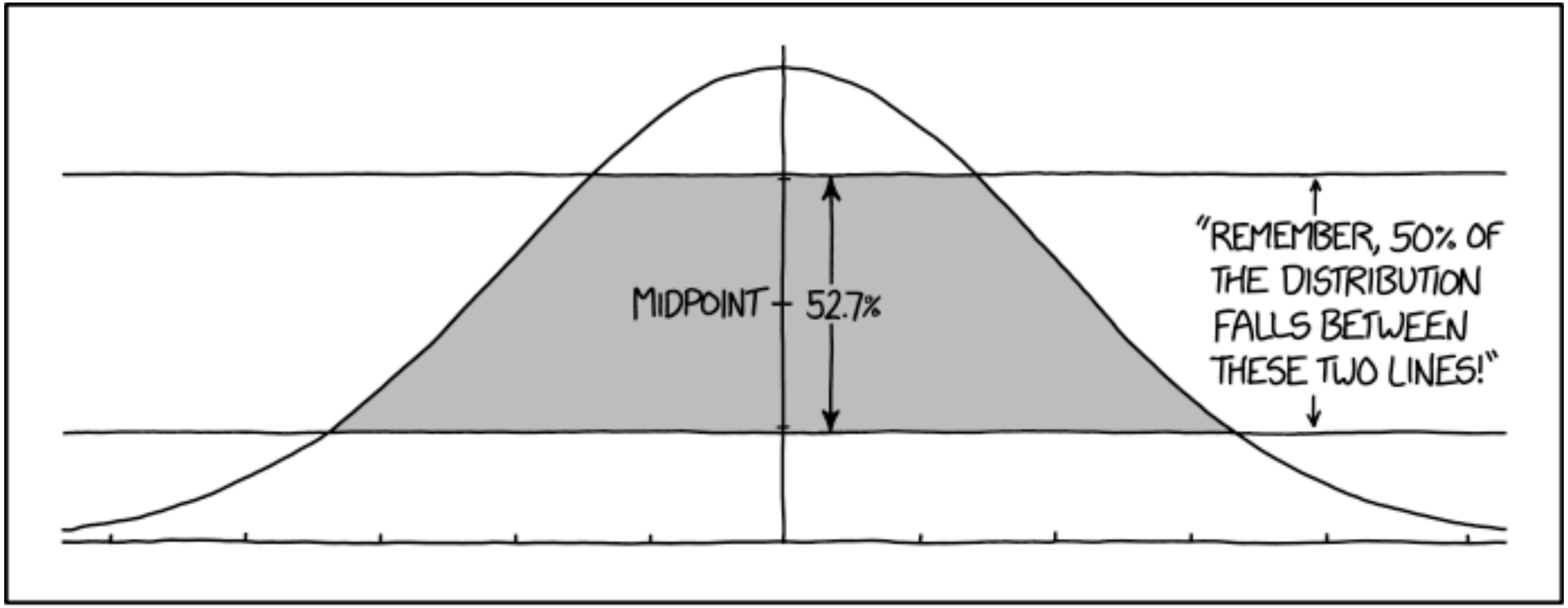


HOW TO ANNOY A STATISTICIAN

Continuous Zoo

CSE 312 Su23
Lecture 13

Continuous Zoo

PDF

$X \sim \mathrm{Unif}(a,b)$	$X \sim \mathrm{Exp}(\lambda)$	$X \sim \mathcal{N}(\mu, \sigma^2)$
$f_X(k) = \frac{1}{b-a}$	$f_X(k) = \lambda e^{-\lambda k}$ for $k \geq 0$	$f_X(k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
$\mathbb{E}[X] = \frac{a+b}{2}$	$\mathbb{E}[X] = \frac{1}{\lambda}$	$\mathbb{E}[X] = \mu$
$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$	$\mathrm{Var}(X) = \frac{1}{\lambda^2}$	$\mathrm{Var}(X) = \sigma^2$

1

2

It's a smaller zoo, but it's just as much fun!

Continuous Uniform Distribution

$X \sim \text{Unif}(a, b)$ (uniform real number between a and b)

PDF: $f_X(k) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq k \leq b \\ 0 & \text{otherwise} \end{cases}$

CDF: $F_X(k) = \begin{cases} 0 & \text{if } k < a \\ \frac{k-a}{b-a} & \text{if } a \leq k \leq b \\ 1 & \text{if } k \geq b \end{cases}$

$\mathbb{E}[X] = \frac{a+b}{2}$

$\text{Var}(X) = \frac{(b-a)^2}{12}$

Exponential

Exponential Random Variable

What is the **amount of time** until the next event occurs?

> The next bus arrives? The next earthquake? The next geyser eruption?

Exponential Random Variable

What is the **amount of time** until the next event occurs?

> The next bus arrives? The next earthquake? The next geyser eruption?

The above examples may remind you of the Poisson distribution!

Poisson process

Exponential Random Variable

The exponential distribution looks at the amount of time until the next event occurs.

Events occur via a Poisson process, i.e.,

- events occur with a constant average rate, λ

- events occur independently of the timing of other events

$$\Omega_X = \underbrace{\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} | x \geq 0\}}_{\text{\# of events}}$$

Exponential CDF

$Y \sim Exp(\lambda)$

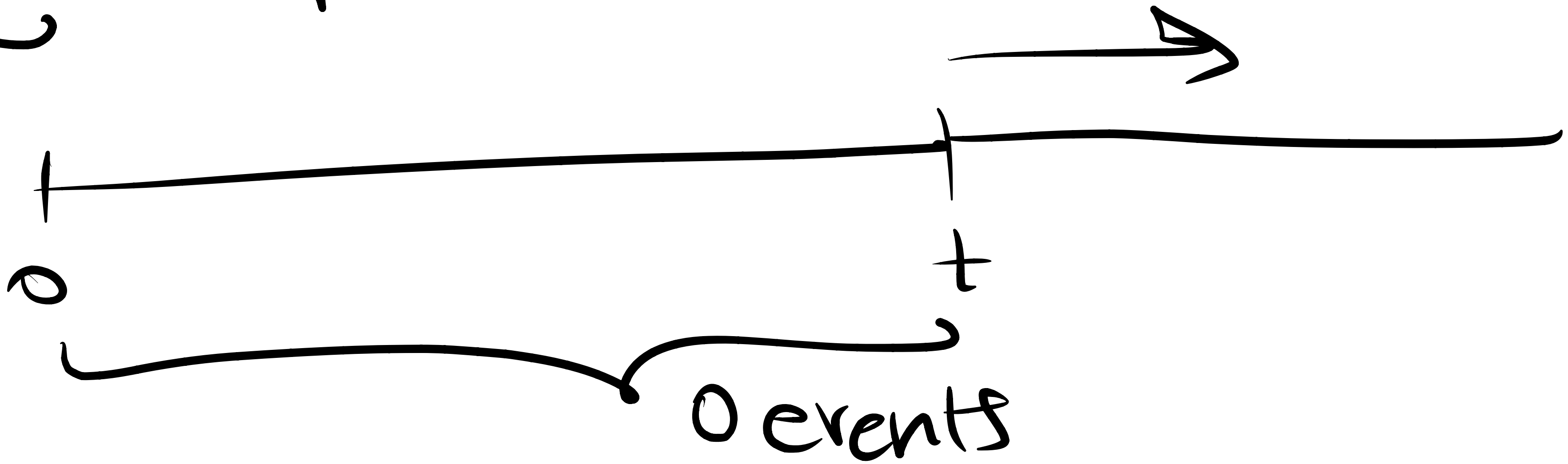
Y is the time until the next event

λ is the average number of events per unit time

$$F_Y(t) = \mathbb{P}(Y \leq t)$$

$$= 1 - \mathbb{P}(Y > t)$$

complement



Exponential CDF

$\lambda - 1s$

$5\lambda - 5s$

$Y \sim Exp(\lambda)$

Y is the time until the next event

λ is the average number of events per unit time

$F_Y(t) = \mathbb{P}(Y \leq t)$

$= 1 - \underbrace{\mathbb{P}(Y > t)}$

$= 1 - \mathbb{P}(X = 0)$

$X \sim Poi(t\lambda)$

$(Y > t)$: It takes over t units of time for the next event. What does that tell us about the **number of events over t units of time**?

Exponential CDF

$Y \sim Exp(\lambda)$

Y is the time until the next event

λ is the average number of events per unit time

$F_Y(t) = \mathbb{P}(Y \leq t)$

$= 1 - \mathbb{P}(Y > t)$

$(Y > t)$: It takes over t units of time for the next event. What does that tell us about the **number of events over t units of time**?

Let $X \sim Poi(\lambda t)$ be the number of events – we are looking at $\mathbb{P}(X = 0)$

Exponential CDF

$Y \sim Exp(\lambda)$

Y is the time until the next event

λ is the average number of events per unit time

$$F_Y(t) = \mathbb{P}(Y \leq t)$$

$$= 1 - \mathbb{P}(Y > t)$$

$$= 1 - \mathbb{P}(X = 0)\ [X \sim Poi(\lambda t)]$$

$$= 1 - e^{-\lambda t}\,\frac{(\lambda t)^0 = 1}{0! = 1}$$

$$= 1 - e^{-\lambda t}\ \text{(for } t \geq 0\text{, } F_Y(t) = 0 \text{ for } t < 0\text{)}$$

Exponential PDF

We know the CDF, $F_Y(t) = \mathbb{P}(Y \leq t) = \underline{1 - e^{-\lambda t}}$

What's the density?

$f_Y(t) =$

Exponential PDF

We know the CDF, $F_Y(t) = \mathbb{P}(Y \leq t) = 1 - e^{-\lambda t}$ $e^{-\lambda t}$

What's the density? $t\lambda$

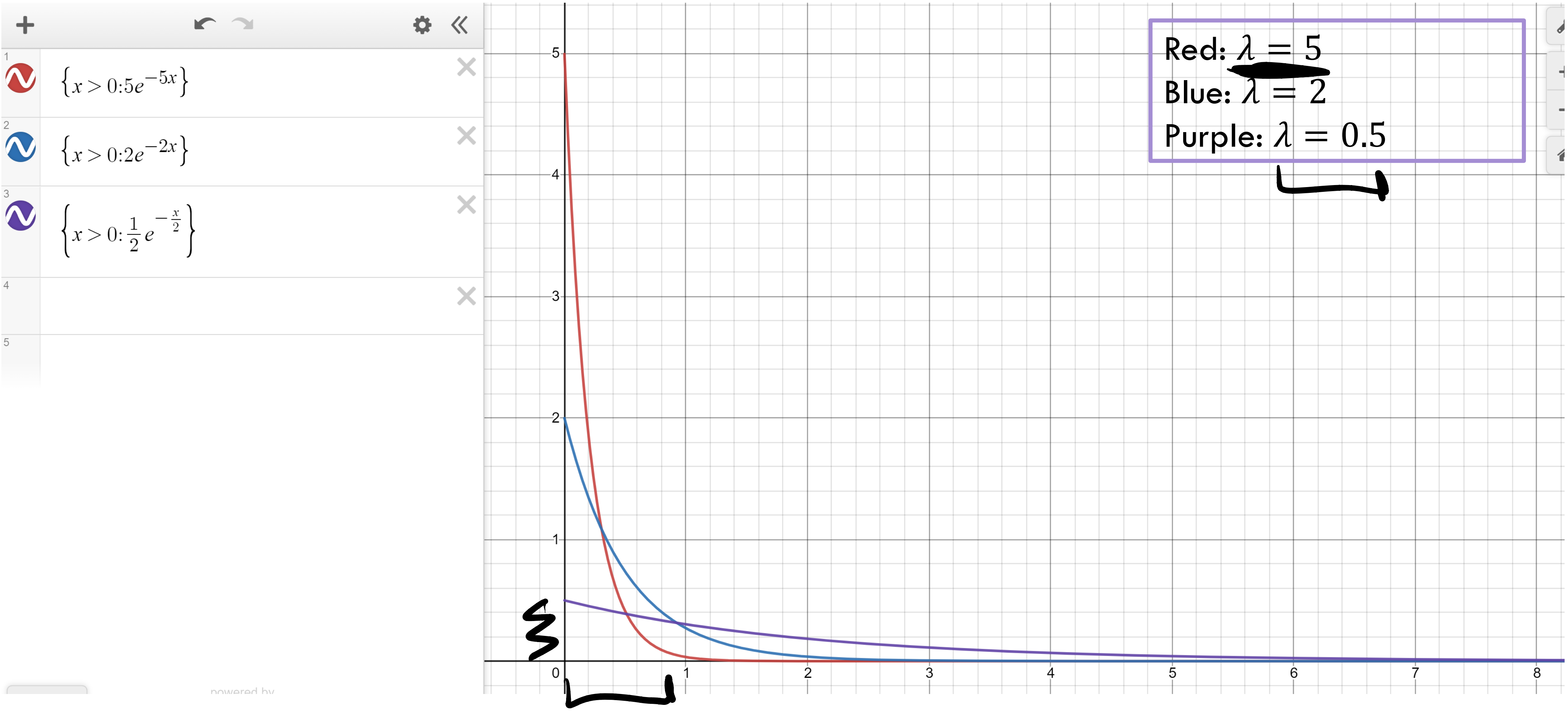
$f_Y(t) = \frac{d}{dt}\left(1 - e^{-\lambda t}\right) = 0 - \frac{d}{dt}\left(e^{-\lambda t}\right) = \lambda e^{-\lambda t}$.

For t ≥ 0 it's that expression

For $t < 0$ it's just $\mathbf{0}$.

Exponential PDF

↑ events



Memorylessness

Memorylessness

A random variable X is memoryless if for all $s, t \geq 0$

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X > s+t \cap X > s)}{\mathbb{P}(X > s)}$$

$$= \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)} = \frac{1 - \mathbb{P}(X \leq s+t)}{1 - \mathbb{P}(X \leq s)}$$

Memorylessness

Memorylessness

A random variable X is memoryless if for all $s, t \geq 0$

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X>s+t \cap X>s)}{\mathbb{P}(X>s)}$$

$$= \frac{\mathbb{P}(X>s+t)}{\mathbb{P}(X>s)} = \frac{1-F_X(s+t)}{1-F_X(s)}$$

$$= \frac{1-(1-e^{-\lambda(s+t)}}{1-(1-e^{-\lambda s})}$$

Memorylessness

Memorylessness

A random variable X is memoryless if for all $s, t \geq 0$

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X>s+t \cap X>s)}{\mathbb{P}(X>s)}$$

$$= \frac{\mathbb{P}(X>s+t)}{\mathbb{P}(X>s)} = \frac{1-F_X(s+t)}{1-F_X(s)}$$

$$= \frac{1-(1-e^{-\lambda(s+t)})}{1-(1-e^{-\lambda s})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$$

$$= 1 - (1 - e^{-\lambda t}) = 1 - \mathbb{P}(X \leq t) = \mathbb{P}(X > t)$$

← start

CDF

Memorylessness

So, an exponential random variable is memoryless: i.e., waiting doesn’t make the next event happen any sooner!

Expectation of an exponential

Let $X \sim \text{Exp}(\lambda)$

$\mathbb{R}$

$\mathbb{E}[X] = \int_{-\infty}^{\infty} z \cdot f_X(z) \mathrm{d}z$

$= \int_0^{\infty} z \cdot \lambda e^{-\lambda z}\, dz$

Expectation of an exponential

Don't worry about the derivation (it's here if you're interested; you're not responsible for the derivation. Just the value)

Let $X \sim \text{Exp}(\lambda)$

$\mathbb{E}[X] = \int_{-\infty}^{\infty} z \cdot f_X(z)\text{d}z$

$\boxed{= \int_0^{\infty} z \cdot \lambda e^{-\lambda z}\, dz}$

Let $u = z;\ dv = \lambda e^{-\lambda z} dz\ (v = -e^{-\lambda z})$

Integrate by parts: $-ze^{-\lambda z} - \int -e^{-\lambda z}\, dz = -ze^{-\lambda z} - \frac{1}{\lambda} e^{-\lambda z}$

Definite Integral: $-ze^{-\lambda z} - \frac{1}{\lambda} e^{-\lambda z} |_{z=0}^{\infty} = (\lim_{z \to \infty} -ze^{-\lambda z} - \frac{1}{\lambda} e^{-\lambda z})\ - (0 - \frac{1}{\lambda})$

By L'Hopital's Rule $(\lim_{z \to \infty} -\frac{z}{e^{\lambda z}} - \frac{1}{\lambda e^{\lambda z}})\ - (0 - \frac{1}{\lambda}) = \left(\lim_{z \to \infty} -\frac{1}{\lambda e^{\lambda z}}\right) + \frac{1}{\lambda} = \frac{1}{\lambda}$

Variance of an exponential

If $\mathrm{X} \sim \mathrm{Exp}(\lambda)$ then $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

Similar calculus tricks will get you there.

Exponential

1 min − 5 accidents.

5 min X

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

Parameter $\lambda \geq 0$ is the average number of events in a unit of time.

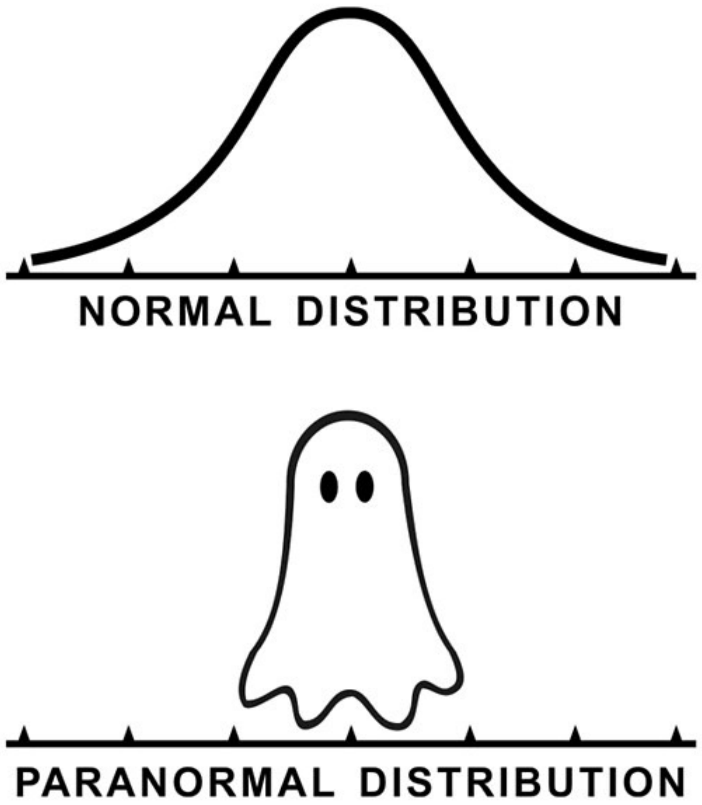
$$f_X(k) = \begin{cases} \lambda e^{-\lambda k} & \text{if } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(k) = \begin{cases} 1 - e^{-\lambda k} & \text{if } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Normal



Normal Random Variable

The **most important** random variable type!

AKA Gaussian RV

Has two parameters:

- mean μ $\mathbb{E}[X]$
- variance σ^2 $\mathrm{Var}(X)$

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

mean

Normal PDF

distance from value to mean.

$X \sim \mathcal{N}(\mu, \sigma^2)$ iff X has the following PDF:

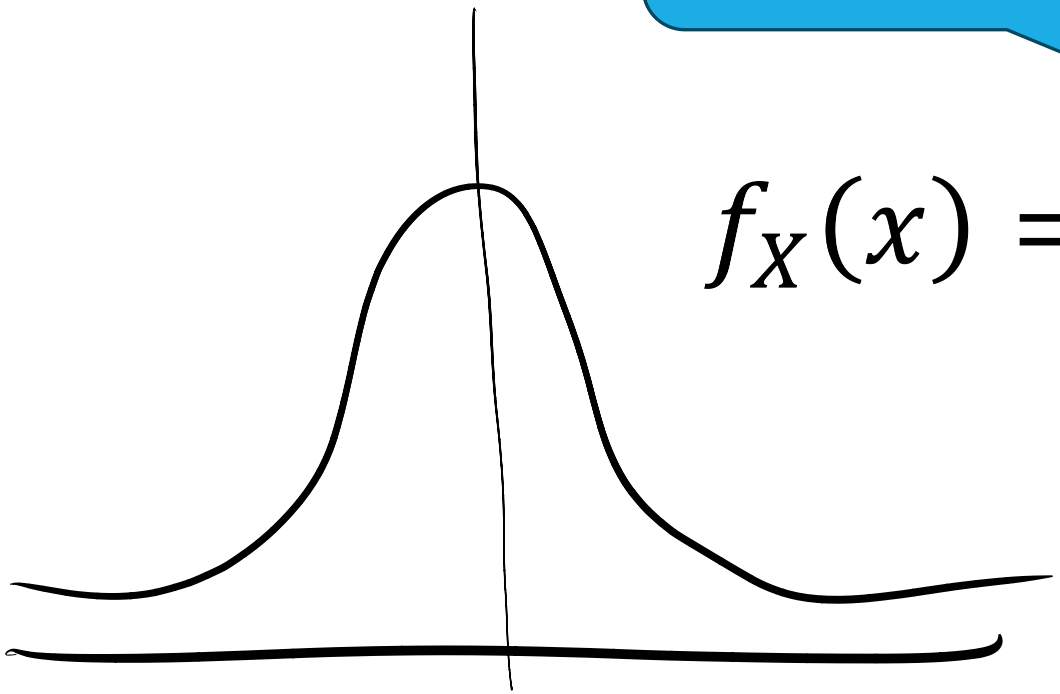
Normalization constant

Symmetric around mean

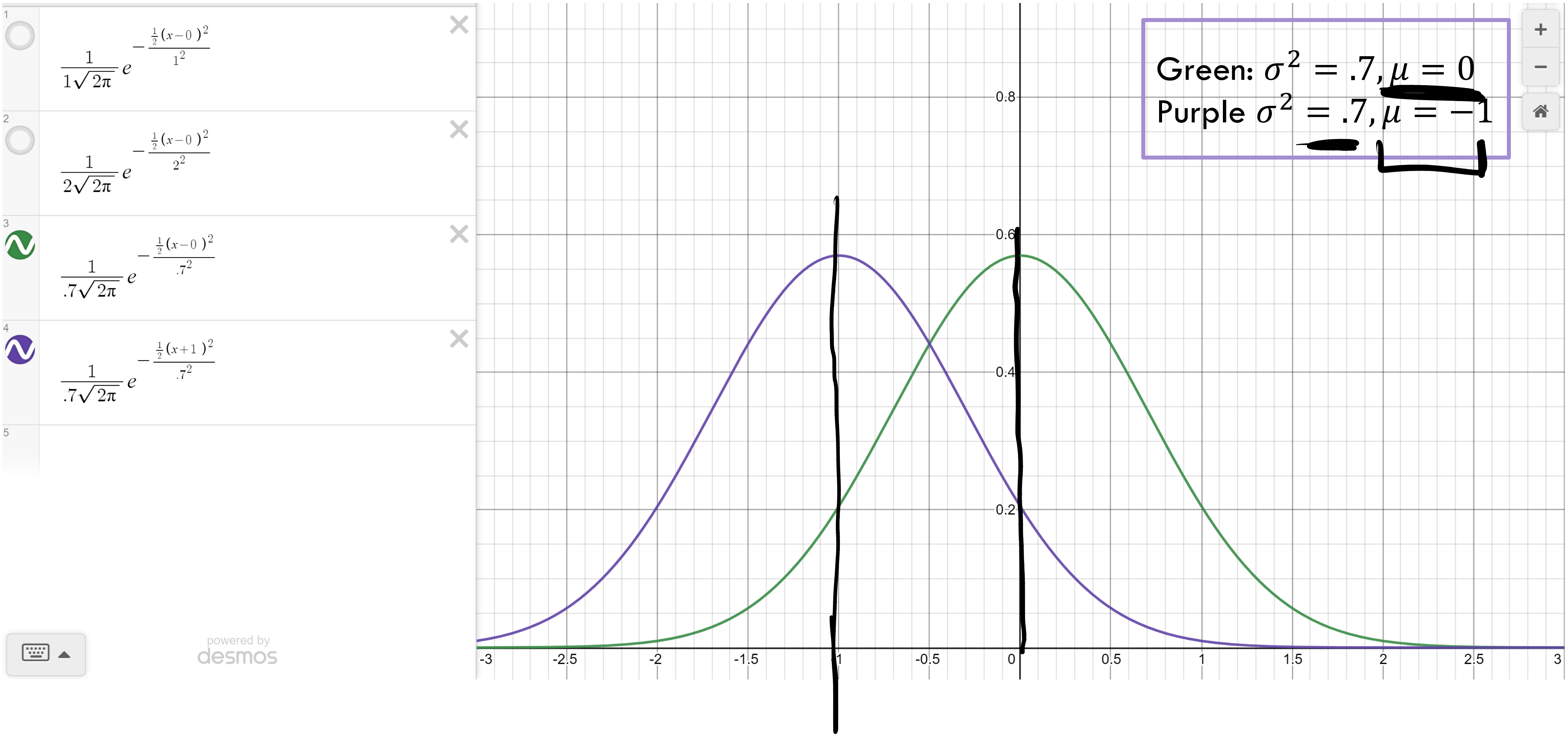
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Variance for spread

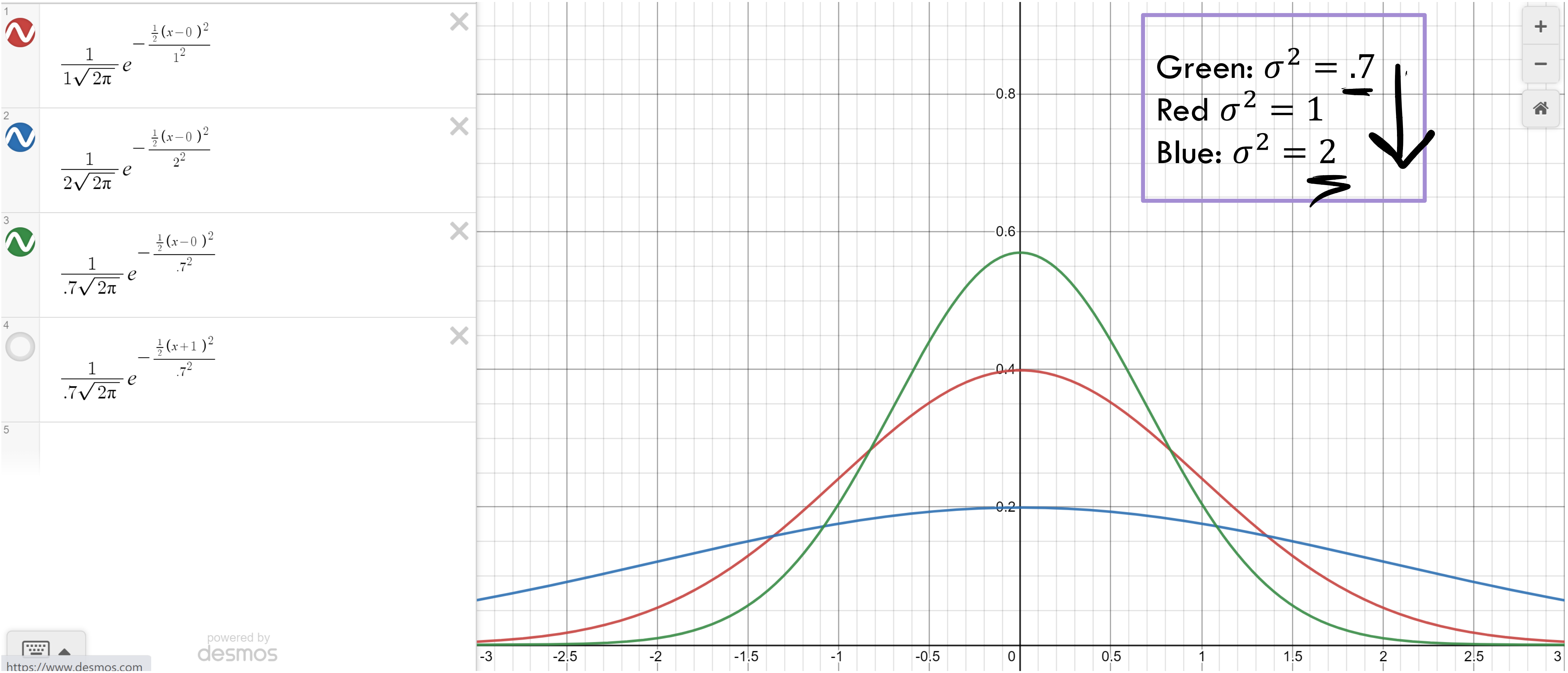
Exponential tail



Changing the mean



Changing the variance



Closure of Normal under Scale and Shift

$3/2X$ $\Omega_{\frac{3}{2}X}: \{0, \frac{3}{2}, \ldots, \frac{3}{2}n\} \neq \text{Bin}$

MGF

When we scale a normal (multiplying by a constant or adding a constant) we get a normal random variable back!

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

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

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

Normals are unique in that you get a NORMAL back.

If you multiply a binomial by 3/2 you don't get a binomial (its support isn't even integers!)

$X \sim \text{Bin}(n, p)$
$\Omega_X: \{0, 1, \ldots, n\}$

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

Closure of Normal under Addition

When we add two independent normal random variables, you get another normal random variable

$\mathbb{E}[Z] = \mathbb{E}[aX + bY + c]$
$= a\mathbb{E}[X] + b\mathbb{E}[Y] + c$

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

Then for $Z = aX + bY + c$, $Z \sim \mathcal{N}(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$

Normals are unique in that you get a NORMAL back.

Ok… what about the CDF?

There is **no closed form** for the normal CDF.

(We still need to do calculations that use $F_X(k) = \mathbb{P}(X \leq k)$ though – so how?)

We have one (1) CDF table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

The CDF table for the **standard normal RV** $Z \sim \mathcal{N}(0,1)$.

Yes, we're going to use a table in 2023. Mainly for consistency in this class.

(In the real world, we have programming libraries like Python's scipy: stats.norm.cdf(x, mean, std))

Standard Normal Random Variable

$\mu \quad \sigma^2$

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

The CDF of the standard normal has a special function: Φ

$\Phi(z) = \mathbb{P}(Z \leq z) = F_Z(z)$ for the standard normal

Fun fact: this table was first computed in 1799 using the Taylor series expansion

How do we get from $X \sim \mathcal{N}(\mu, \sigma^2)$ to $Z \sim \mathcal{N}(0,1)$?

We standardize!

Let $Z = \frac{X-\mu}{\sigma}$.

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\mathbb{E}[X-\mu] = \frac{1}{\sigma}(\mathbb{E}[X]-\mu) = \frac{1}{\sigma}\cdot 0 = 0$$

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X-\mu) = \frac{1}{\sigma^2}Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

The normal is closed under scale & shift: $Z \sim \mathcal{N}(0,1)$.

Table of Standard Normal CDF

Z

$\mathbb{P}(Z \le 2.47)$

$\Phi(2.47)$

The way we'll evaluate the CDF of a normal is to:

1. Convert to a standard normal 0, 1
2. Round the "z-score" to the hundredths place.
3. Look up the value in the table.

[Link: Z table on course homepage]

Common names: Z-table, phi-table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	$\boxed{0.84134}$	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	$\boxed{0.97725}$	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	$\boxed{0.99324}$	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

$\mathbb{P}(X \le k)$

$\mathbb{P}(Z \le 1) \rightarrow \underline{1.00}$

1

How to use the table?

We'll use the notation $\Phi(z)$ to mean $F_X(z)$ where $X \sim \mathcal{N}(0,1)$.

$\mu \ \sigma^2$

Let $Y \sim \mathcal{N}(5,4)$ what is $\mathbb{P}(Y \leq 9)$?

$\mathbb{P}(Y \leq 9)$

$= \mathbb{P}\left(\boxed{\frac{Y-5}{2}} \leq \boxed{\frac{9-5}{2}}\right)$ (algebra on both sides) standardize

$= \mathbb{P}(Z \leq \frac{9-5}{2})$ where Z is $\mathcal{N}(0,1)$.

$= \mathbb{P}\left(Z \leq \frac{9-5}{2}\right) = \Phi(2.00) = 0.97725.$

2

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

Another use…

Let $Y \sim \mathcal{N}(5,4)$ what is $\mathbb{P}(Y > 9)$?

$\mathbb{P}(Y > 9)$

$= \mathbb{P}\left(\frac{Y-5}{2} > \frac{9-5}{2}\right)$

$= \mathbb{P}(Z > \frac{9-5}{2})$ where Z is $\mathcal{N}(0,1)$.

$= 1 - \mathbb{P}\left(Z \leq \frac{9-5}{2}\right) = 1 - \Phi(2.00) = 1 - 0.97725 = .02275.$

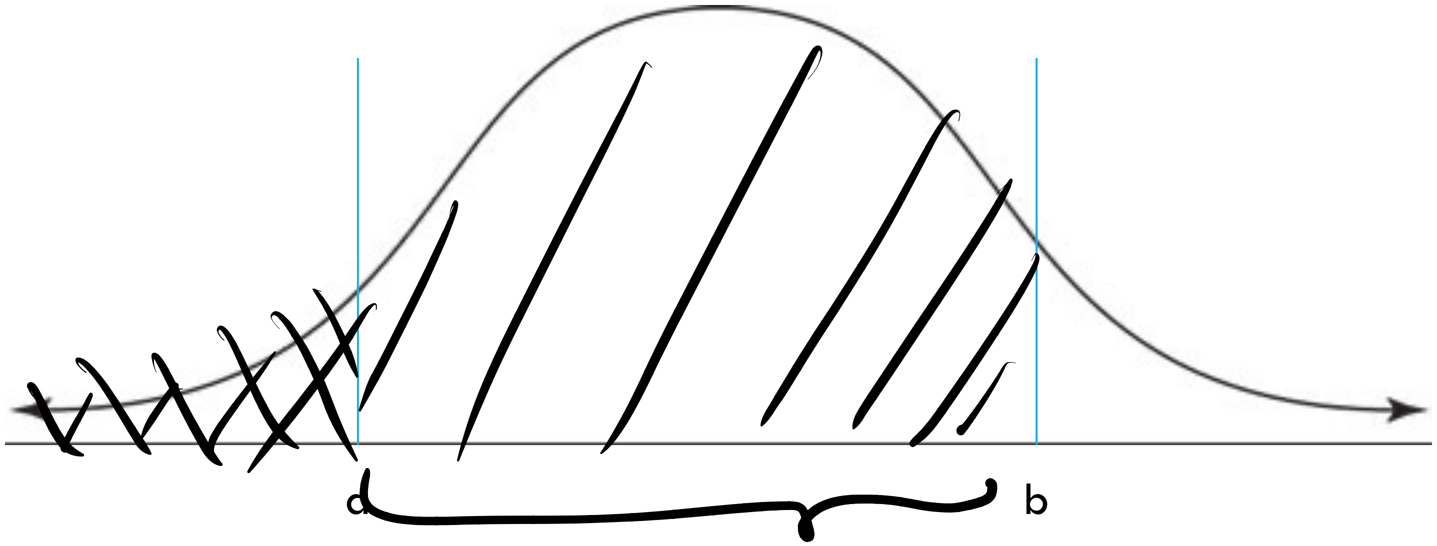
Another trick

$\int_a^b$

How do we find $\mathbb{P}(a \leq X \leq b)$ using this table?

Can we express it in terms of the CDF?

$$= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$$
$$= F_X(b) - F_X(a)$$



Symmetry

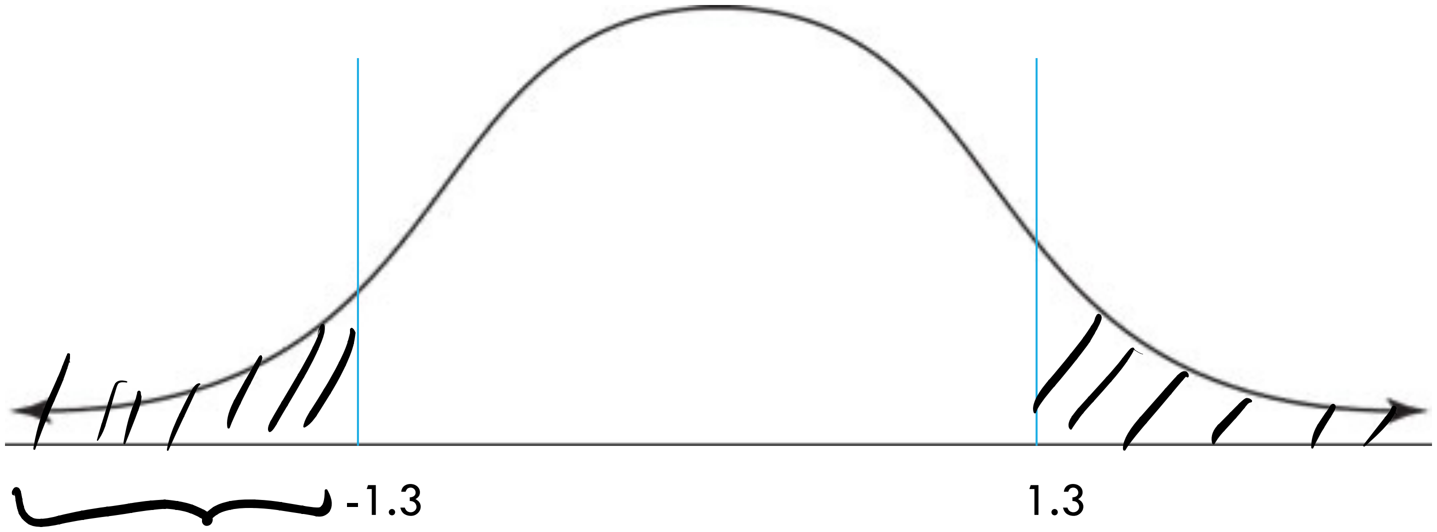
How do we find $\mathbb{P}(X \leq -1.3)$? Our table only has CDFs for positive values! 😱

Recall: the normal distribution is symmetric.

$\mathbb{P}(X \leq -1.3) = \mathbb{P}(X \geq 1.3)$ symmetry

$= 1 - \mathbb{P}(X \leq 1.3)$

$= 1 - F_X(1.3)$



Normal (aka Guassian)

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

Parameter μ is the expectation; σ^2 is the variance.

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

$F_X(k)$ has no closed form. Use the table.

$\mathbb{E}[X] = \mu$

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

Key Normal Calculations

CDF notation $F_Z(k)$

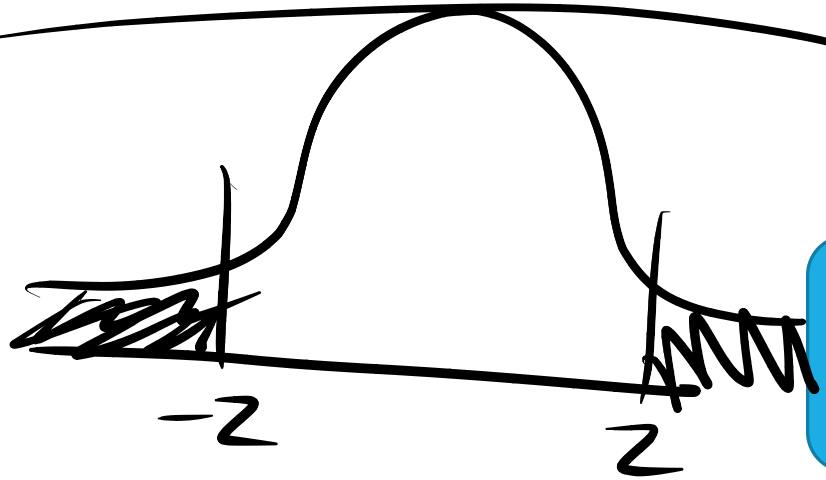
$Z \sim \mathcal{N}(0,1)$

1. $\mathbb{P}(Z \leq z)$ $F_Z(z)$
2. $\mathbb{P}(Z < z)$ $F_Z(z)$ $\mathbb{P}(Z=z)$
3. $\mathbb{P}(Z \geq z)$ $1-(2) = 1 - F_Z(z)$
4. $\mathbb{P}(Z \leq -z)$ $\mathbb{P}(Z \geq z) = 1 - F_Z(z)$

$\mathbb{P}(Z \geq -z)$

$\mathbb{P}(y < Z < z)$

$\mathbb{P}(|Z| > z)$



Your turn!

Key Normal Calculations

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

$\mathbb{P}(Z \leq z)$

$\mathbb{P}(Z < z)$

$\mathbb{P}(Z \geq z)$

$\mathbb{P}(Z \leq -z)$

5. $\mathbb{P}(Z \geq -z)$

6. $\mathbb{P}(y < Z < z)$

7. $\mathbb{P}(|Z| > z)$

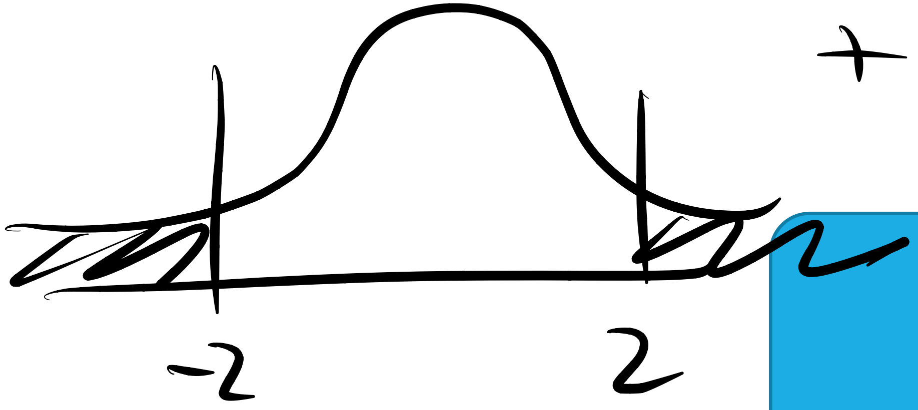
5. $\mathbb{P}(Z \geq -z) = 1-(1-F_Z(z))$
$= F_Z(z)$

6. $F_Z(z) - F_Z(y)$

$2(\mathbb{P}(Z<-z))$

7. $\mathbb{P}(|Z|>z) = \mathbb{P}(\underline{Z>z}) + \mathbb{P}(\underline{Z<-z})$

$\underline{2(1-F_Z(z))}$

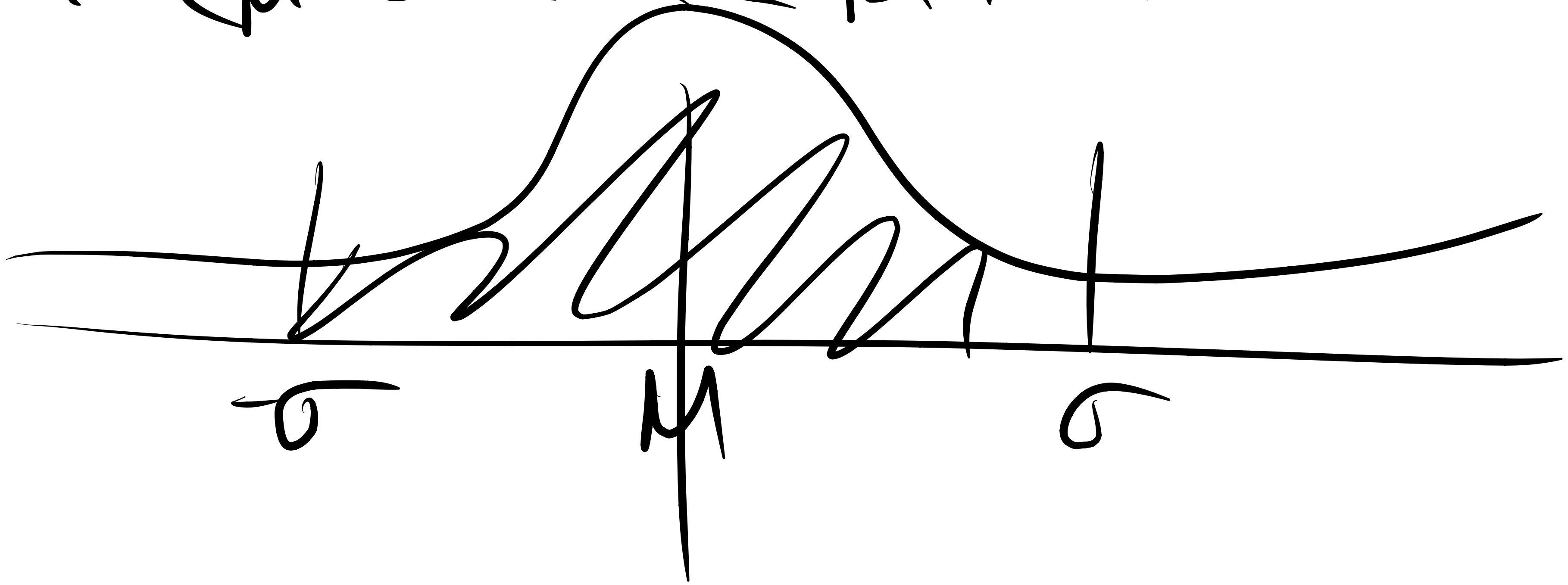


Your turn!

‘Within one standard deviation’

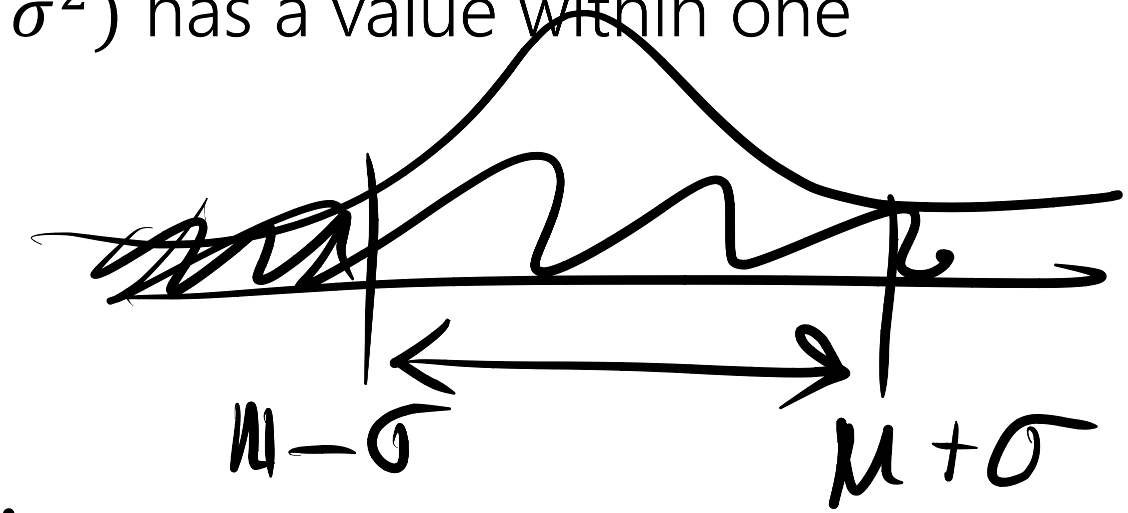
What is the probability that $X \sim \mathcal{N}(\mu, \sigma^2)$ has a value within one standard deviation of its mean?

$\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma)$



‘Within one standard deviation’

What is the probability that $X \sim \mathcal{N}(\mu, \sigma^2)$ has a value within one standard deviation of its mean?



$$\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma)$$

$$= \mathbb{P}(X \leq \mu + \sigma) - \mathbb{P}(X \leq \mu - \sigma)$$

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

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

symmetry

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

$$= 2\Phi(1) - 1 = 2 \cdot 0.84134 - 1 = 0.68268$$

In real life

You’ll sometimes hear statisticians refer to the “68-95-99.7 rule” which is the probability of being within 1, 2, or 3 standard deviations of the mean.

Extra

You said the normal RV was most important? And conveniently didn't mention that ever again? What happened to defending your argument?

μ, σ

Patience, grasshoppers

We have spent 7 (!) lectures looking at random variables.

Why?

We want to **model real-life situations** with probability distributions.

Fancy version: A Guassian maximizes entropy when we only know the mean and variance of a dataset.

TLDR; it is the probability distribution that best represents our knowledge of mean and variance

But also… that doesn’t mean it’s always the most accurate.

Easy to use, though.

You should always be critical of the models we make.

Central limit theorem

You may have often heard it said that many things are normally distributed.

Next class, we'll look at why ☺

combinatorics
probability
random variables

modelling
uncertainty theory
ML