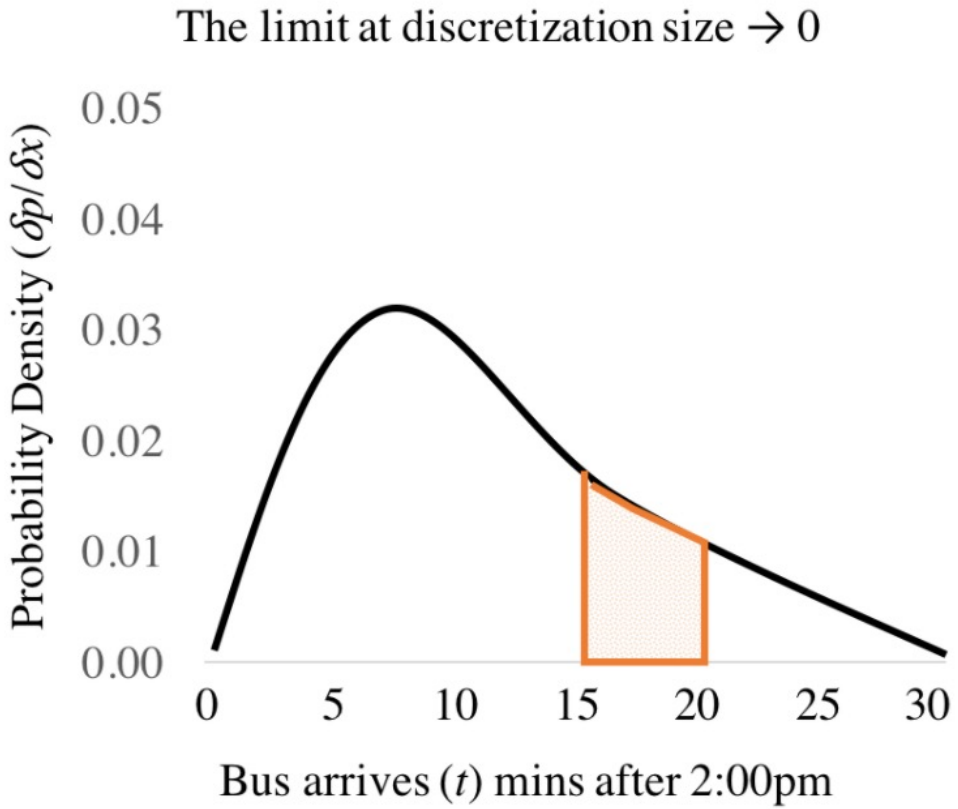
Review: Probability Density Function

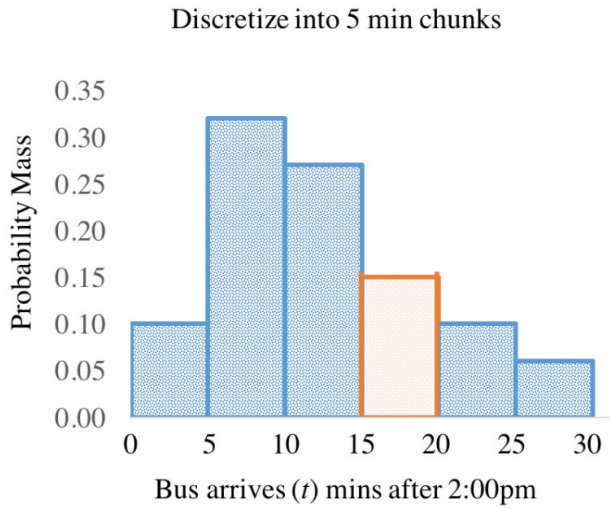
The **probability density function** (PDF) of a continuous random variable represents the **derivative** of probability at a given point.

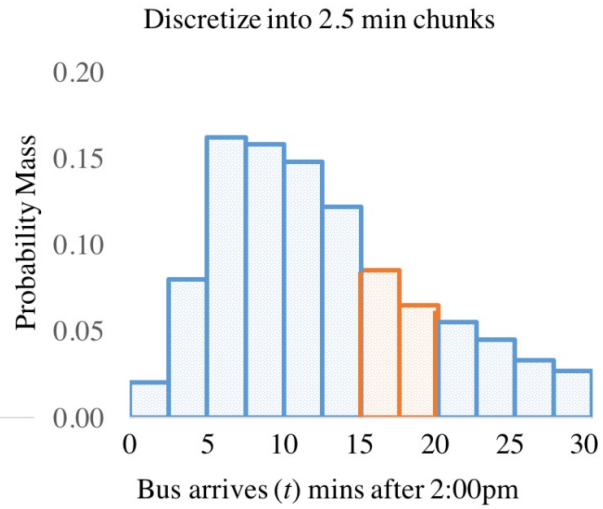
Units of probability *divided by units of X*.
Integrate it to get probabilities!

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

Review: Probability Density Function







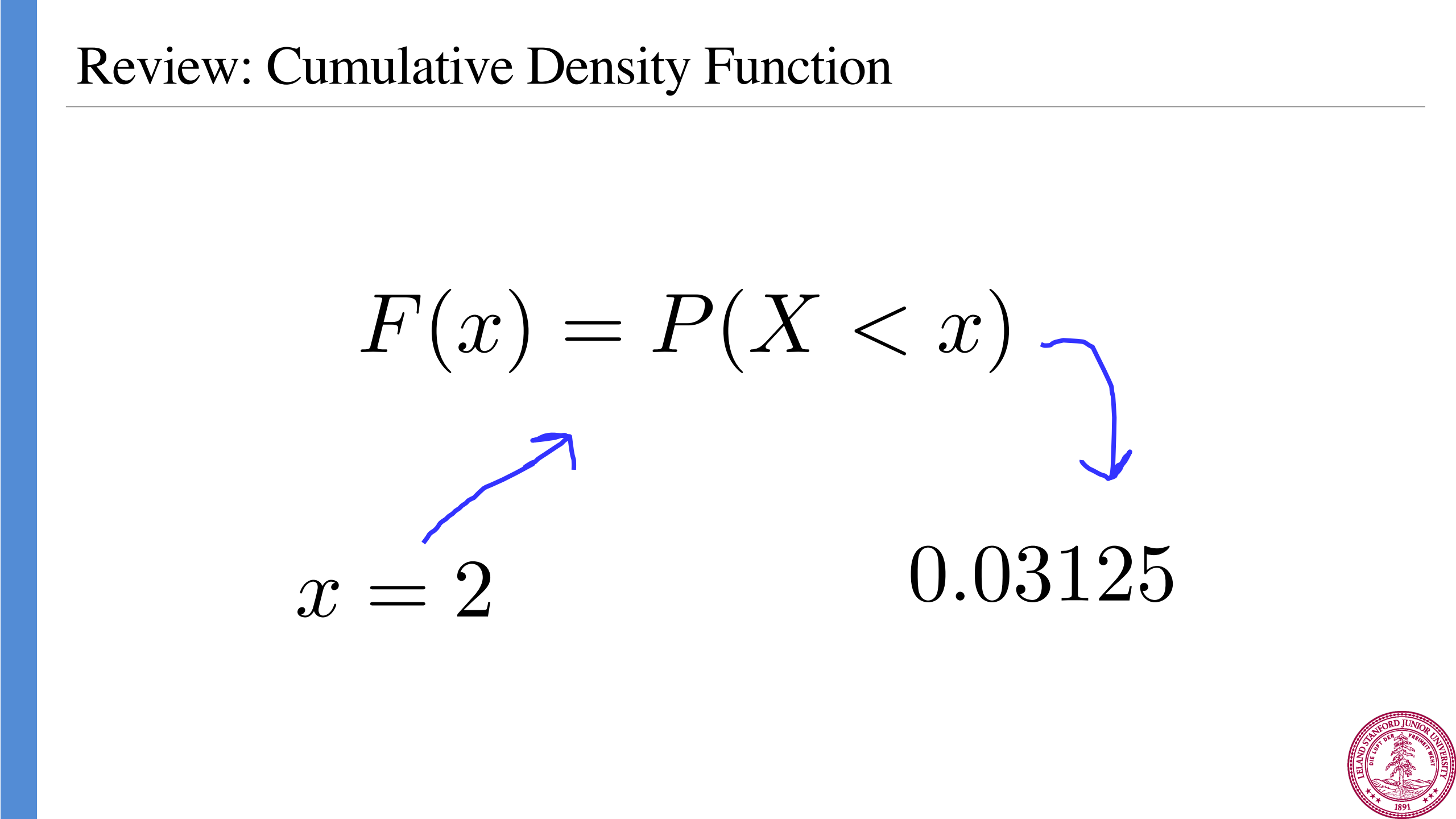
Is there a way to avoid integrals?

Review: Cumulative Density Function

$$F(x) = P(X < x)$$

$x = 2$

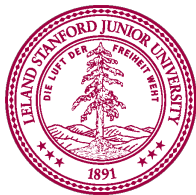
0.03125



Cumulative Density Function

$$F(x) = P(X < x)$$

Probability Query	Solution	Explanation
$\Pr(X < a)$	$F(a)$	That is the definition of the CDF
$\Pr(X \leq a)$	$F(a)$	Trick question. $\Pr(X = a) = 0$
$\Pr(X > a)$	$1 - F(a)$	$\Pr(X < a) + \Pr(X > a) = 1$
$\Pr(a < X < b)$	$F(b) - F(a)$	$F(a) + \Pr(a < X < b) = F(b)$



Probability of Earthquake in Next 4 Years?

Based on historical data, earthquakes of magnitude 8.0+ happen at a **rate of 0.002** per year*. What is the probability of **an major earthquake in the next 4 years?**

Y = Years until the next earthquake of magnitude 8.0+

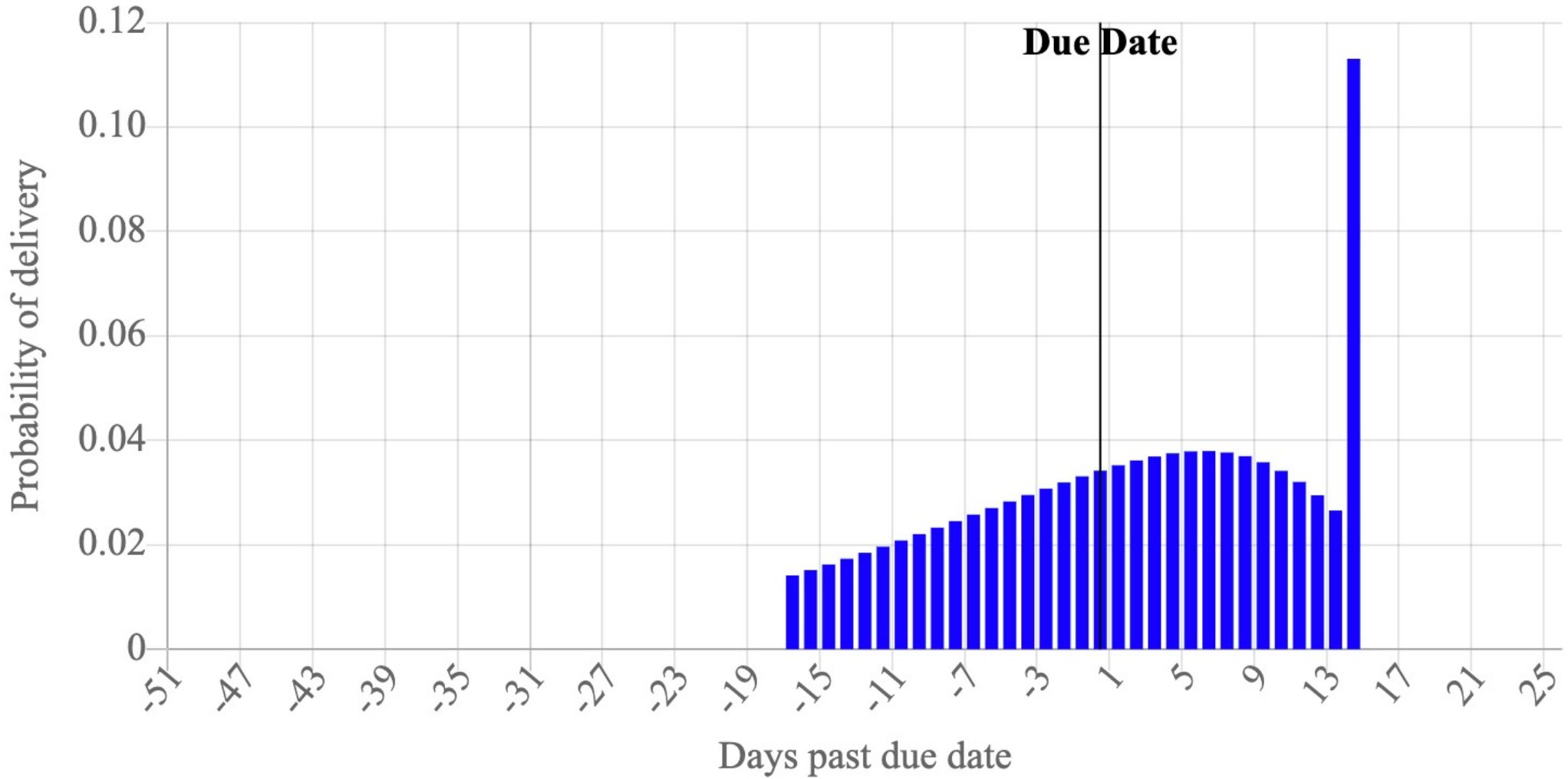
$$Y \sim \text{Exp}(\lambda = 0.002) \qquad F(y) = 1 - e^{-0.002y}$$

$$P(Y < 4) = F(4)$$
$$= 1 - e^{-0.002 \cdot 4}$$
$$\approx 0.008$$

Be ready!

*According to USGS, 2015

[For Fun] How to Represent Baby Delivery Day?



/Review

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 $$E[X] = \mu$$

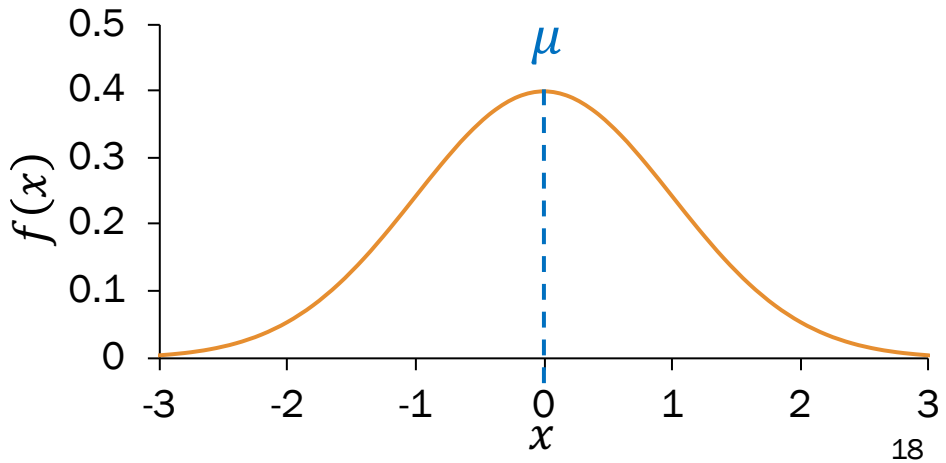
Variance $$\text{Var}(X) = \sigma^2$$

Other names: **Gaussian** random variable

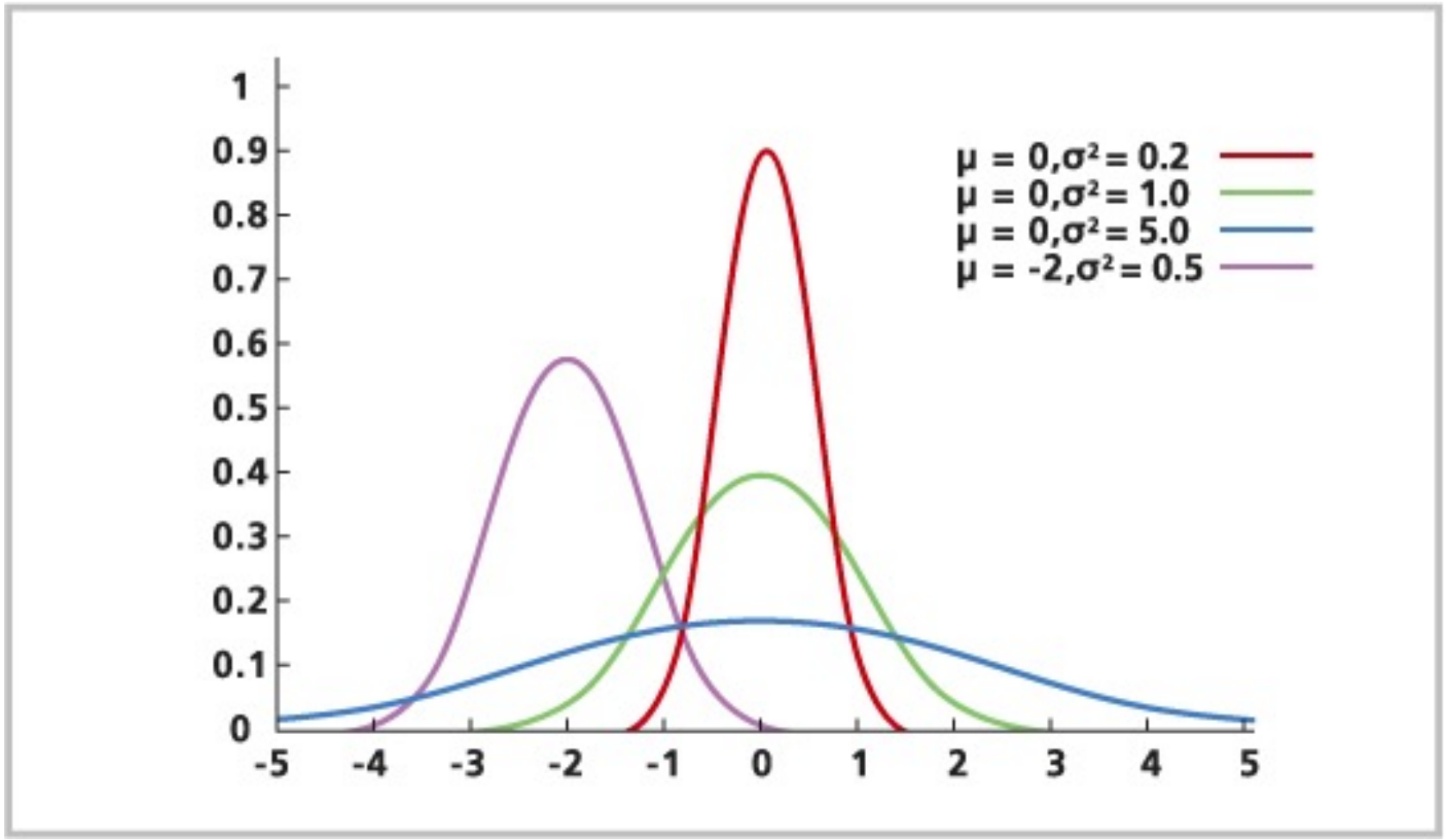
mean

variance

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

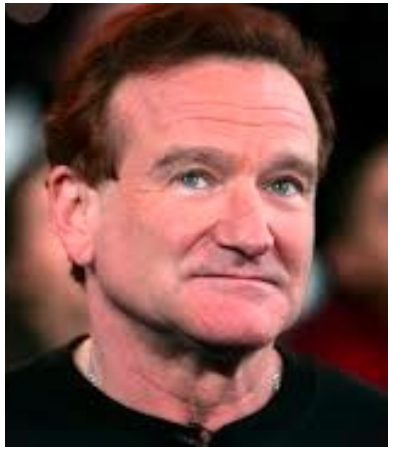


NormCore: A Few Normal Examples



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?

These are log-normal

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

Most noise is assumed normal

- Most noise in the world is Normal

- Often results from the sum of many random variables

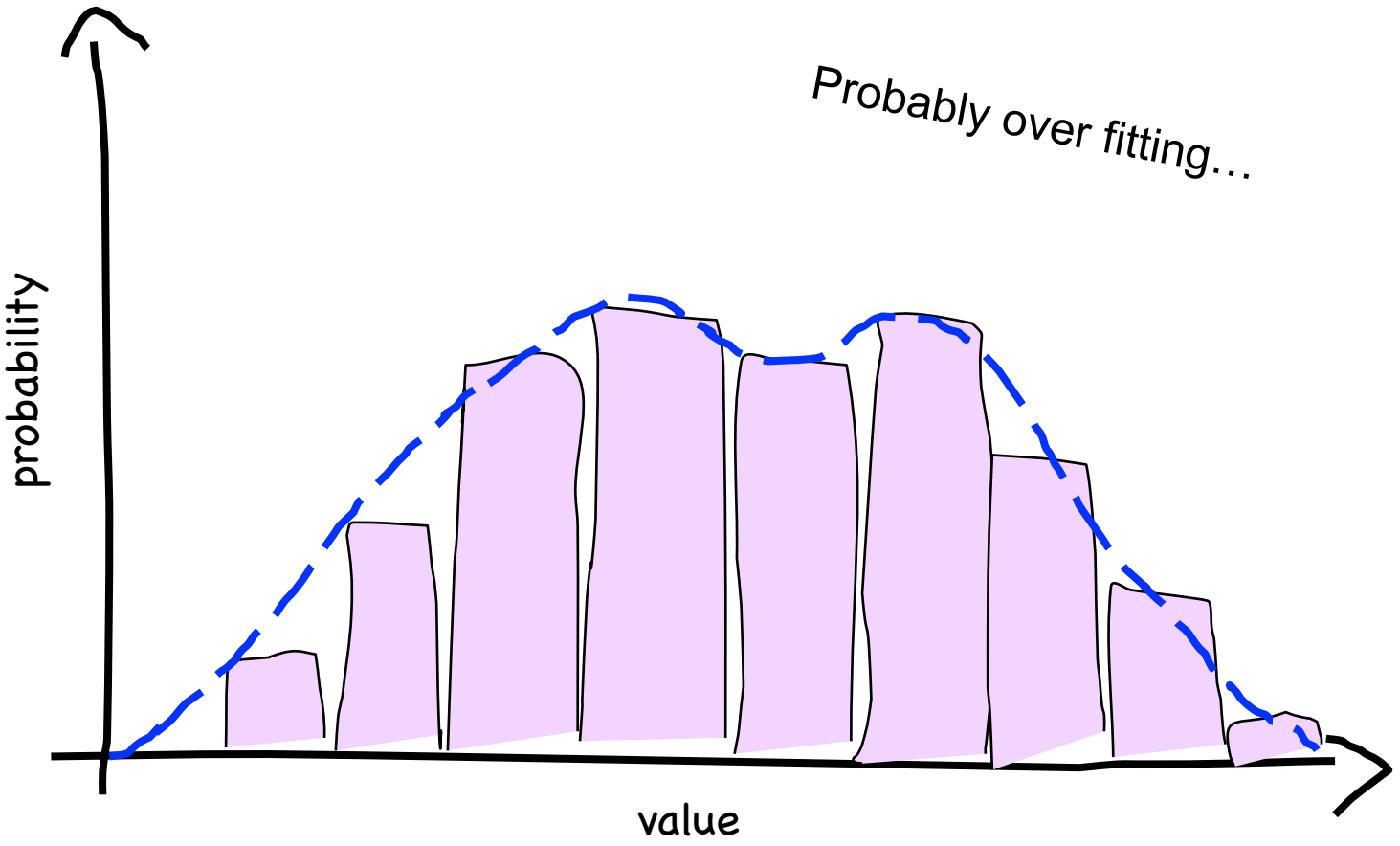
Only if they are equally weighted and independent

- Sample means are distributed normally

That's what they want you to believe...

"The simplest explanation is usually the best one"

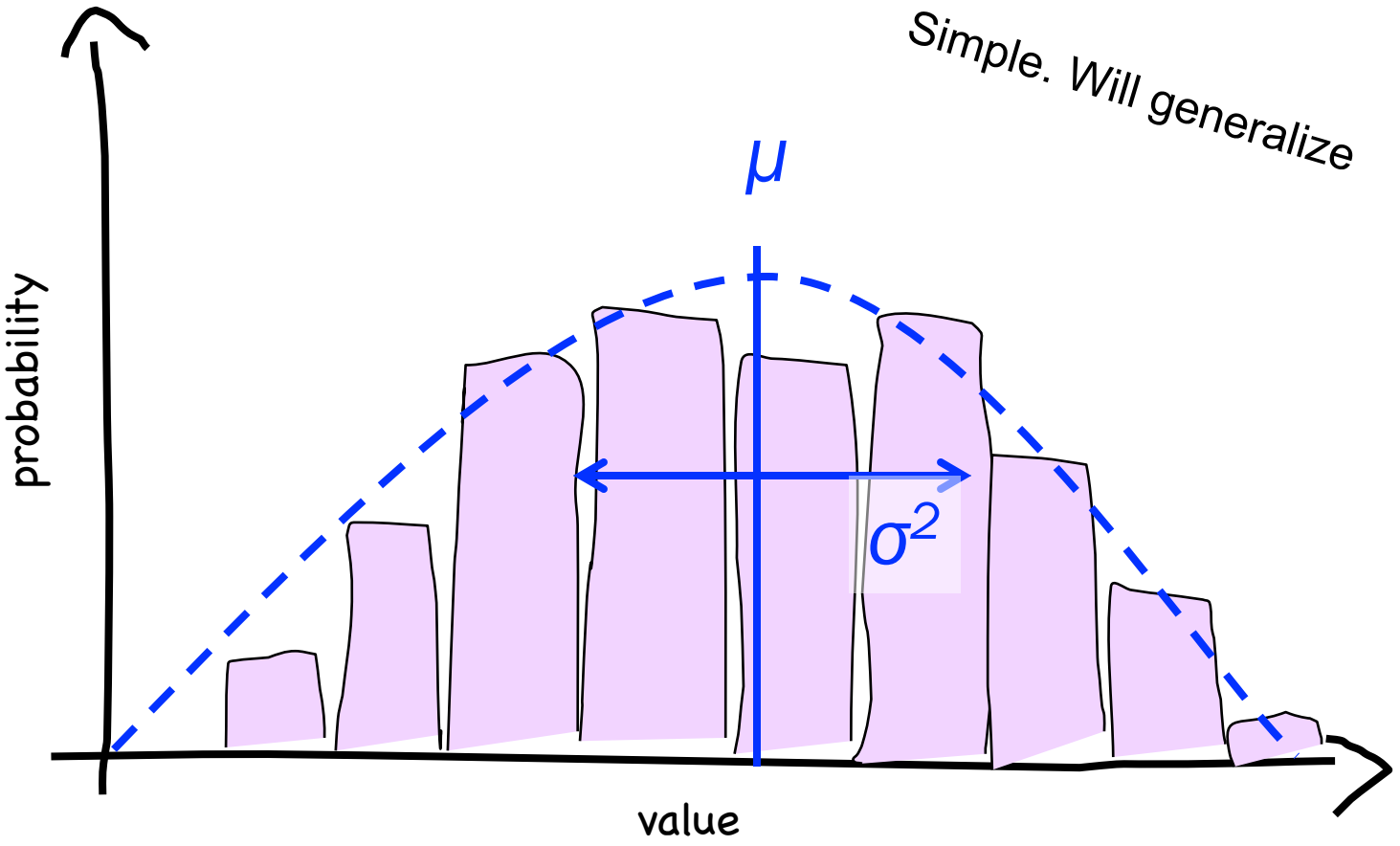
Complexity is Tempting



* That describes the training data, but will it **generalize**?

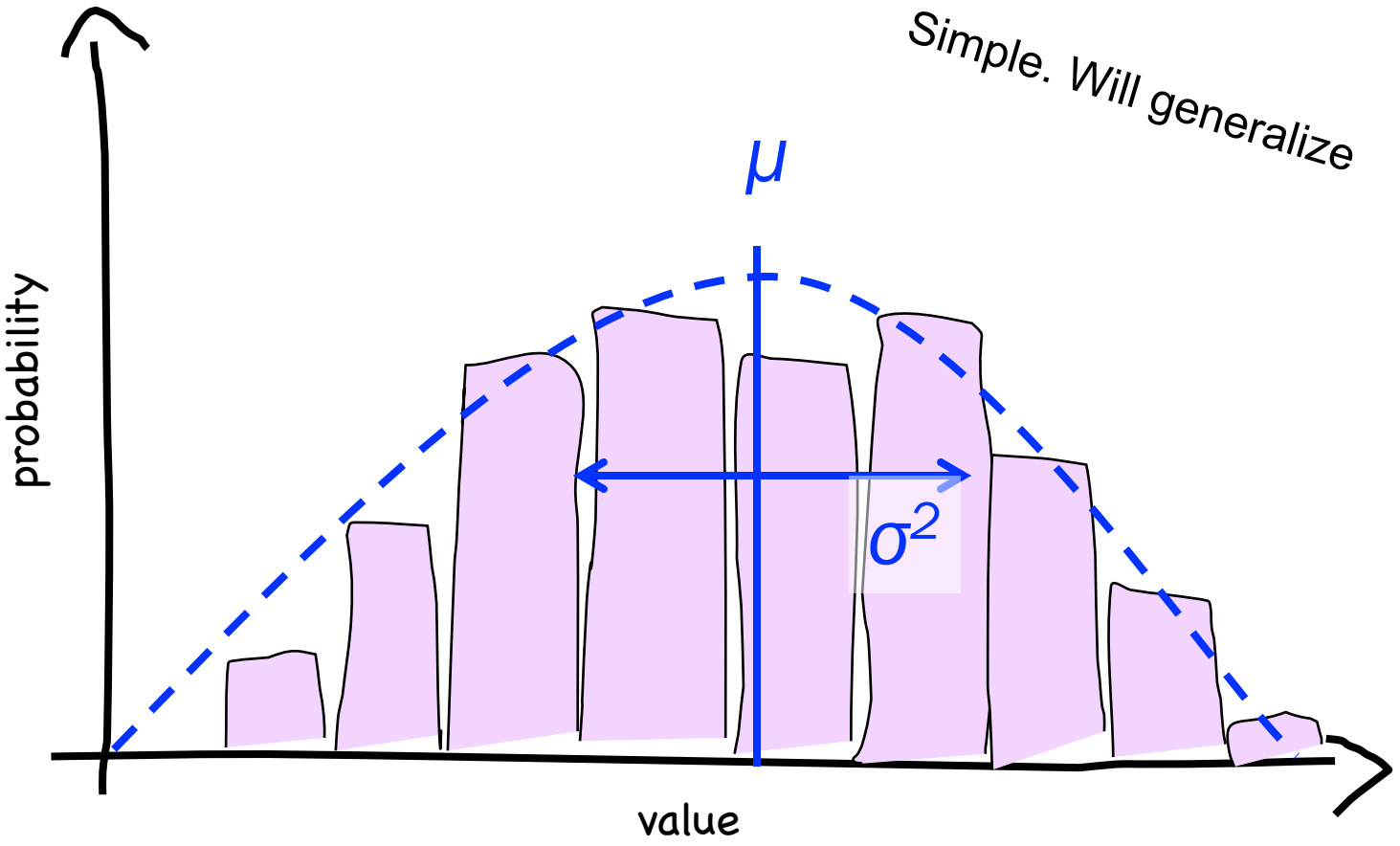
Fewest Assumptions

$$\mathrm{H}(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log \mathrm{P}(x_i)$$



* A Gaussian **maximizes entropy** for a given mean and variance

Fewest Assumptions



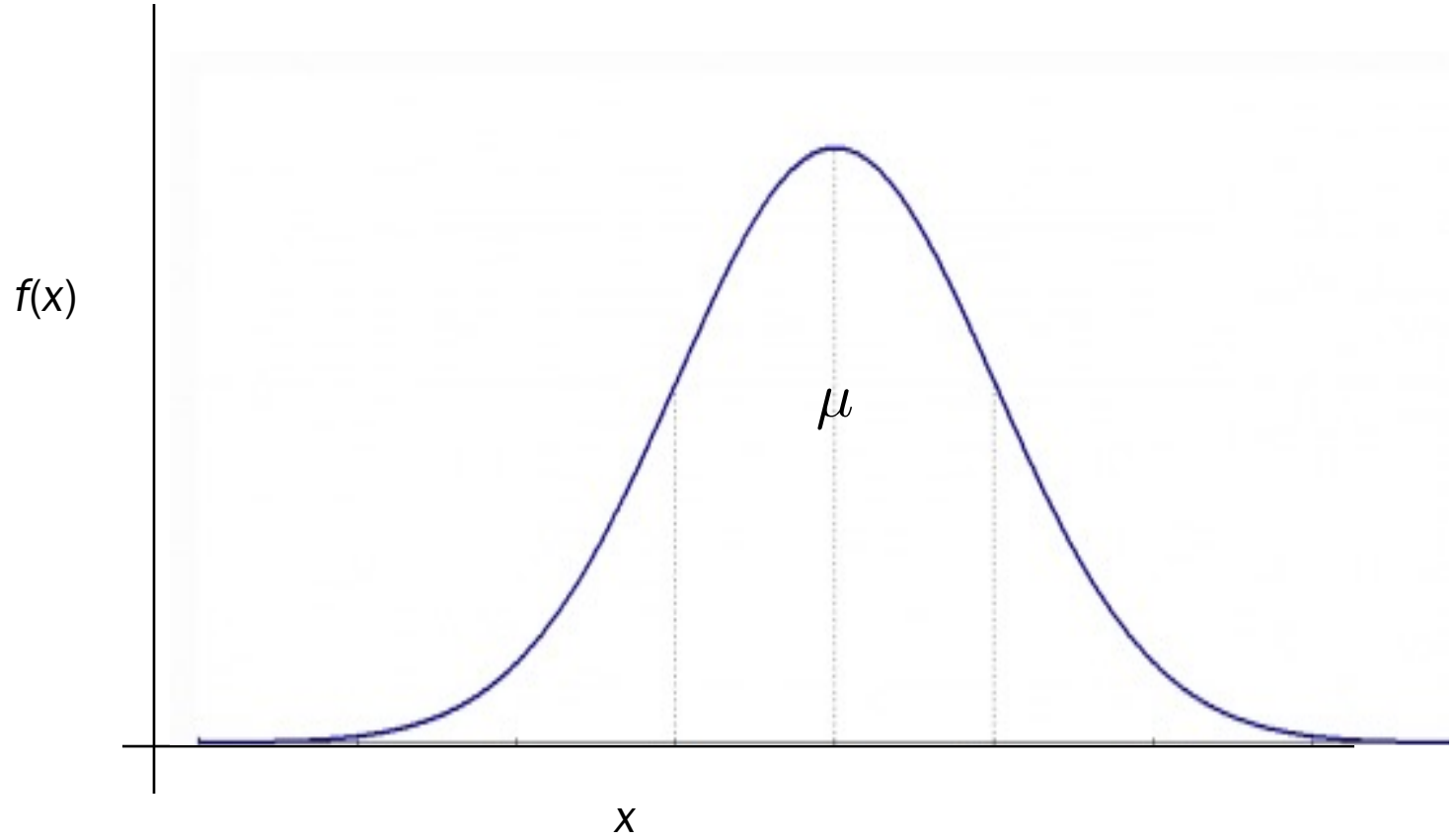
* A Gaussian makes the **fewest assumptions** after matching mean and variance.

But also, it is easy to use

Normal Probability Density Function

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

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



Anatomy of a Beautiful Equation

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

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

"exponential"

the distance to the mean

probability density at *x*

a constant

sigma shows up twice

Campus bikes

You spend some minutes, X, traveling between classes.

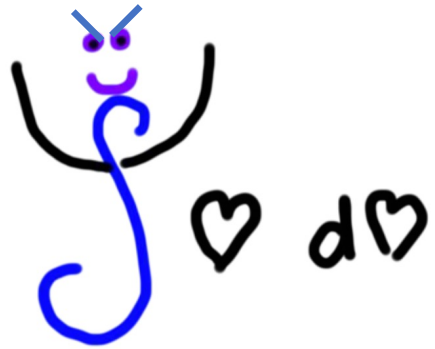
- Average time spent: $\mu = 4$ minutes
- Variance of time spent: $\sigma^2 = 2$ minutes^2

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

No closed form for the integral

No closed form for F(x)

Numerically Solved CDF

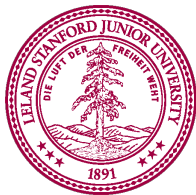
$$\mathcal{N}(\mu, \sigma^2)$$

A function that has been solved for numerically

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

The cumulative density function of any normal

* We are going to spend the next few slides getting here



Linear Transform of Normal is Normal

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

If $Y = aX + b$ then *Y* is also Normal

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

$$Var(Y) = Var(aX + b) = a^2 Var(X) = a^2\sigma^2$$

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

Special Linear Transform

If $Y = aX + b$ then *Y* is also Normal

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

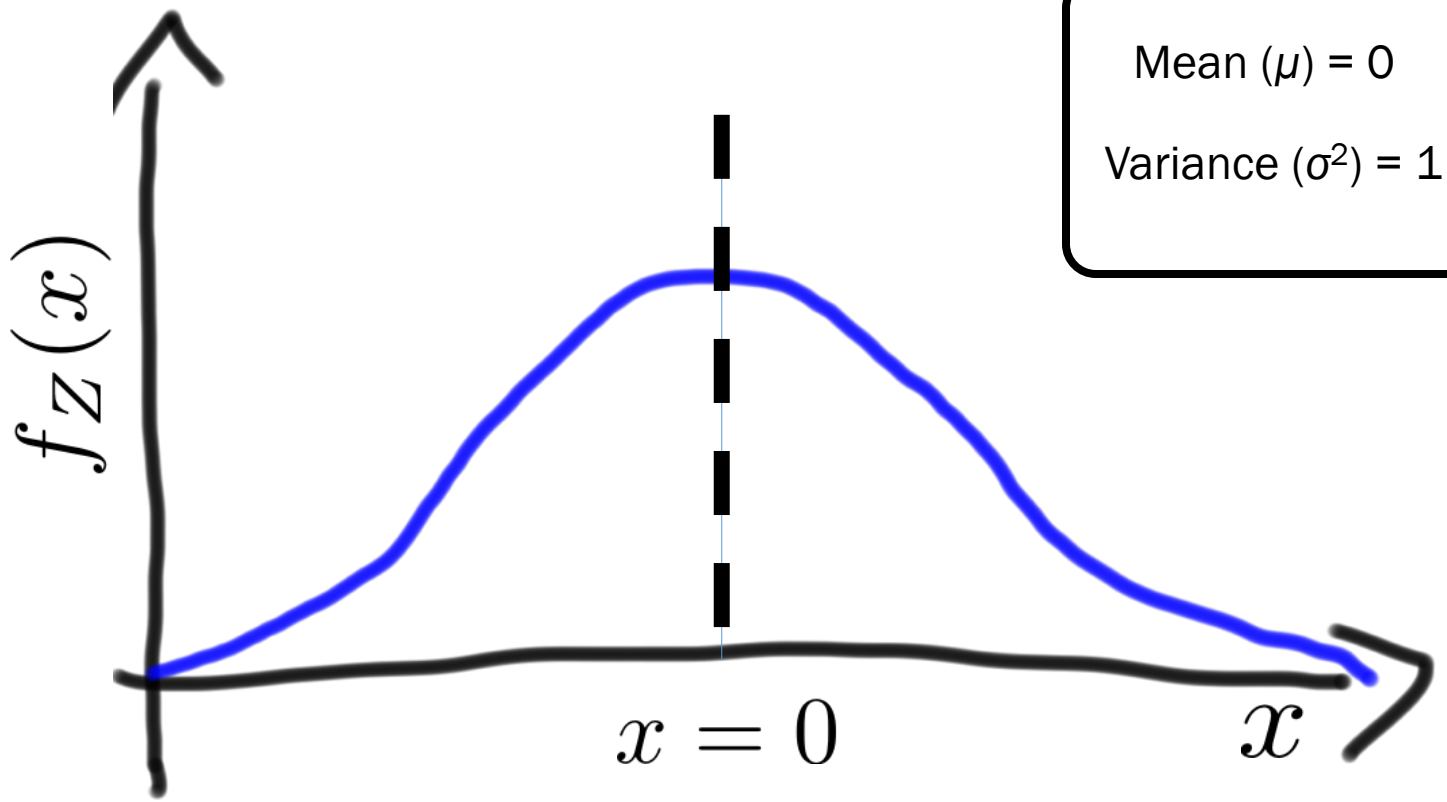
There is a special case of linear transform for any *X*:

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \qquad a = \frac{1}{\sigma} \qquad b = -\frac{\mu}{\sigma}$$

$$\begin{aligned} Z &\sim \mathcal{N}(a\mu + b, a^2\sigma^2) \\ &\sim \mathcal{N}(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}) \\ &\sim \mathcal{N}(0, 1) \end{aligned}$$

The Standard Normal

$$Z \sim N(\mu = 0, \sigma^2 = 1)$$



Mean (μ) = 0

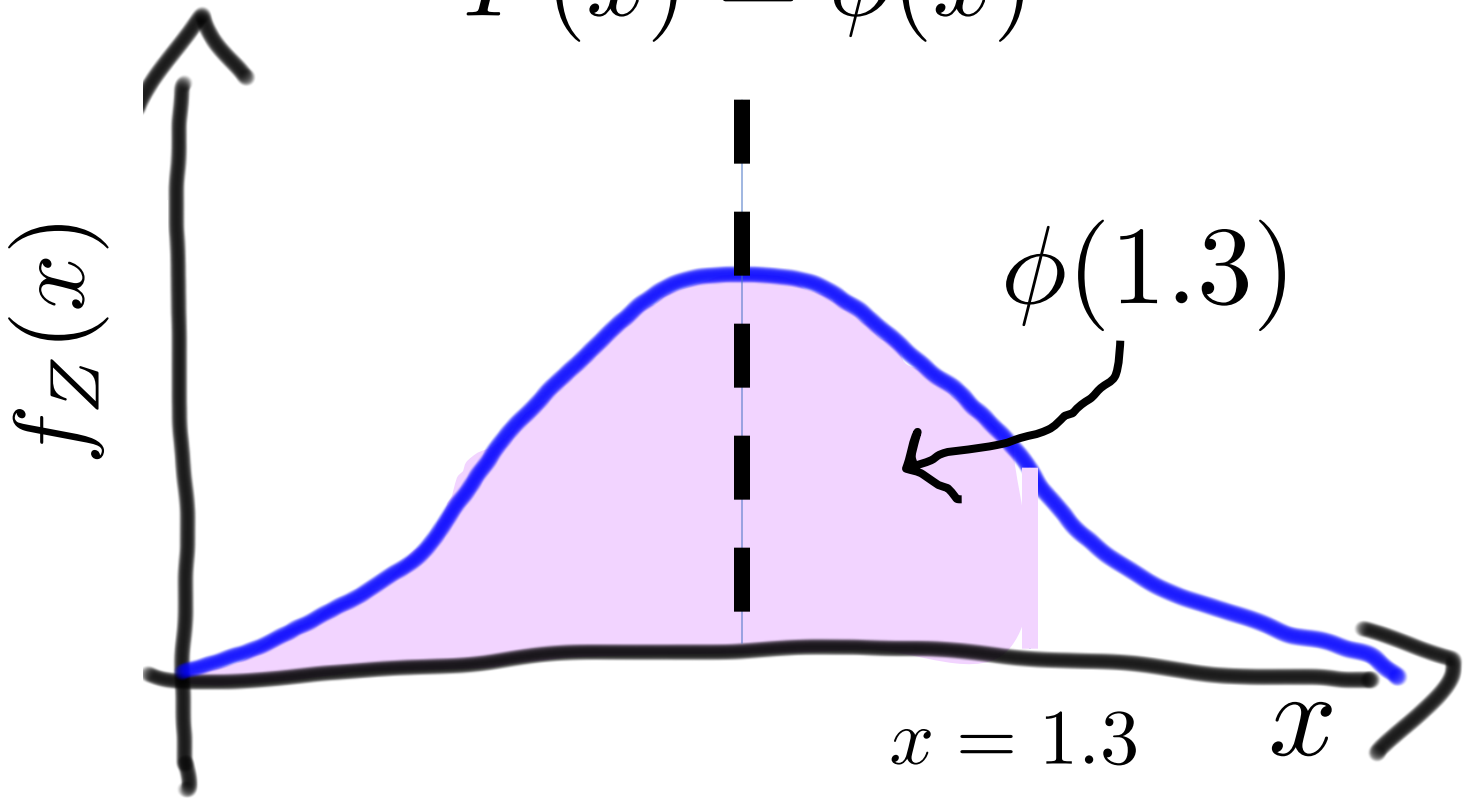
Variance (σ^2) = 1

*This is the probability density function for the standard normal

Phi

$$Z \sim N(\mu = 0, \sigma^2 = 1)$$

$$F(x) = \phi(x)$$

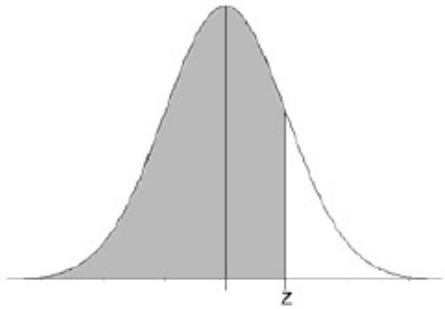


*This is the probability density function for the standard normal

Using Table of ϕ

Standard Normal Cumulative Probability Table

$$\Phi(1.31) = 0.7054$$

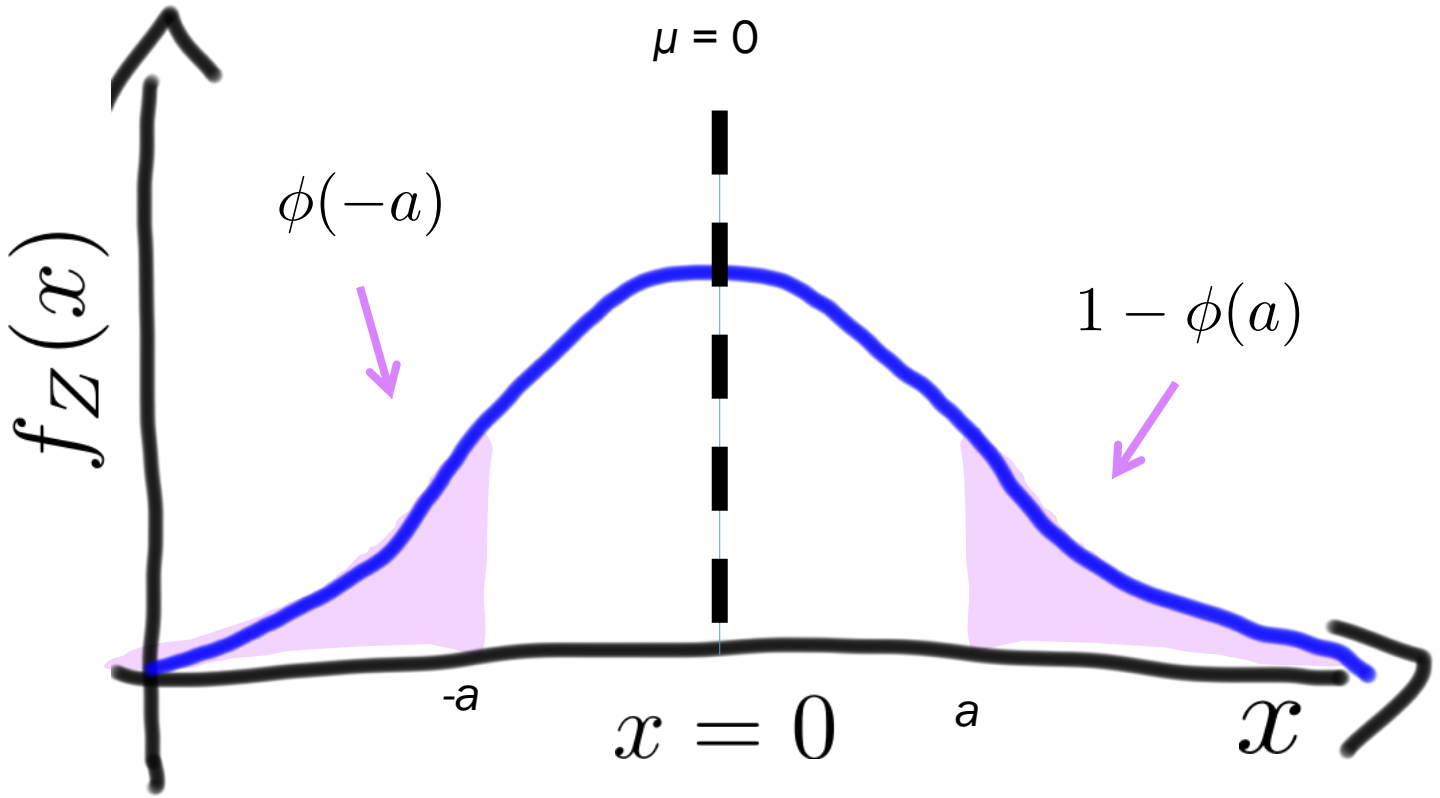


Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

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
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
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	0.8907	0.8925	0.8944	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

Symmetry of Phi

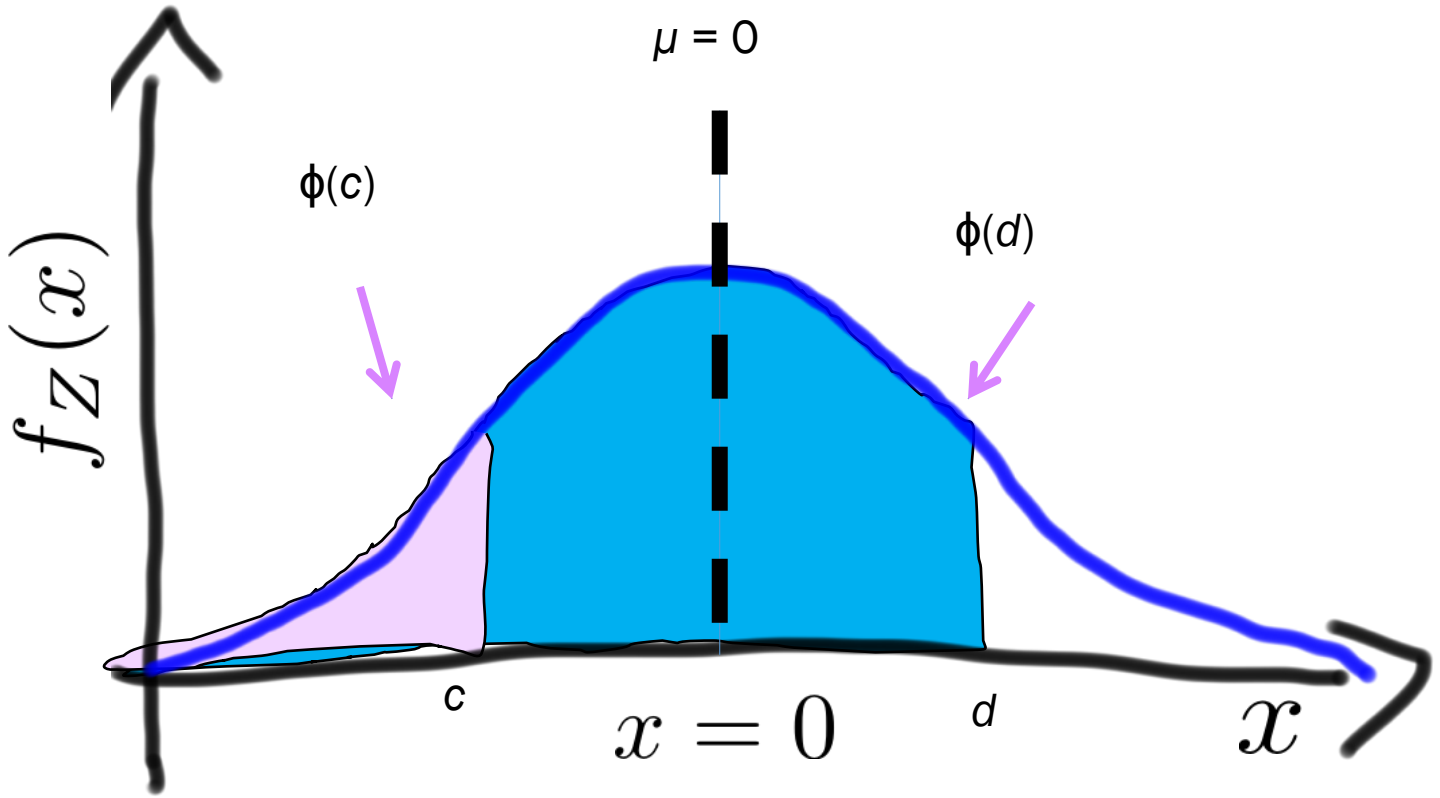
$$\phi(a) = 1 - \phi(a)$$



*This is the probability density function for the standard normal

Interval of Phi

$$P(c < Z < d) = \phi(d) - \phi(c)$$



Compute F(x) via Transform

$$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2) \qquad Z = \frac{X - \mu}{\sigma}$$

Use *Z* to compute *F*(x)

$$F_X(x) = P(X \leq x)$$

$$= P(X - \mu \leq x - \mu)$$

$$= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right)$$

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

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

For normal distribution, F(x) is computed using the phi transform.

And here we are

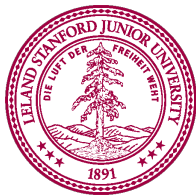
$\mathcal{N}(\mu, \sigma^2)$

CDF of Standard Normal: A function that has been solved for numerically

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

The cumulative density function (CDF) of any normal

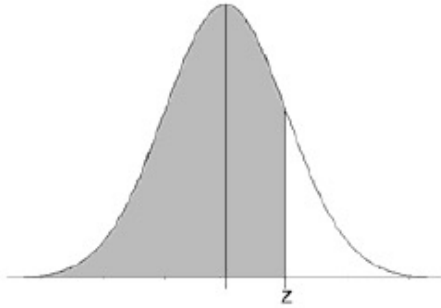
Table of $\Phi(Z)$ values in textbook, p. 201 and handout



Using the Phi Table

Standard Normal Cumulative Probability Table

$\Phi(0.54) = 0.7054$



Cumulative probabilities for POSITIVE z-values are shown in the following table:

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
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
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	0.8907	0.8925	0.8944	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

Table is kinda old school

Using Programming Library

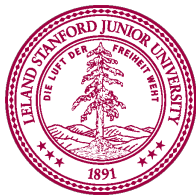
Every modern programming language has a normal library

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

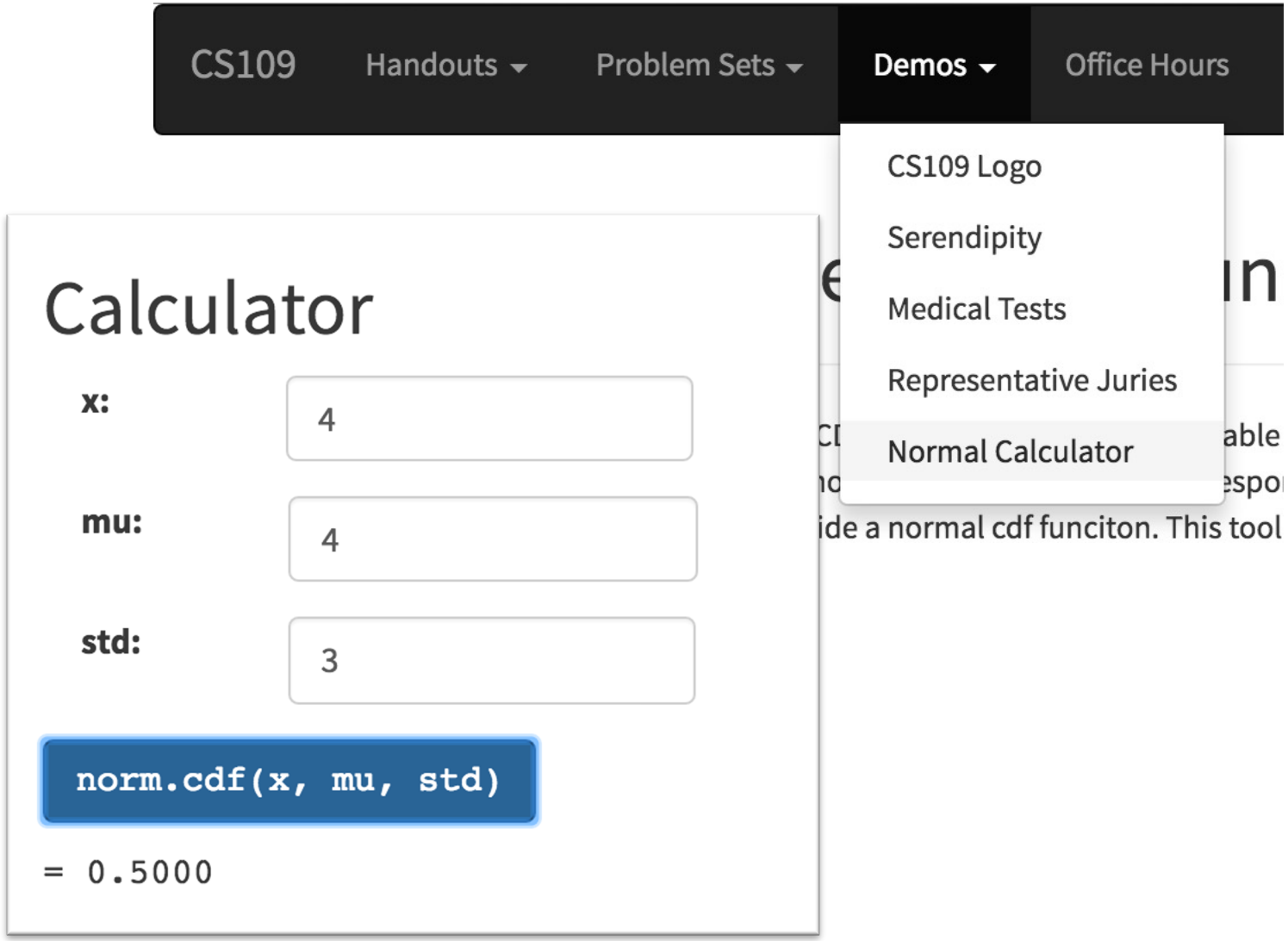
$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

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

* This is from Python's scipy library



I made one for you



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

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

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

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

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

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?

1. $P(X > 0)$ (we just did this)
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 $\Phi(-z) = 1 - \Phi(z)$

Get your Gaussian On

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

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

- 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

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

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

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

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

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

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

$$\approx 0.1337$$

Imagine you are taking a quiz…
With no computer!!!

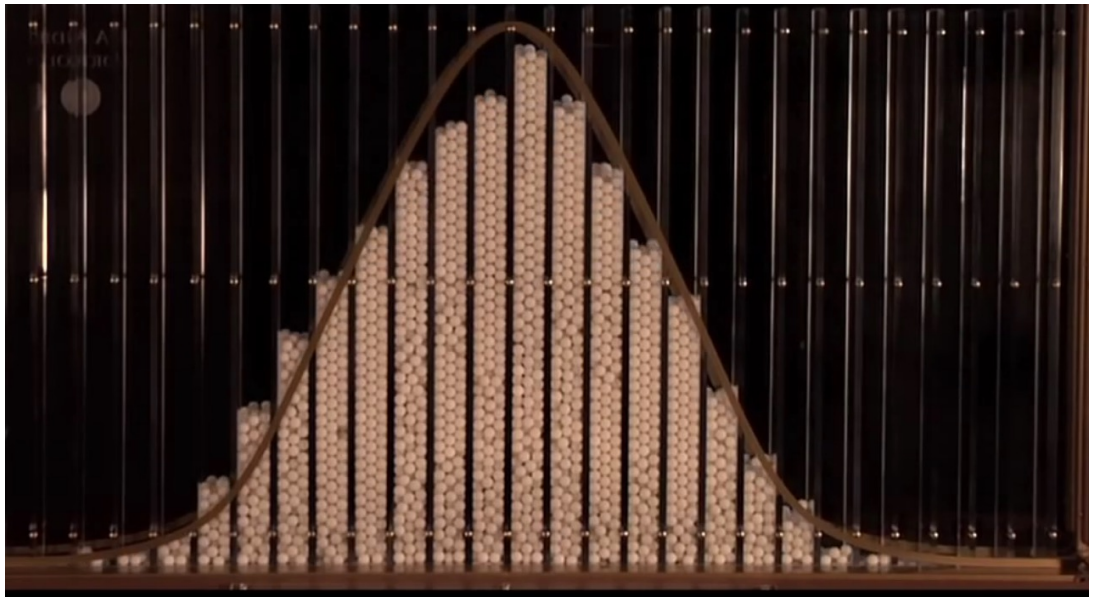
Website Testing

100 people are given a new website design

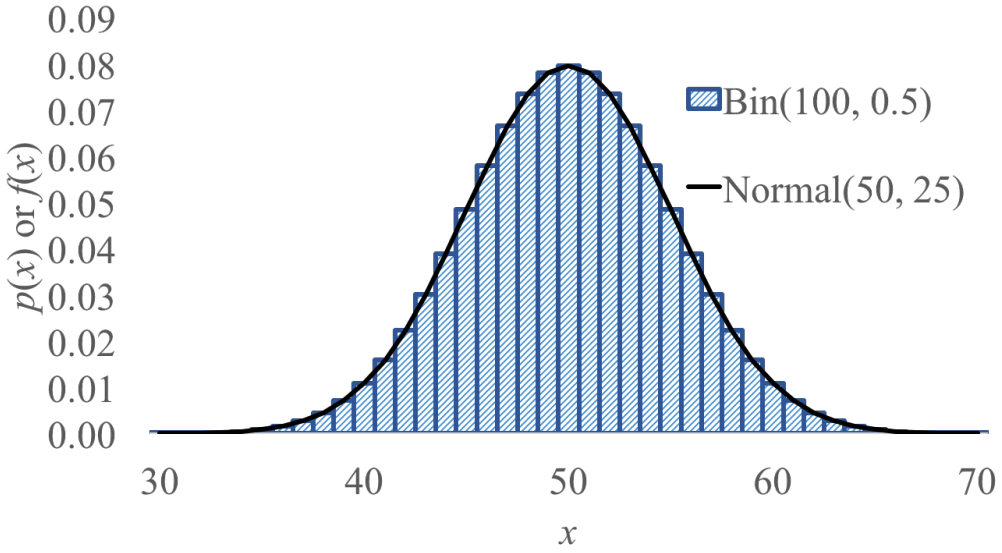
- X = # people whose time on site increases
- CEO will endorse new design if X ≥ 65 What is P(CEO endorses change| it has no effect)?
- X ~ Bin(100, 0.5). Want to calculate P(X ≥ 65)
- Give a numerical answer…

$$P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} (0.5)^i (1 - 0.5)^{100-i}$$

Don’t worry, Normal approximates Binomial



Galton Board



(We’ll explain ***why*** in 2 weeks’ time)

Website testing

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

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

Approach 1: Binomial

Define

$X \sim \text{Bin}(n = 100, p = 0.5)$

Want: $P(X \geq 65)$

Solve

$P(X \geq 65) \approx 0.0018$

Approach 2: approximate with Normal

Define

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

$\mu = np = 50$

$\sigma^2 = np(1-p) = 25$

$\sigma = \sqrt{25} = 5$

Solve

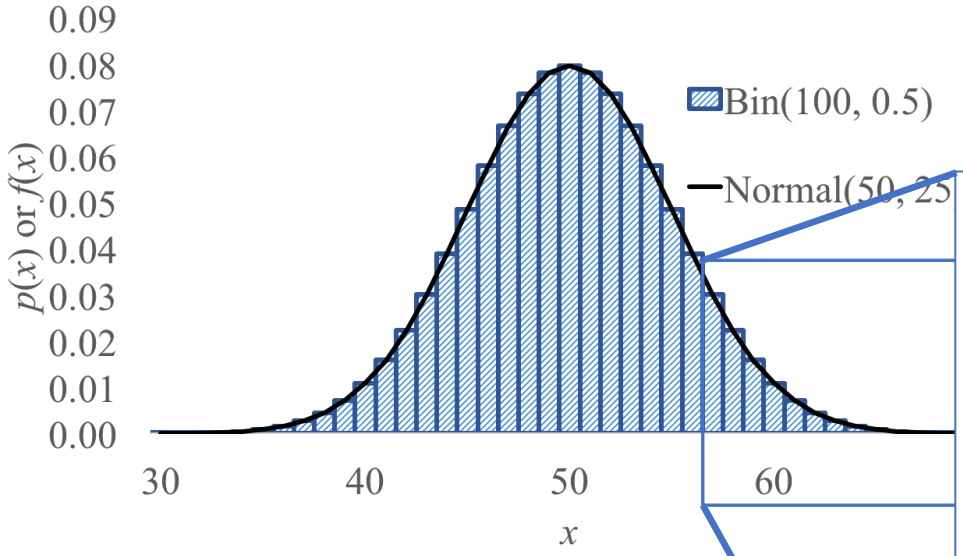
$P(X \geq 65) \approx P(Y \geq 65) = 1 - F_Y(65)$ 😐

$= 1 - \Phi\left(\frac{65-50}{5}\right) = 1 - \Phi(3) \approx 0.0013$?

⚠️⚠️ (this approach is missing something important)

Website testing (with continuity correction)

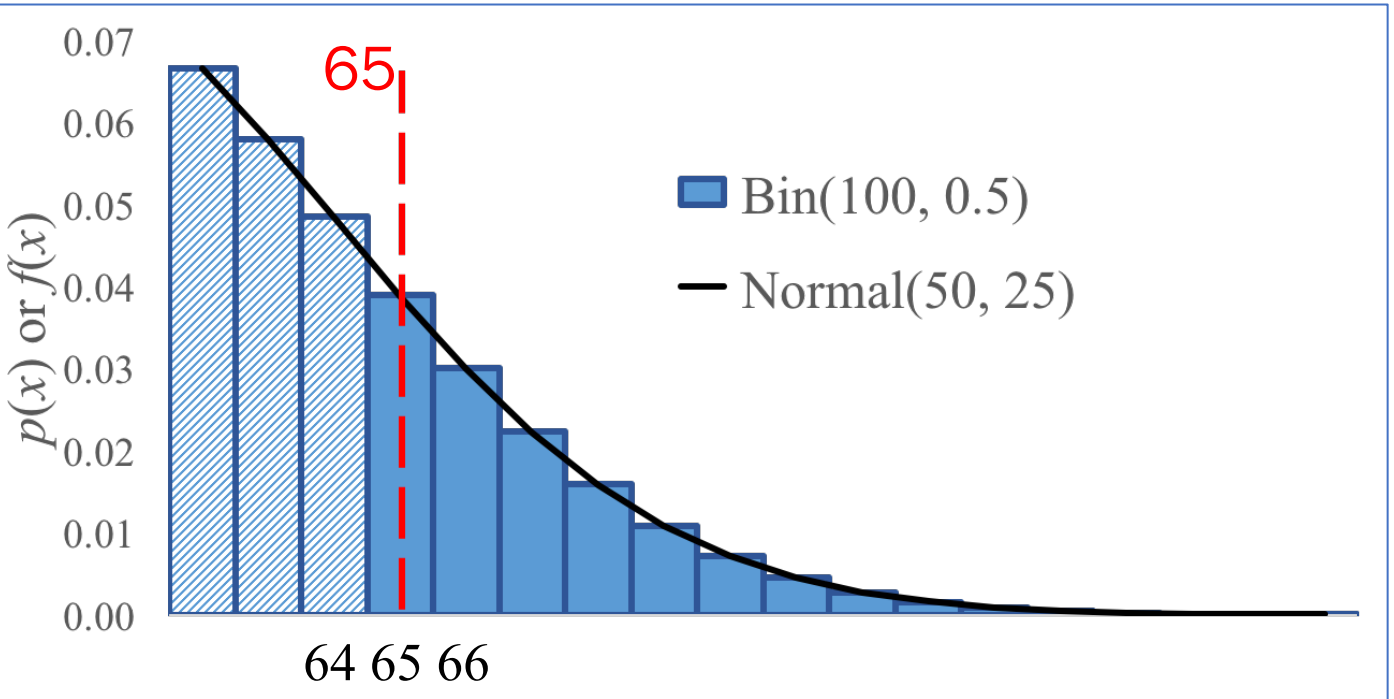
In our website testing, $Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$.



$P(X \geq 65)$ Binomial

$\approx P(Y \geq 64.5)$ Normal

≈ 0.0018 ✅ the better Approach 2

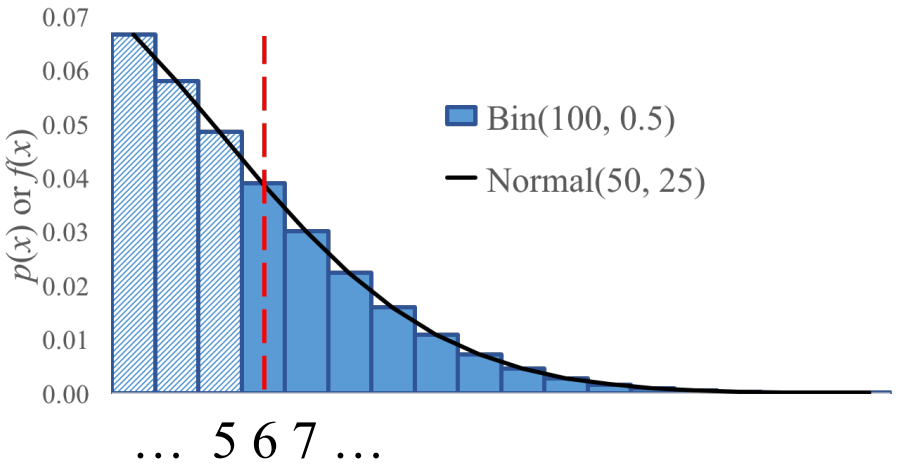


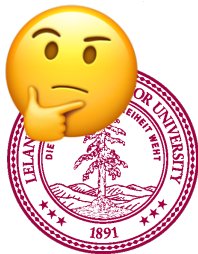
You must perform a **continuity correction** when approximating a Binomial RV with a Normal RV.

Continuity correction

If $Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

Discrete (e.g., Binomial) probability question		Continuous (Normal) probability question
$P(X = 6)$		
$P(X \geq 6)$		
$P(X > 6)$		
$P(X < 6)$		
$P(X \leq 6)$		

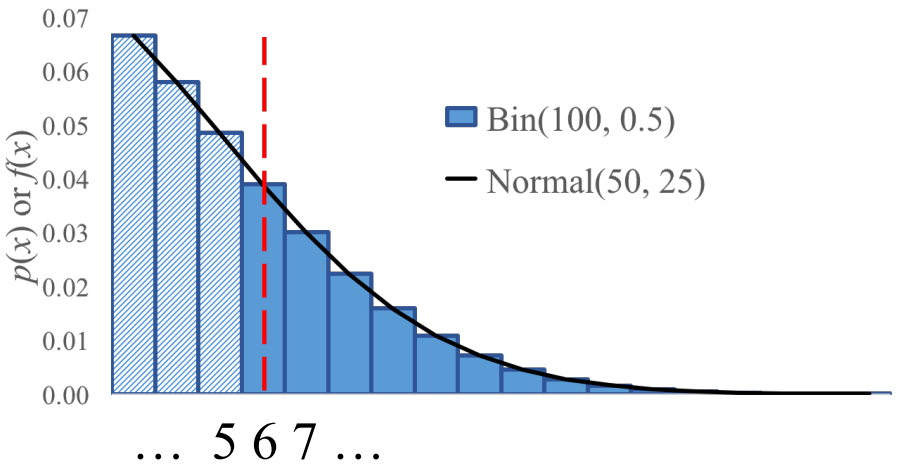




Continuity correction

If $Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	$P(5.5 \leq Y \leq 6.5)$
$P(X \geq 6)$	$P(Y \geq 5.5)$
$P(X > 6)$	$P(Y \geq 6.5)$
$P(X < 6)$	$P(Y \leq 5.5)$
$P(X \leq 6)$	$P(Y \leq 6.5)$



Normal Approximation

Who gets to approximate?

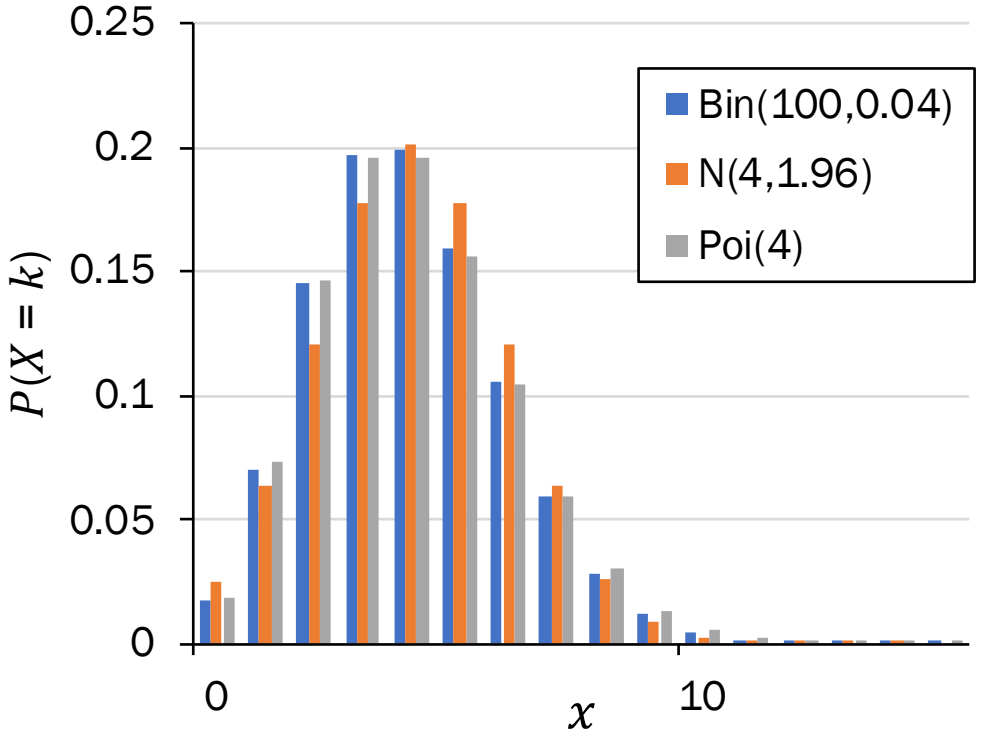
$$X \sim \text{Bin}(n, p)$$
$$E[X] = np$$
$$\text{Var}(X) = np(1-p)$$

$$Y \sim \text{Poi}(\lambda)$$
$$\lambda = np$$

?

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$
$$\mu = np$$
$$\sigma^2 = np(1-p)$$

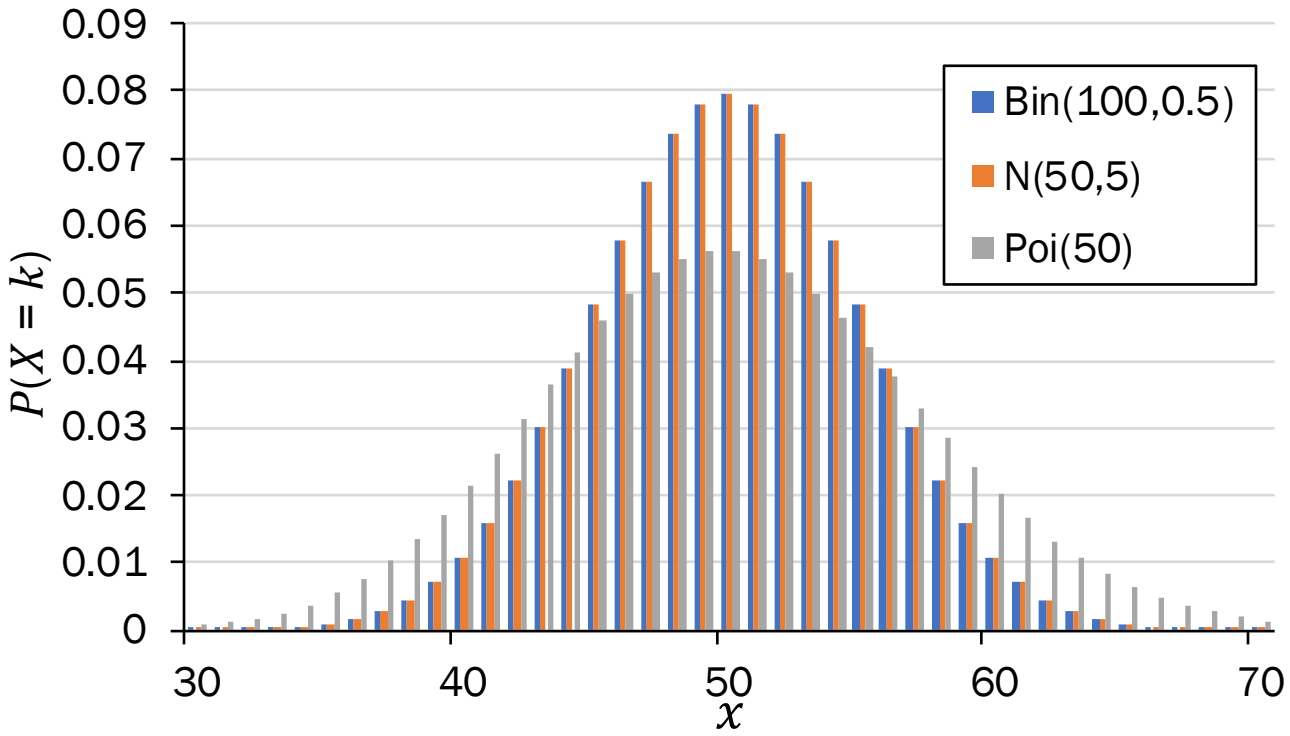
Who gets to approximate?



Poisson approximation

n large (> 20), p small (< 0.05)

slight dependence okay



Normal approximation

n large (> 20), variance large ($np(1-p) > 10$)

independence

1. If there is a choice, either is fine.
2. When using Normal to approximate a discrete RV, use a continuity correction.

Super Question:

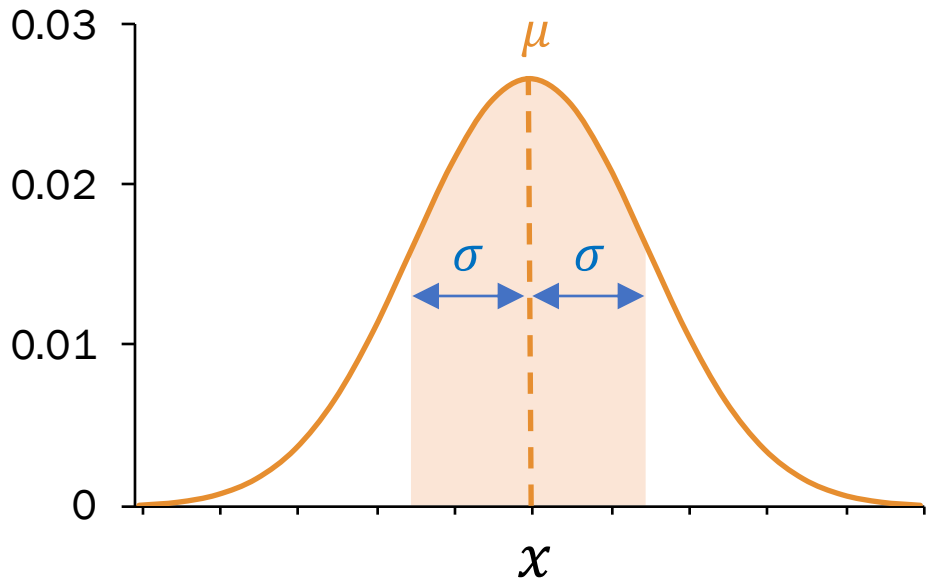
Why Be Normal? 68% rule

You may have heard the statement:

"68% of the class will fall within 1 standard deviation of the exam average."

In general, this is only true of normal distributions:

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



$$P(|X-\mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$$

$$= F(\mu+\sigma) - F(\mu-\sigma)$$

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

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

$$= 2\Phi(1) - 1 \quad \approx 2(0.8413) - 1 = 0.6826$$

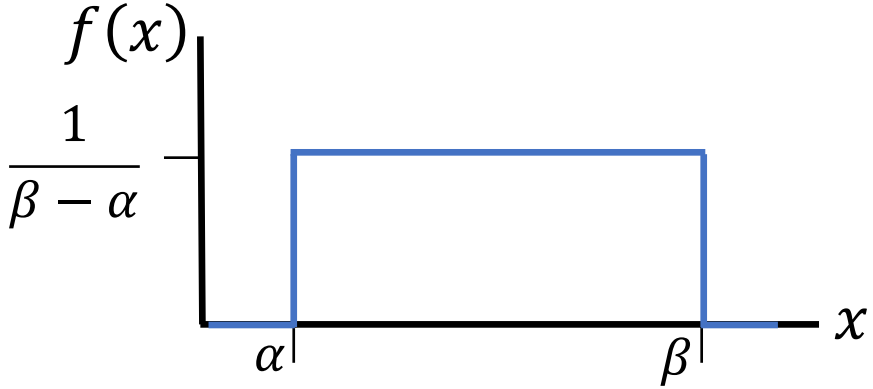
Why Be Normal? 68% rule

You may have heard the statement:

"68% of the class will fall within 1 standard deviation of the exam average."

In general, this is only true of normal distributions:

Counterexample: Let $X \sim \text{Uni}(\alpha, \beta)$.



$$\mu = E[X] = \frac{\alpha + \beta}{2}$$

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12} \Rightarrow \sigma = \text{SD}(X) = \frac{\beta - \alpha}{\sqrt{12}}$$

$$P(|X - \mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma)$$

$$= \frac{1}{\beta - \alpha} \cdot [(\mu + \sigma) - (\mu - \sigma)]$$

$$= \frac{1}{\beta - \alpha}[2\sigma] = \frac{1}{\beta - \alpha} \cdot \left[2 \cdot \frac{\beta - \alpha}{\sqrt{12}}\right]$$

$$= 2/\sqrt{12} \approx 0.58$$

Gaussian Sampling and ELO ratings

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

Gaussian Sampling and ELO ratings

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

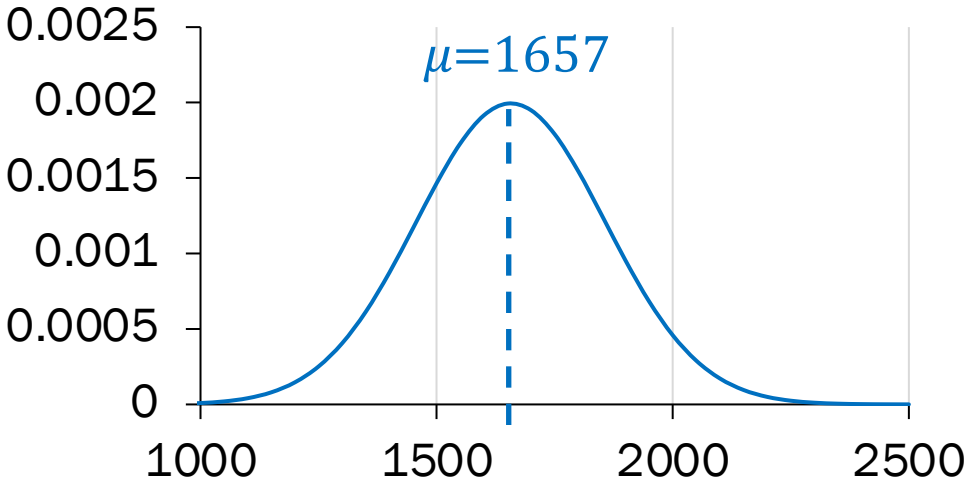
- Each game, a team has ability $A \sim \mathcal{N}(S, 200^2)$.
- The team with the higher sampled ability wins.

What is the probability that Warriors win this game?

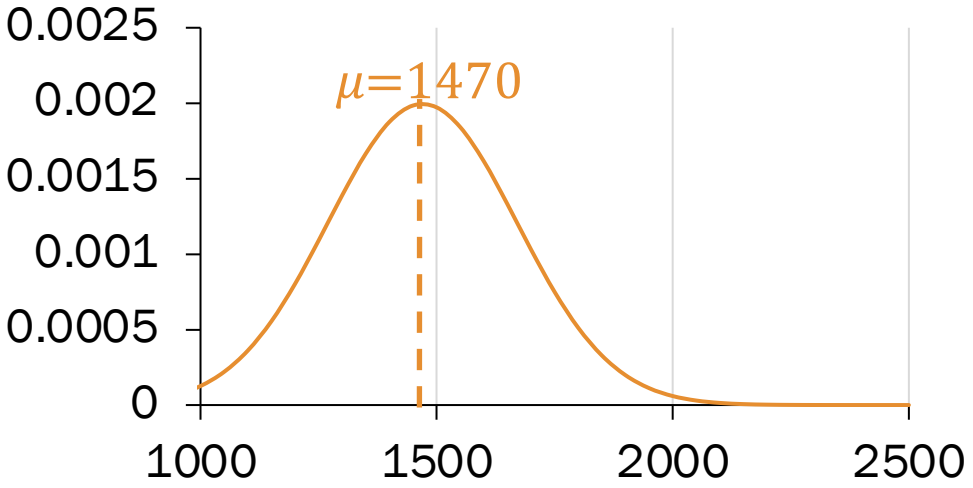
Want: $P(\text{Warriors win}) = P(A_W > A_O)$

Arpad Elo

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



Opponent's $A_O \sim \mathcal{N}(S = 1470, 200^2)$



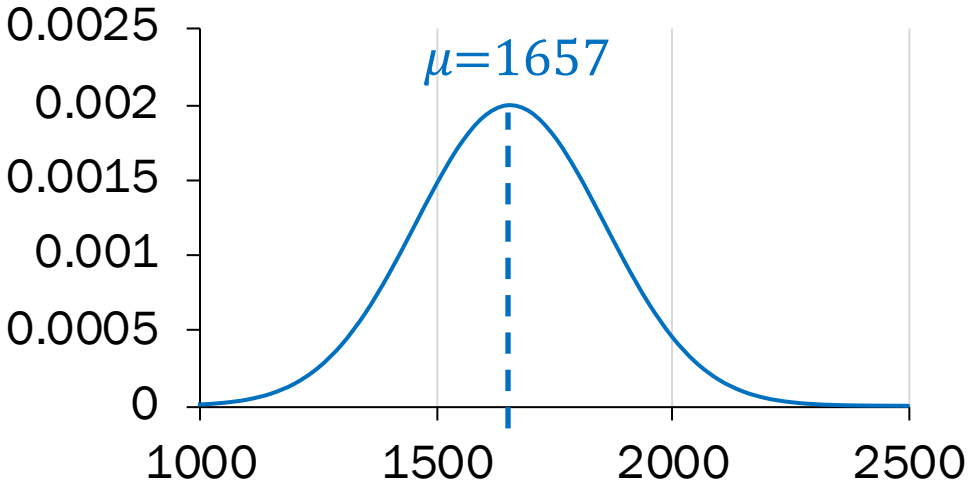
Gaussian Sampling and ELO ratings

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

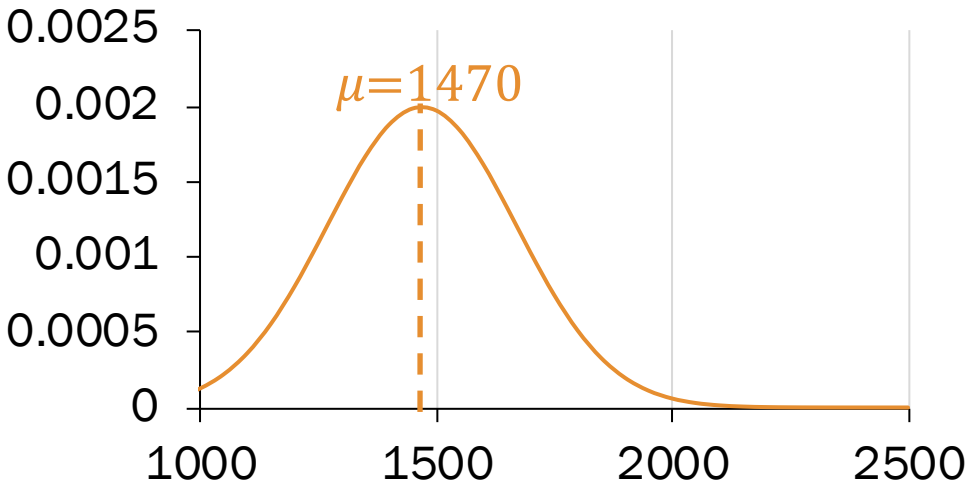
```
from scipy import stats
WARRIORS_ELO = 1657
OPPONENT_ELO = 1470
STDEV = 200
NTRIALS = 10000

nSuccess = 0
for i in range(NTRIALS):
  w = stats.norm.rvs(WARRIORS_ELO, STDEV)
  o = stats.norm.rvs(OPPONENT_ELO, STDEV)
  if w > o:
    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 $A_O \sim \mathcal{N}(S = 1470, 200^2)$



Is there a better way?

$$P(A_W > A_O)$$

- This is a probability of an event involving ***two continuous*** 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!

Anatomy of a Beautiful Equation

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

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

"exponential"

the distance to the mean

probability density at *x*

a constant

sigma shows up twice

Does it look less scary like this?

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

This means "e to the power of" and is common function in code math libraries

$$f(x) \propto \frac{1}{\sigma} \cdot \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

This means "proportional to". There is a constant but there are many cases where we don't care what it is!

What if you had to take the log of this function?

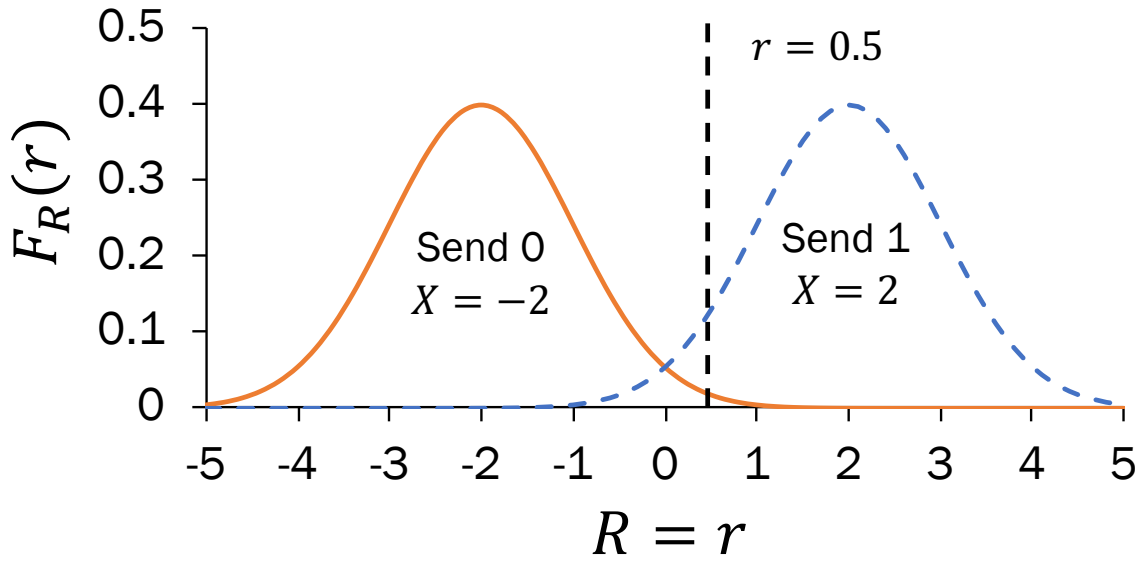
Extra practice problem

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



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)?

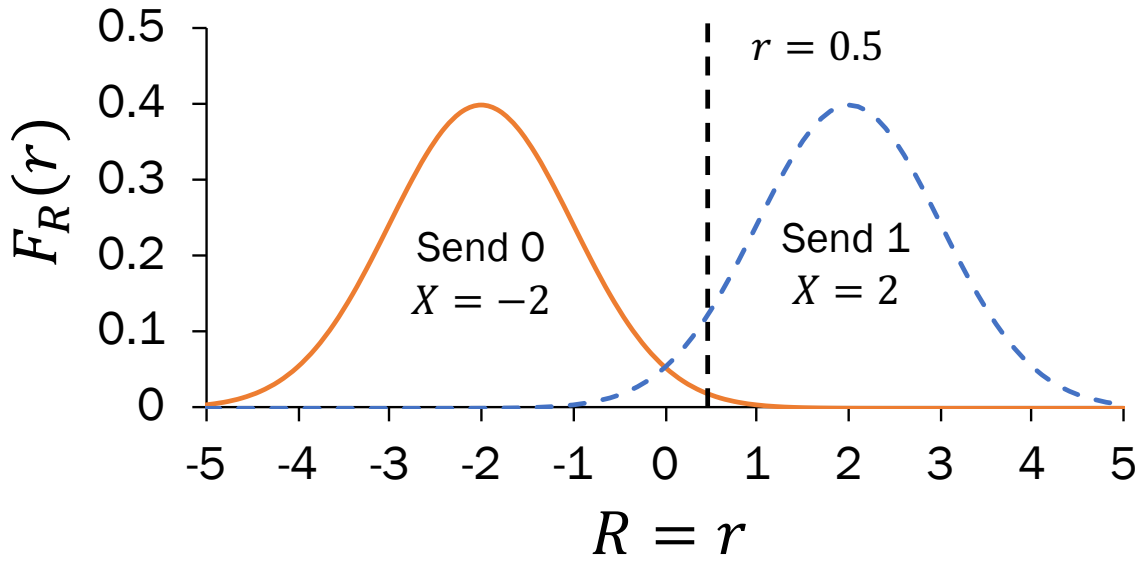
These probabilities are unequal. Why might this be useful?

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



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

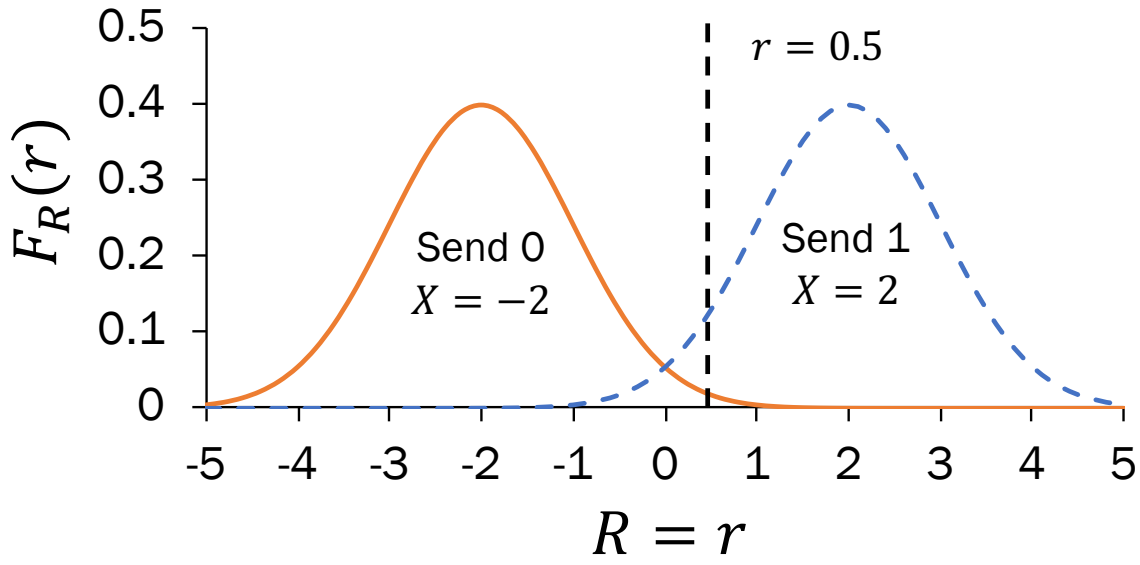
$$= \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$

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



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)?

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

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