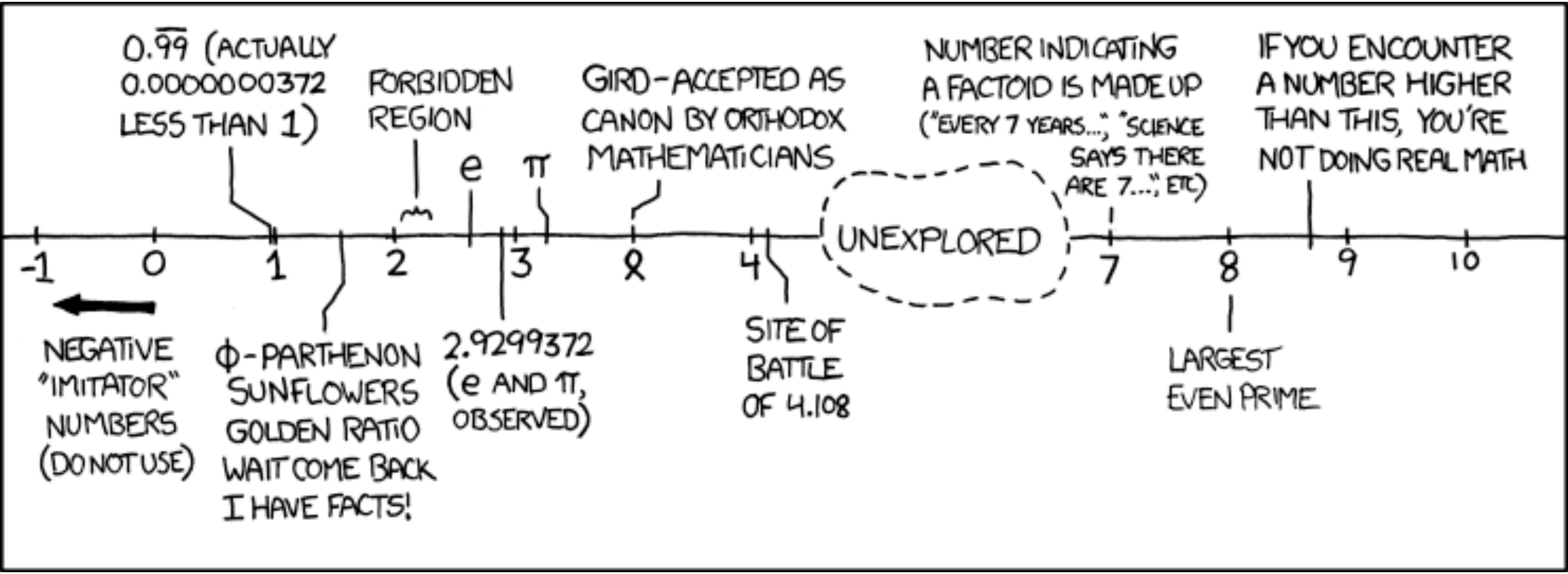


Continuous Probability

CSE 312 Su23
Lecture 12

https://xkcd.com/899/

Announcements

HW4 is out, and due next Tuesday (homeworks will be due weekly on Tuesdays going forward)

Today

Continuous Probability

- Probability Density Function
- Cumulative Distribution Function

Goal for today is to get intuition on what's different in the continuous case. Your goal today is to start building up a gut-feeling of what's happening.

ASK QUESTIONS, (always, but today especially).

Continuous Random Variables

From Discrete to Continuous

So far, we have looked at random variables that have countable (countably finite or infinite) sample spaces – for e.g., the number of heads in n coin flips

$|\Omega| = 2^{100}$

Geometric

$\{1, \ldots, \infty\}$

From Discrete to Continuous

So far, we have looked at random variables that have countable (countably finite or infinite) sample spaces – for e.g., the number of heads in n coin flips

What if we wanted to look at things like…

The height of a randomly chosen person

How long until the next bus shows up

We need to define random variables with an uncountably infinite sample space (for e.g., the set of all real numbers $\mathbb{R}$)

Aside: Why?

Wait, we're computer scientists. Computers don't do real numbers, why should we?

double

Continuous random variables will be a useful model for enormous sample spaces. The math will be easier.

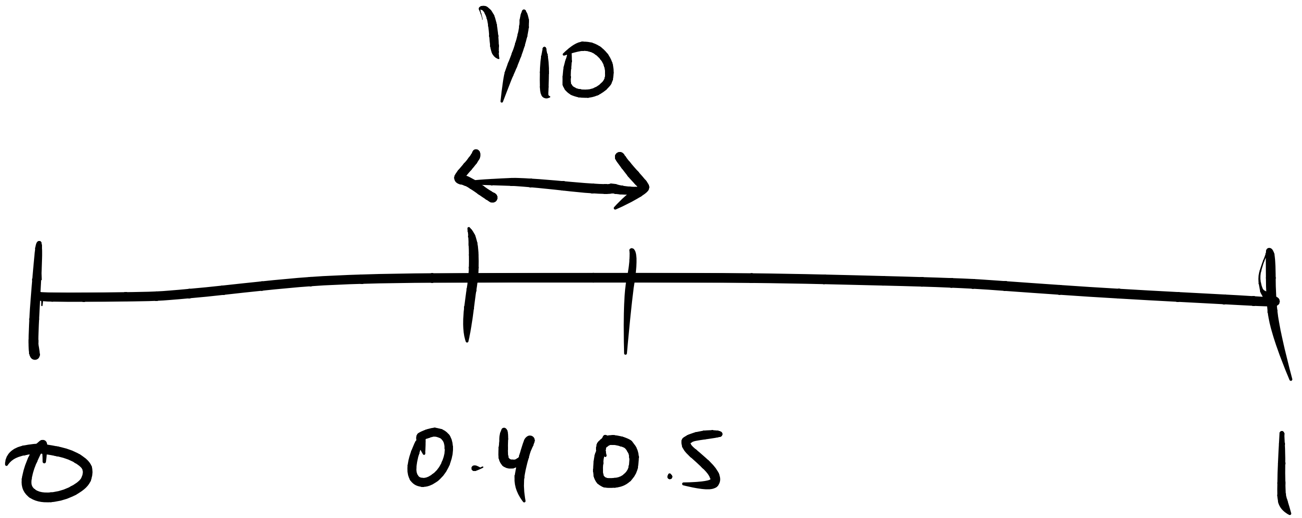
Example: polling a large population.

The sample space is actually discrete! But we're going to round the result anyway. Make it continuous first for easier math, then round.

Why Need New Rules?

We want to choose a uniformly random real number between 0 and 1.

What's the probability the number is between 0.4 and 0.5?



$$\mathbb{P}(0.4 \le X \le 0.5) = \frac{1}{10}$$

Why Need New Rules?

We want to choose a uniformly random real number between 0 and 1.

What's the probability the number is between 0.4 and 0.5?

For discrete random variables, we'd ask for $\frac{|E|}{|\Omega|} = \frac{\infty}{\infty}$

Why Need New Rules?

We want to choose a uniformly random real number between $\mathbf{0}$ and $\mathbf{1}$.

What's the probability the number is between $\mathbf{0.4}$ and $\mathbf{0.5}$?

For discrete random variables, we'd ask for $\frac{|E|}{|\Omega|}$

So we get $\frac{\infty}{\infty}$

The mathematical tools to get consistent answers from expressions like those is calculus.

PDFs

Let’s start with the PMF

RV value

For discrete RVs, we started with the PMF $p_X(x) = \mathbb{P}(X = x)$.

What might that look like for a continuous RV?

What is the probability that a person’s height is exactly 5.678123589 feet?

What is the probability that a bus arrives exactly 6 hours, 23 minutes and 2976427909 seconds from now?

$$\mathbb{P}(H = 5.67) = \frac{|E|}{|\Omega|} = \frac{1}{\infty} \to 0$$

Let's start with the PMF

For discrete RVs, we started with the PMF $p_X(x) = \mathbb{P}(X = x)$.

What might that look like for a continuous RV?

What is the probability that a person's height is exactly 5.678123589 feet?

What is the probability that a bus arrives exactly 6 hours, 23 minutes and 2976427909 seconds from now?

Remember, for discrete random variables, we look for $\frac{|E|}{|\Omega|}$. But with continuous random variables, $\mathbb{P}(X = 5.678) = \frac{1}{\infty}$?

Probability density functions (PDF)

So… it doesn't make sense to have $\mathbb{P}(X = x)$ for a continuous RV; let's try to create an analogous function.

For Continuous RVs, the analogous object is the "probability density function" (PDF)

We write $f_X(k)$ instead of $p_X(k)$

Let's try to maintain as many rules as we can:

Discrete	Continuous
$p_Y(k) \geq 0$	$f_X(k) \geq 0$
$\sum_k p_Y(k) = 1$	$\int_{-\infty}^{\infty} f_X(k)\ \mathrm{d}k = 1$

non-negativity

normalization

Probability density function

Idea: make it "work right" for **events** since single outcomes don't make sense

$$\underbrace{\mathbb{P}(X_{cont} = k) = 0}_{\omega \in \Omega}$$

$$\mathbb{P}(0.4 \leq X \leq 0.5) = \frac{\text{size of the range for event}}{\text{size of the entire sample space}}$$

Uniform PDF

$\int_{0.3}^{0.4} f_X(k)\,dk = 0.1$

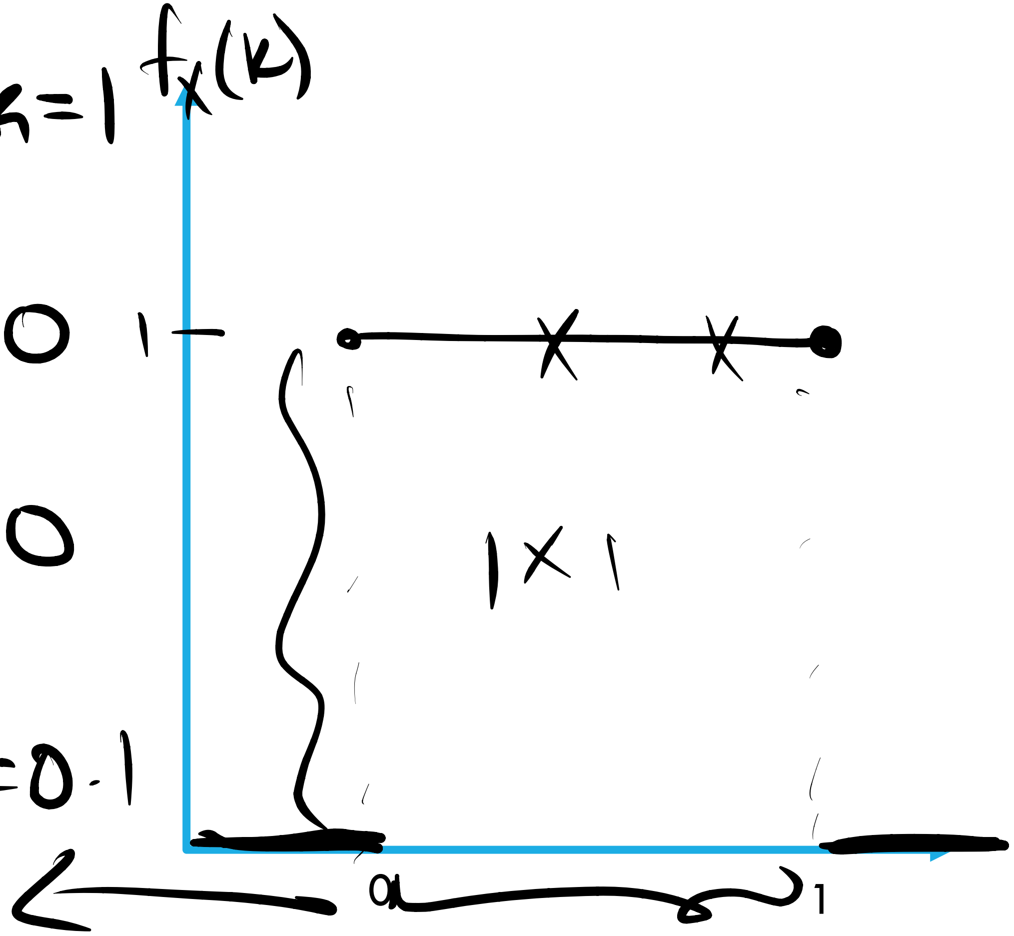
$X \sim Unif(0,1)$ - pick a random number between 0 and 1

$\mathbb{P}(0 \le X \le 1) = 1 \rightarrow \int_0^1 f_X(k)\,dk = 1$

$\mathbb{P}(X < 0) = 0 \rightarrow \int_{-\infty}^{0} f_X(k)\,dk = 0$

$\mathbb{P}(X > 1) = 0 \rightarrow \int_{1}^{\infty} f_X(k)\,dk = 0$

$\mathbb{P}(0.4 \le X \le 0.5) = 0.1 = \int_{0.4}^{0.5} f_X(k)\,dk = 0.1$



Uniform PDF

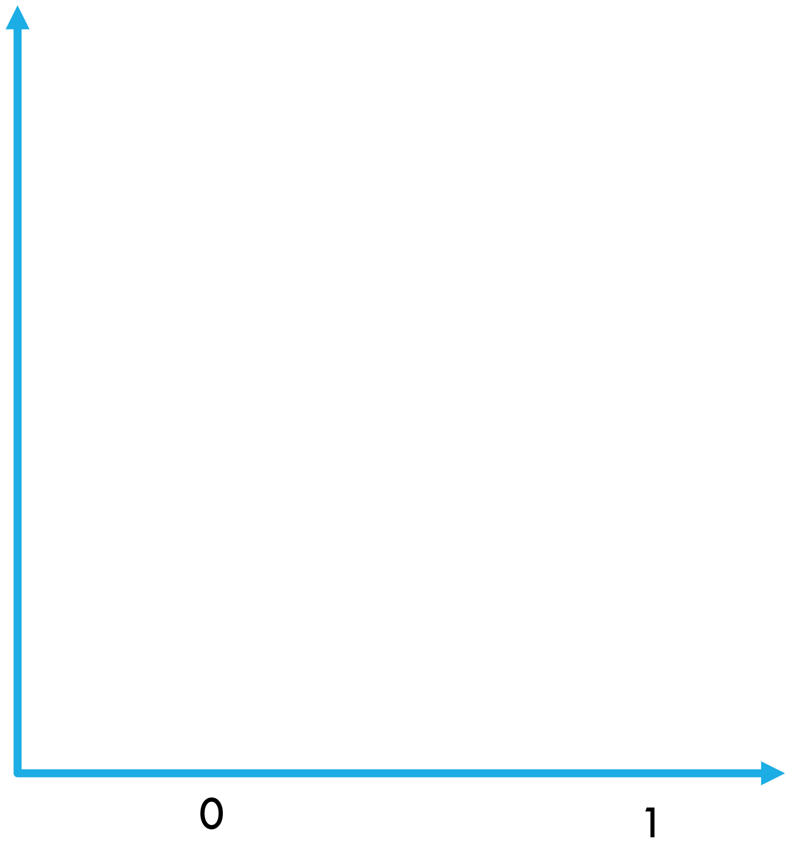
$X \sim Unif(0,1)$ - pick a random number between 0 and 1

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

$\mathbb{P}(X < 0) = 0$

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

$\mathbb{P}(0.4 \leq X \leq 0.5) = 0.1$



Uniform PDF

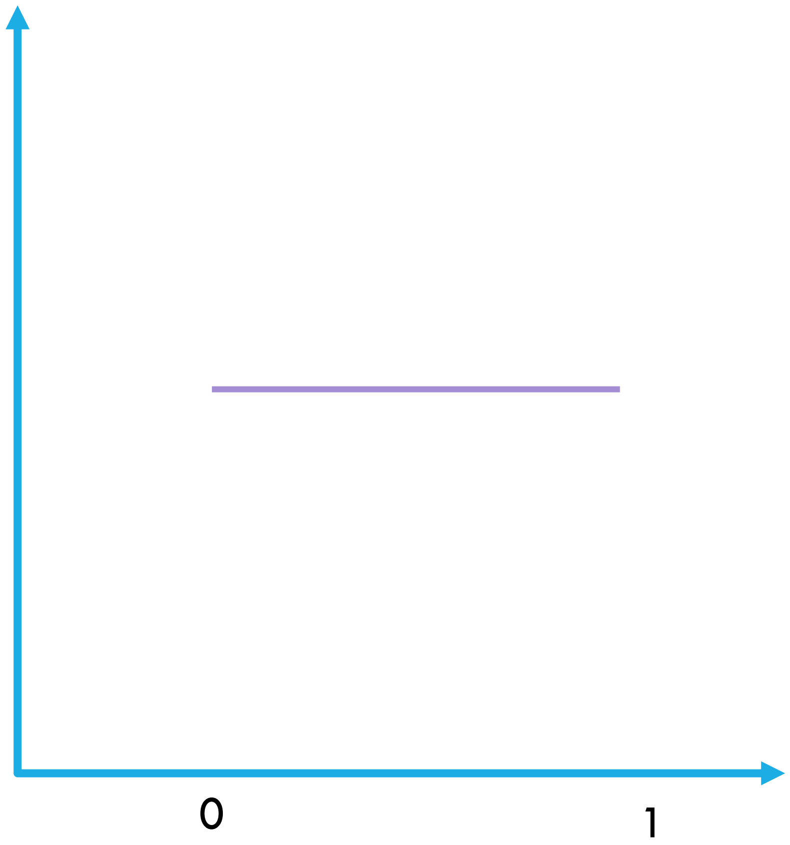
$X \sim Unif(0,1)$ - pick a random number between 0 and 1

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

$\mathbb{P}(X < 0) = 0$

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

$\mathbb{P}(0.4 \leq X \leq 0.5) = 0.1$



Probability density functions

Key idea: integrating the PDF gives us probabilities

But the PDF itself does not give us probabilities

The number that best represents $\mathbb{P}(X = 0.1)$ is 0.

This is different from $f_X(0.1) = 1$

For continuous probability spaces:
Impossible events have probability 0,
but some probability 0 events might be possible.

Probability density functions

Key idea: integrating the PDF gives us probabilities

But the PDF itself does not give us probabilities

The number that best represents $\mathbb{P}(X = 0.1)$ is 0.

This is different from $f_X(0.1) = 1$

So…what is $f_X(x)$???

For continuous probability spaces:
Impossible events have probability 0,
but some probability 0 events might be possible.

Using the PDF

$X = 0.2$

Take the event: $X \approx 0.2$

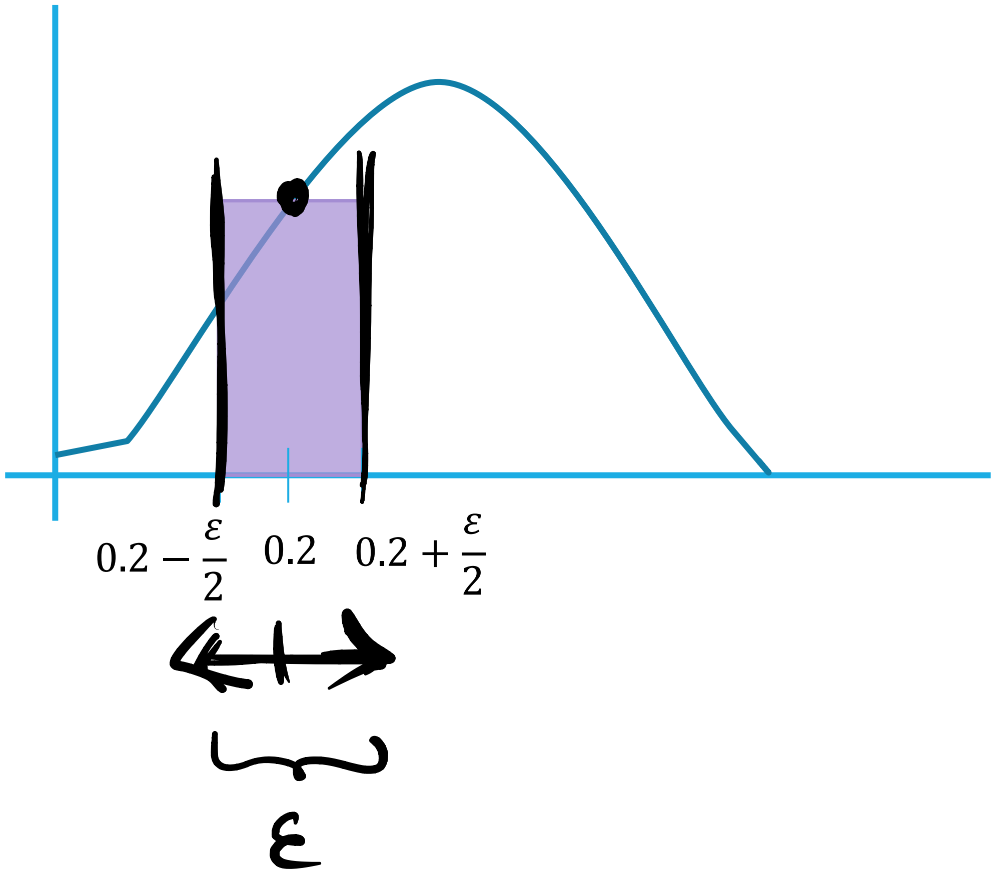
For a small ε:

$\mathbb{P}(X \approx 0.2) = \mathbb{P}(0.2 - \frac{\varepsilon}{2} \leq X \leq 0.2 + \frac{\varepsilon}{2})$

$= \int_{0.2-\frac{\varepsilon}{2}}^{0.2+\frac{\varepsilon}{2}} f_X(z)\, \mathrm{d}z \approx \textit{height} \cdot \textit{width}$

$= f_X\,(0.2) \cdot \varepsilon$

In general, $\mathbb{P}(X \approx q) = f_X(q) \cdot \varepsilon$.



Using the PDF

$$\mathbb{P}(X \underbrace{\approx 0.5}) = f_X(0.5) \cdot \varepsilon$$

Take the event: $X \approx 0.2$

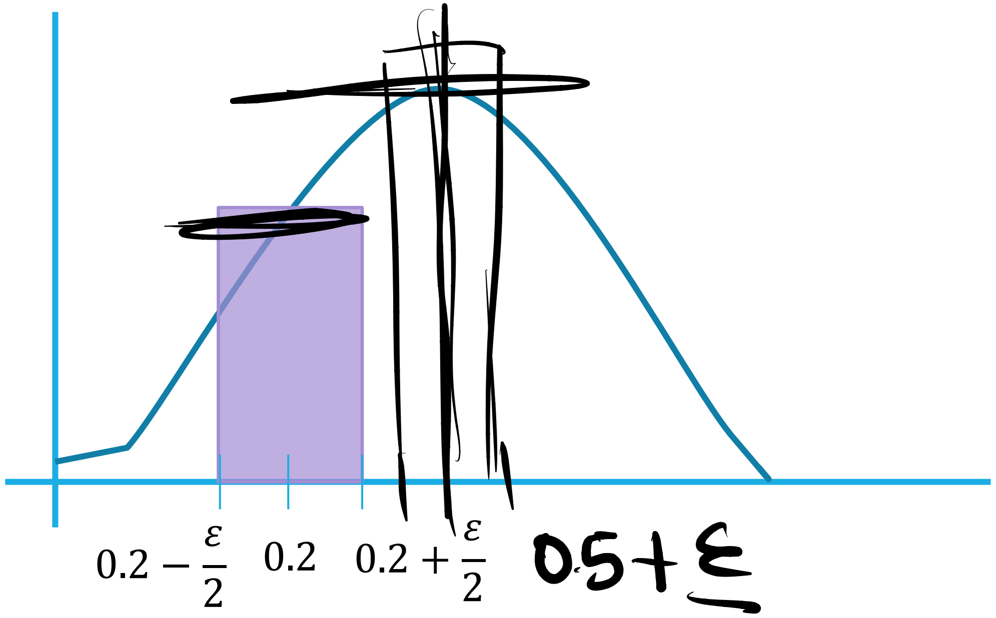
$$\frac{\mathbb{P}(X \approx 0.5)}{\mathbb{P}(X \approx 0.2)} = \frac{f_X(0.5) \cdot \cancel{\varepsilon}}{f_X(0.2) \cdot \cancel{\varepsilon}}$$

For a small ε:

$$\mathbb{P}(X \approx 0.2) = \mathbb{P}(0.2 - \tfrac{\varepsilon}{2} \leq X \leq 0.2 + \tfrac{\varepsilon}{2})$$

$$= \int_{0.2-\frac{\varepsilon}{2}}^{0.2+\frac{\varepsilon}{2}} f_X(z)\,\mathrm{d}z \approx \textit{height} \cdot \textit{width}$$

$$= f_X\,(0.2) \cdot \varepsilon$$



$0.5 - \frac{\varepsilon}{2}$ 0.5

In general, $\mathbb{P}(X \approx q) = f_X(q) \cdot \varepsilon$.

So, what’s the PDF?

It’s the number that when integrated over gives the probability of an event.

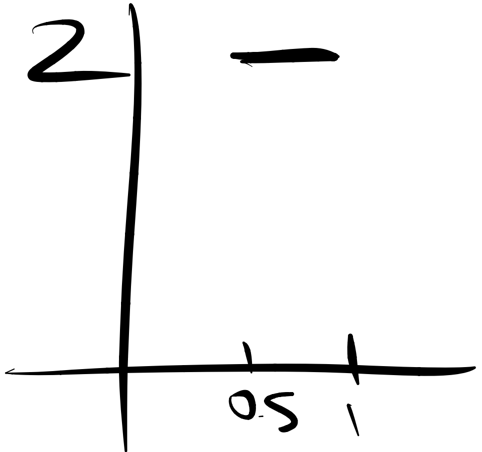
Equivalently, it’s a number such that:

- it is always non-negative

- integrating it over all real numbers gives 1.

- comparing $f_X(k)$ and $f_X(\ell)$ gives the relative chances of X being near k or ℓ.

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

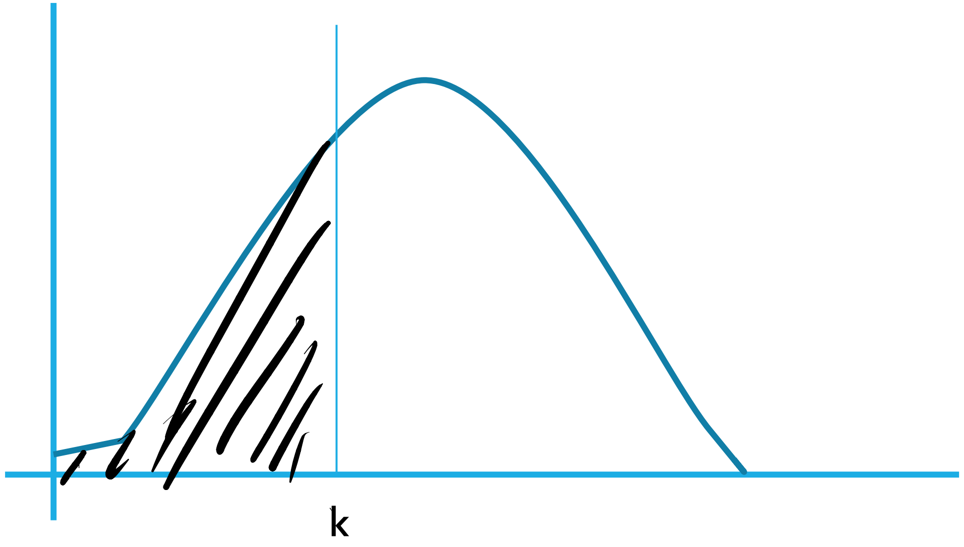


CDFs

What's a CDF?

The Cumulative Distribution Function $F_X(k) = \mathbb{P}(\boldsymbol{X} \le \boldsymbol{k})$ is analogous to the CDF for discrete variables.

$$F_X(k) = \mathbb{P}(X \le k) = \int_{-\infty}^{k} f_X(z)\,\mathrm{dz}$$



What's a CDF?

The Cumulative Distribution Function $F_X(k) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{k})$ is analogous to the CDF for discrete variables.

$$F_X(k) = \underbrace{\mathbb{P}(X \leq k)}_{\text{CDF}} = \int_{-\infty}^{k} \underbrace{f_X(z)}_{\text{PDF}}\, \mathrm{d}z$$

$$\text{CDF} \rightleftarrows \text{PDF}$$

So how do I get from CDF to PDF? Taking the derivative!

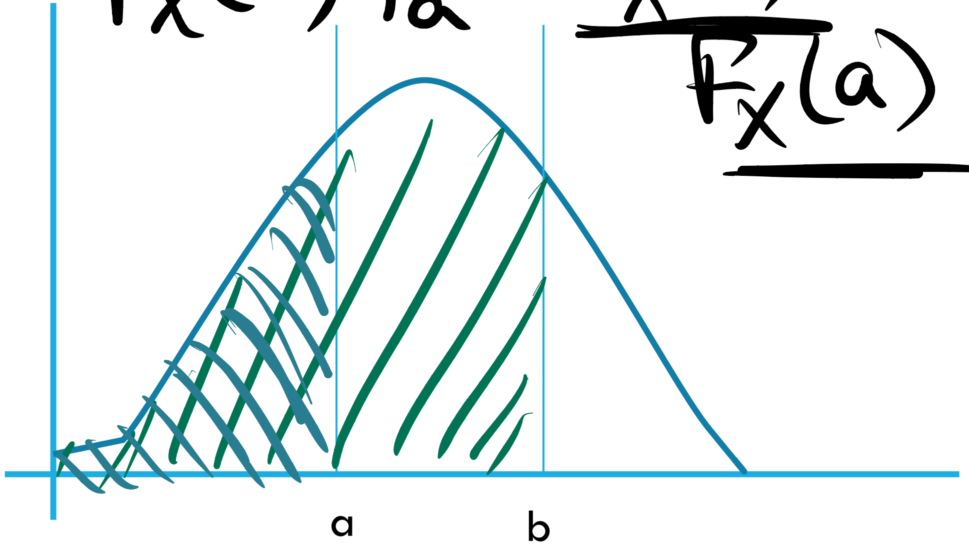
$$\frac{\mathrm{d}}{\mathrm{d}k} F_X(k) = \frac{\mathrm{d}}{dk}\left(\int_{-\infty}^{k} f_X(z)\, \mathrm{d}z\right) = f_X(k)$$

Some other properties of the CDF

How would we calculate $\mathbb{P}(a \leq X \leq b)$ using the CDF?

$$\underbrace{\int_a^b f_X(k)\,dk}_{F_X(k)\big|_a^b} = F_X(b) - F_X(a)$$

We are interested in the area under the curve between *a* and *b*… which is the area to the left of b ($F_X(b)$) minus the area to the left of a ($F_X(a)$)!



Giving us… $F_X(b) - F_X(a)$

$$\mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$$
$$= \mathbb{P}(a < X \leq b) + 0$$

Technical notes

The CDF is a monotone increasing function: so, if $a \leq b, F_x(a) \leq F_X(b)$

$$\underbrace{\lim_{v \to -\infty} F_X(v)} = \underbrace{\mathbb{P}(X \leq -\infty)} = 0$$

$$\lim_{v \to \infty} F_X(v) = \mathbb{P}(X \leq \infty) = 1$$

Expectation & Variance

What about expectation?

For a random variable X, we define:

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

$$\mathbb{E}[X] = \sum_{k} k \cdot p_X(k)$$

As for discrete RVs, LOTUS still applies:

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

$$\mathbb{E}[X^3] = \int_{-\infty}^{\infty} z^3 \cdot f_X(z)\, dz$$

Just replace summing over the PMF with integrating the PDF.

It still represents the average value of X.

Linearity of Expectation

Still true!

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

For all X, Y; even if they're continuous.

Won't show you the proof – for just $\mathbb{E}[aX + b]$, it's

$\mathbb{E}[aX + b] = \int_{-\infty}^{\infty}[aX(k) + b]f_X(k)\,\mathrm{d}k$

$= \int_{-\infty}^{\infty} aX(k)f_X(k)dk + \int_{-\infty}^{\infty} bf_X(k)dk$

$= a\int_{-\infty}^{\infty} X(k)f_X(k)dk + b\int_{-\infty}^{\infty} f_X(k)dk$

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

Variance

No surprises here

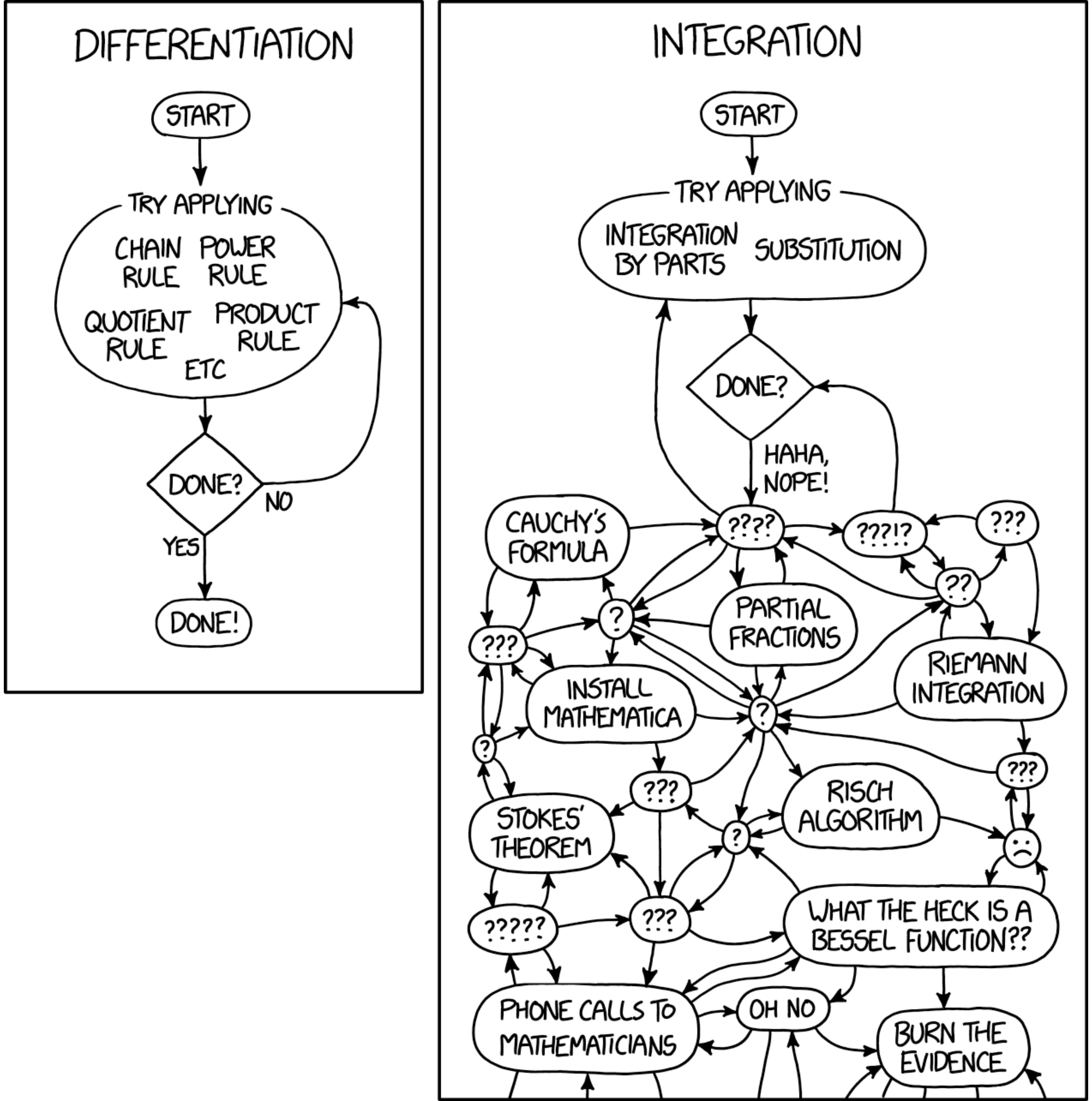
$$\mathbf{Var}(X) = \underbrace{\mathbb{E}[X^2]}_{\text{LOTUS}} - \underbrace{(\mathbb{E}[X])^2}_{\text{Def } \mathbb{E}[X]} = \int_{-\infty}^{\infty} f_X(k)(X(k) - \mathbb{E}[X])^2 \, \mathrm{d}k$$

Comparing Discrete and Continuous

	Discrete Random Variables	Continuous Random Variables
Probability 0	Equivalent to impossible	All impossible events have probability 0, but not conversely.
Relative Chances	PMF: $p_X(k) = \mathbb{P}(X = k)$	PDF: $f_X(k)$ gives chances relative to $f_X(k')$
Events	Sum over PMF to get probability	Integrate PDF to get probability
Convert between CDF and PMF	Sum up PMF to get CDF. Look for "breakpoints" in CDF to get PMF.	Integrate PDF to get CDF. Differentiate CDF to get PDF.
$\mathbb{E}[X]$	$\sum_{\omega} X(\omega) \cdot f_X(\omega)$	$\int_{-\infty}^{\infty} z \cdot f_X(z)\,\mathrm{d}z$
$\mathbb{E}[g(X)]$	$\sum_{\omega} g\big(X(\omega)\big) \cdot f_X(\omega)$	$\int_{-\infty}^{\infty} g(z) \cdot f_X(z)\,\mathrm{d}z$
$\mathrm{Var}(X)$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} (z - \mathbb{E}[X])^2 f_X(z)\,\mathrm{d}z$

e | ln

x^2



https://xkcd.com/2117/

Uniform PDF

$X \sim Unif(a, b)$

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

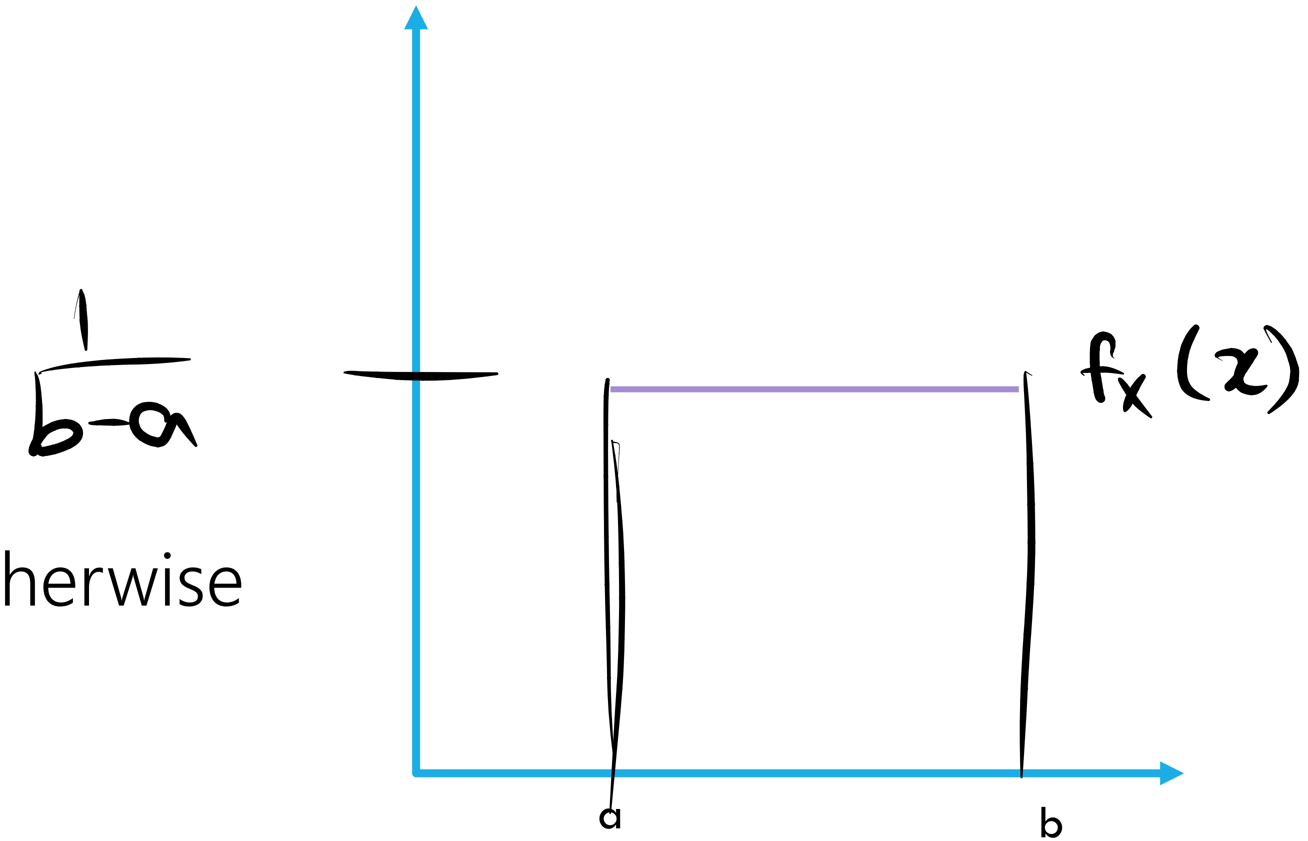
$\int_a^b f_X(x) dx = 1$

$(b-a) f_X(x) = 1$

$f_X(x) = \frac{1}{b-a}$ $for\ x \in [a, b]$, 0 otherwise

$f_X(x) \cdot (b-a) = 1$

$f_X(x) = \frac{1}{b-a}$



Uniform CDF

$$f_X(z) = \begin{cases} \frac{1}{b-a} & [a,b] \\ 0 & \text{o.w.} \end{cases}$$

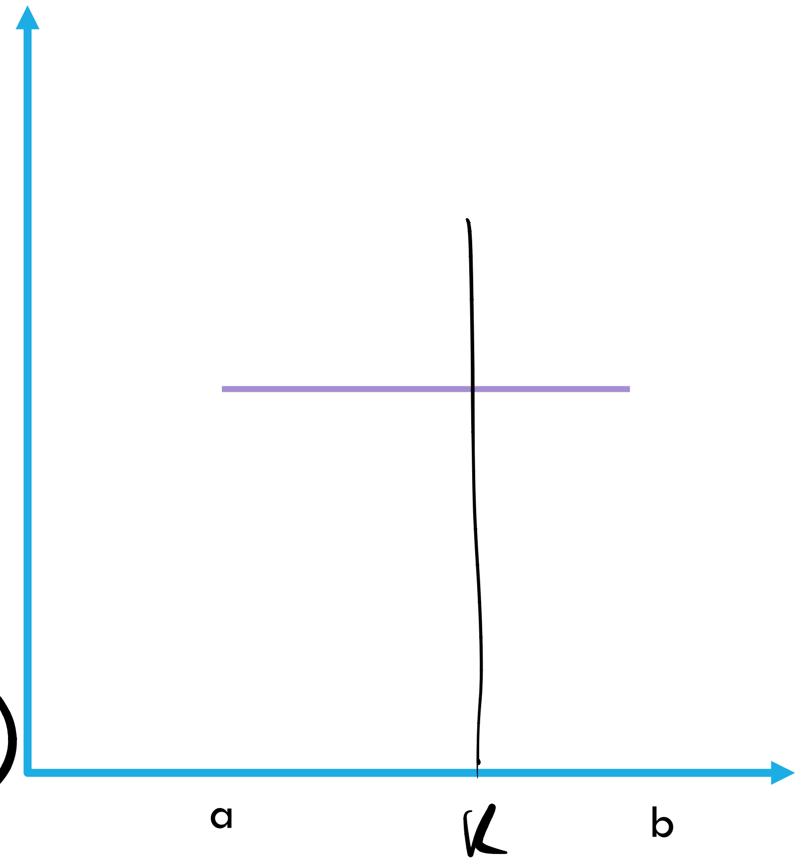
$X \sim Unif(a,b)$

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

$$= \int_{-\infty}^{k} f_X(z)\,dz$$

$$= \int_{-\infty}^{a} f_X(z)\,dz + \int_{a}^{k} f_X(z)\,dz$$

$$= \int_{a}^{k} f_X(z)\,dz = F_X(k) - F_X(a)$$



Uniform CDF

$X \sim Unif(a, b)$

$F_X(k) = \int_{-\infty}^{k} f_X(x)\, dx$

$= \int_a^k f_X(x) dx$ $=$

$k \cdot a \left(\frac{1}{b-a}\right)$

$F_X(x) = \frac{1}{b-a} \cdot (k-a) for\ x \in [a, b],$

$0\ for\ x < a, 1\ for\ x > b$

Let’s calculate an expectation

Let X be a uniform random number between a and b.

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

$$= \underbrace{\int_{-\infty}^{a} z f_X(z)\,dz}_{0} + \int_{a}^{b} z \underbrace{f_X(z)}_{\frac{1}{b-a}}\,dz + \underbrace{\int_{b}^{\infty} z f_X(z)\,dz}_{0}$$

$$= \int_a^b \frac{1}{b-a}\,dz = \frac{z^2}{2(b-a)}\Bigg|_a^b = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)}$$

Let's calculate an expectation

Let X be a uniform random number between a and b.

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

$$= \int_{-\infty}^{a} z \cdot 0\, \mathrm{d}z + \int_{a}^{b} z \cdot \frac{1}{b-a}\, \mathrm{d}z + \int_{b}^{\infty} z \cdot 0\, \mathrm{d}z$$

$$= 0 + \int_{a}^{b} \frac{z}{b-a}\, \mathrm{d}z + 0$$

$$= \left.\frac{z^2}{2(b-a)}\right|_{z=a}^{b} = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \boxed{\frac{a+b}{2}}$$

What about $\mathbb{E}[g(X)]$

Let $X \sim \text{Unif}(a, b)$, what about $\mathbb{E}[X^2]$?

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} z^2 f_X(z)\,\mathrm{d}z$$

$$= \int_{-\infty}^{a} z^2 \cdot 0\,\mathrm{d}z + \int_{a}^{b} z^2 \cdot \frac{1}{b-a}\,\mathrm{d}z + \int_{b}^{\infty} z^2 \cdot 0\,\mathrm{d}z$$

$$= 0 + \underbrace{\int_{a}^{b} z^2 \cdot \frac{1}{b-a}\,\mathrm{d}z}_{} + 0$$

$$\frac{1}{(b-a)} \frac{z^3}{3} \Big|_a^b$$

What about $\mathbb{E}[g(X)]$

Let $X \sim \mathrm{Unif}(a, b)$, what about $\mathbb{E}[X^2]$?

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} z^2 f_X(z) \mathrm{dz}$$

$$= \int_{-\infty}^{a} z^2 \cdot 0 \ \mathrm{dz} + \int_{a}^{b} z^2 \cdot \frac{1}{b-a} \mathrm{dz} + \int_{b}^{\infty} z^2 \cdot 0 \ \mathrm{dz}$$

$$= 0 + \int_{a}^{b} z^2 \cdot \frac{1}{b-a} \mathrm{dz} + 0$$

$$= \frac{1}{b-a} \cdot \frac{z^3}{3}\bigg|_{z=a}^{b} = \frac{1}{b-a}\left(\frac{b^3}{3} - \frac{a^3}{3}\right) = \frac{1}{3(b-a)} \cdot (b-a)(a^2+ab+b^2)$$

$$= \frac{a^2+ab+b^2}{3}$$

Let’s assemble the variance

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{4(a^2+ab+b^2)}{12} - \frac{3(a^2+2ab+b^2)}{12}$$

$$= \frac{a^2-2ab+b^2}{12}$$

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

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

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

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