CSCI 3022

intro to data science with probability & statistics

Lecture 9
September 27, 2017

1. Continuous random variables
   - Intuition
   - Uniform
   - Normal
   - Exponential

HW #2 due Friday!
- a few notes on Piazza
- pls get most current game_log.csv from github.

Department of Computer Science
UNIVERSITY OF COLORADO BOULDER

Dan Larremore

Last time on CSCI 3022:

- **Def**: a discrete random variable X is a function that maps the elements of the sample space Ω to a finite number of values $a_1, a_2, \ldots, a_k$, or an infinite number of values $a_1, a_2, \ldots$

- **Def**: a probability mass function (PMF) is the map between the random variable's values and the probabilities of those values. Outcomes have masses.

$$\underbrace{f_X(a)}_{\text{PMF}} = P(X = a)$$

- **Def**: a cumulative distribution function (CDF) is a function whose value at a point *a* is the cumulative sum of probability masses up until *a*.

$$F_X(a) = P(X \leq a) = \sum_{x \leq a} f_X(\cancel{a})$$

PMF

Continuous random variables

- Many real-life random processes must be modeled by random variables that can take on continuous (i.e. not discrete) values. Some examples include:
  - people's heights: $(0, \infty)$
  - final grades in a course: $[0, 100]$
  - the time between buses arriving at the stop: $(0, \infty)$
- What are some other examples?

Frequencies of sound $(0, \infty)$

Super G times $[t_{min}, \infty)$

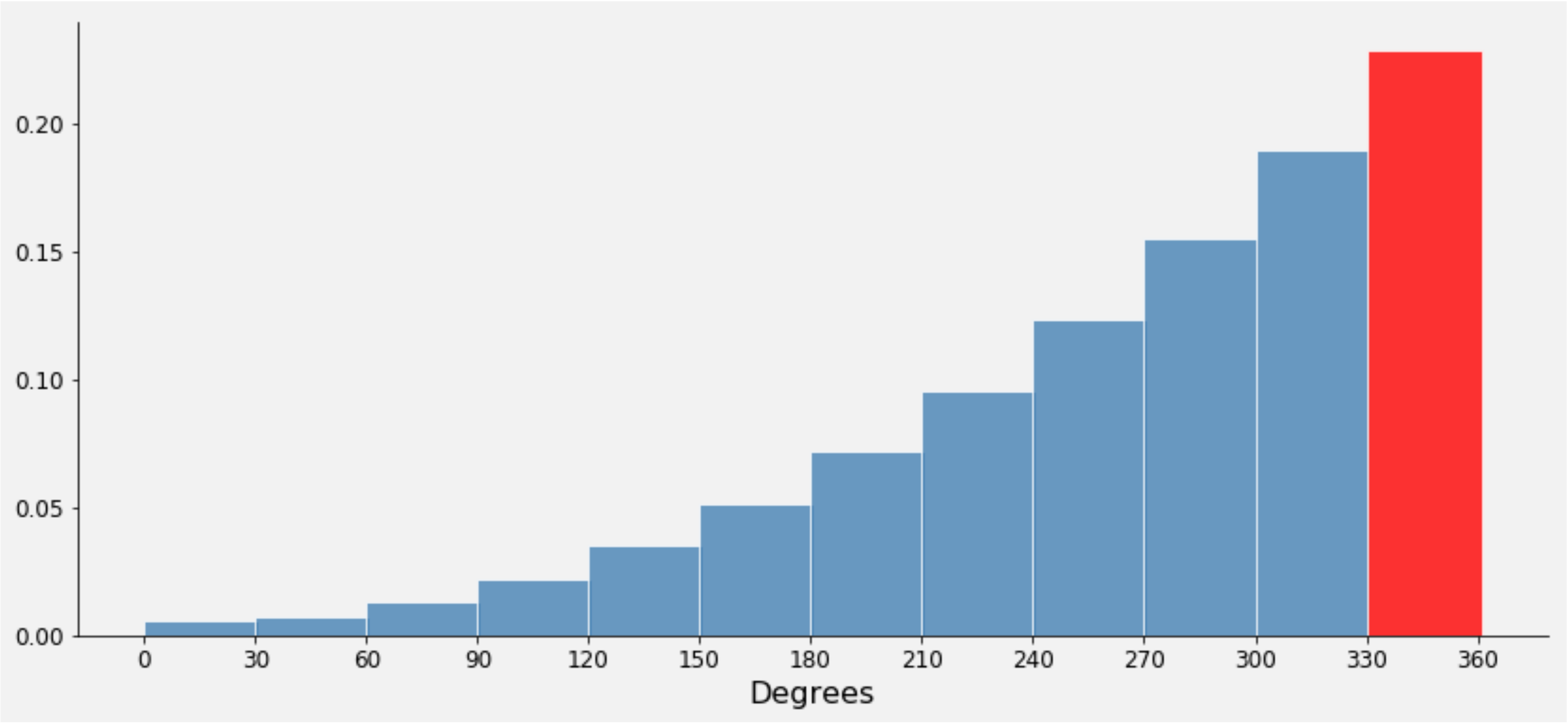
Temperature K $[0, \infty)$

Intuition pump

- **Example**: Suppose you spin the wheel on a game show. Unfortunately the wheel is in **disrepair** and the closer it gets to 360 Degrees the more likely it is to stop! Let X be the random variable describing the angle in Degrees at which the wheel stops.

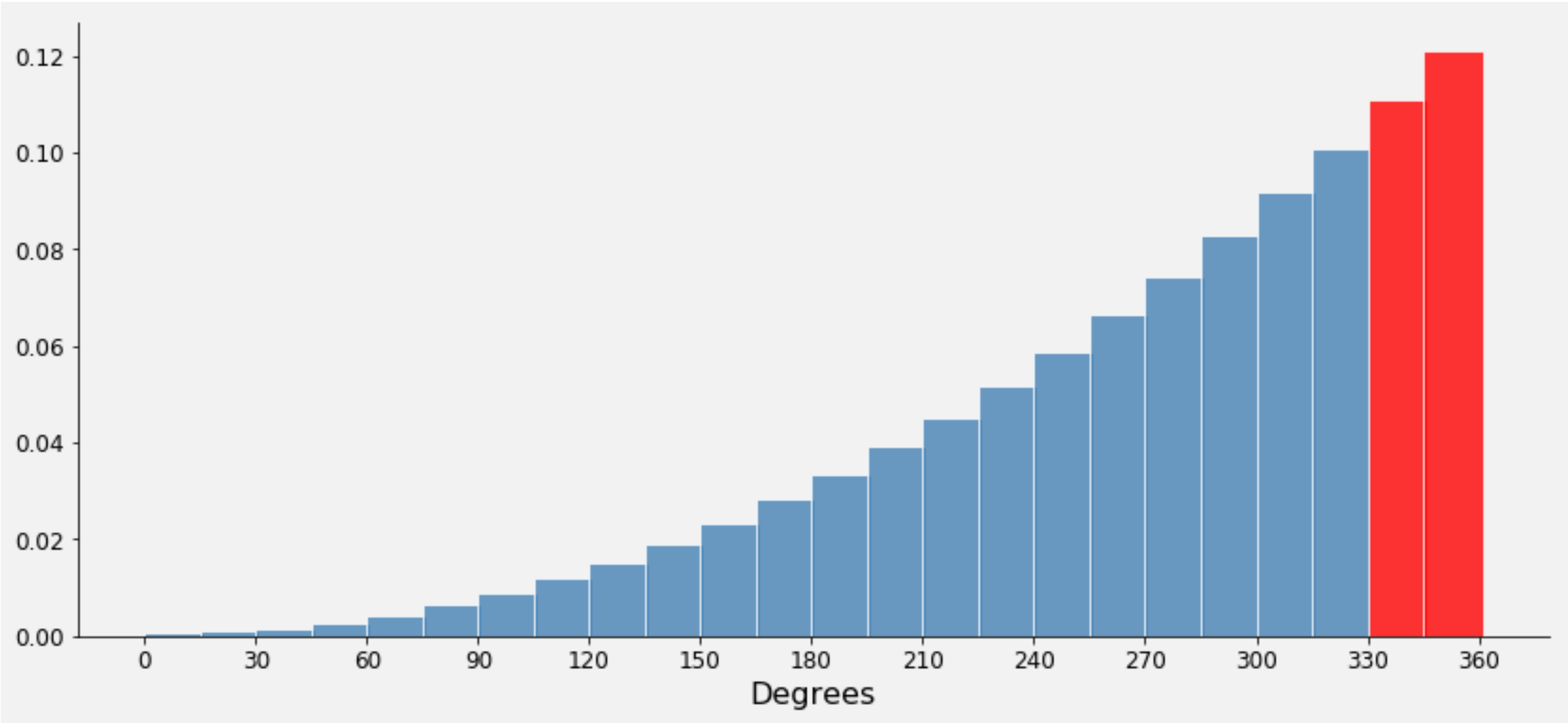
Intuition pump

- **Example**: Suppose you spin the wheel on a game show. Unfortunately the wheel is in **disrepair** and the closer it gets to 360 Degrees the more likely it is to stop! Let X be the random variable describing the angle in Degrees at which the wheel stops.



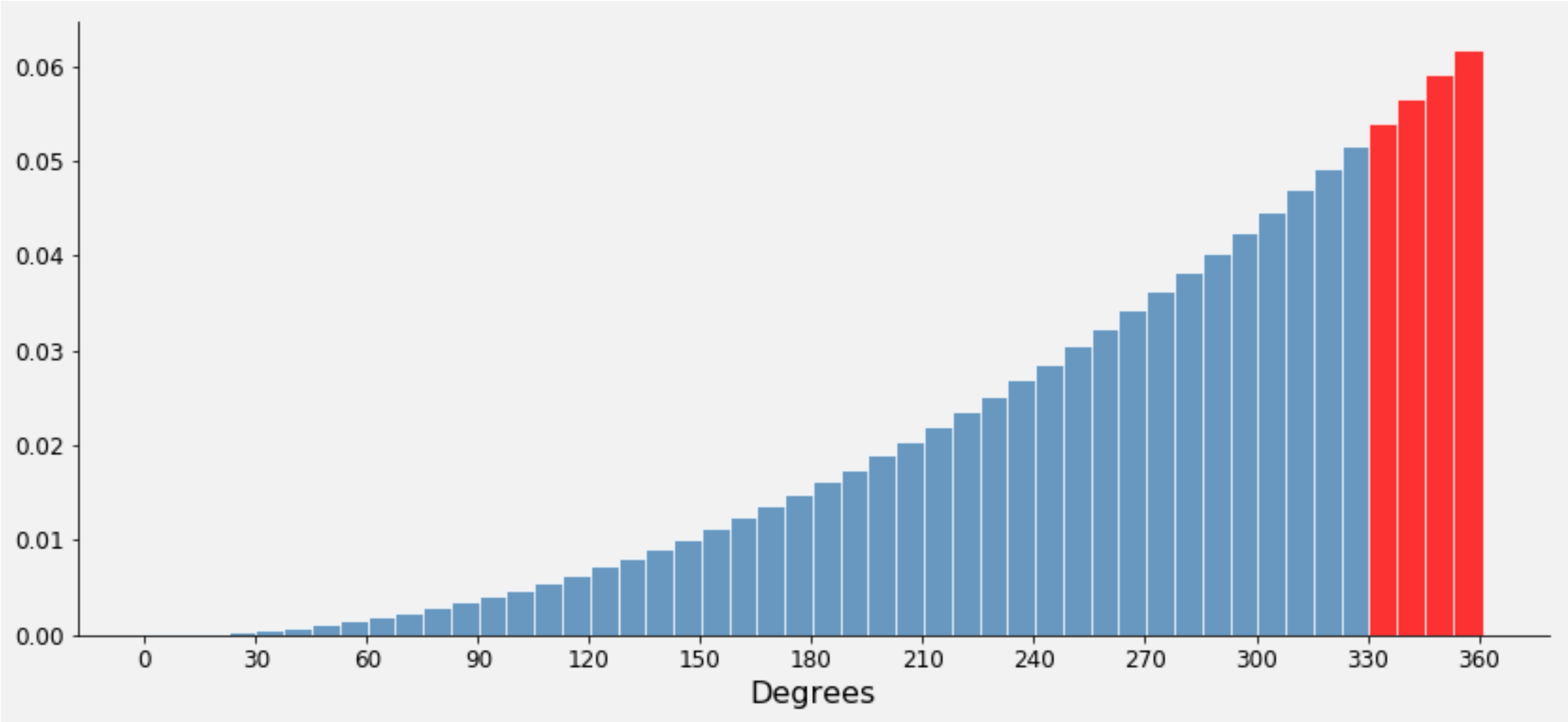
Intuition pump

- **Example**: Suppose you spin the wheel on a game show. Unfortunately the wheel is in **disrepair** and the closer it gets to 360 Degrees the more likely it is to stop! Let X be the random variable describing the angle in Degrees at which the wheel stops.



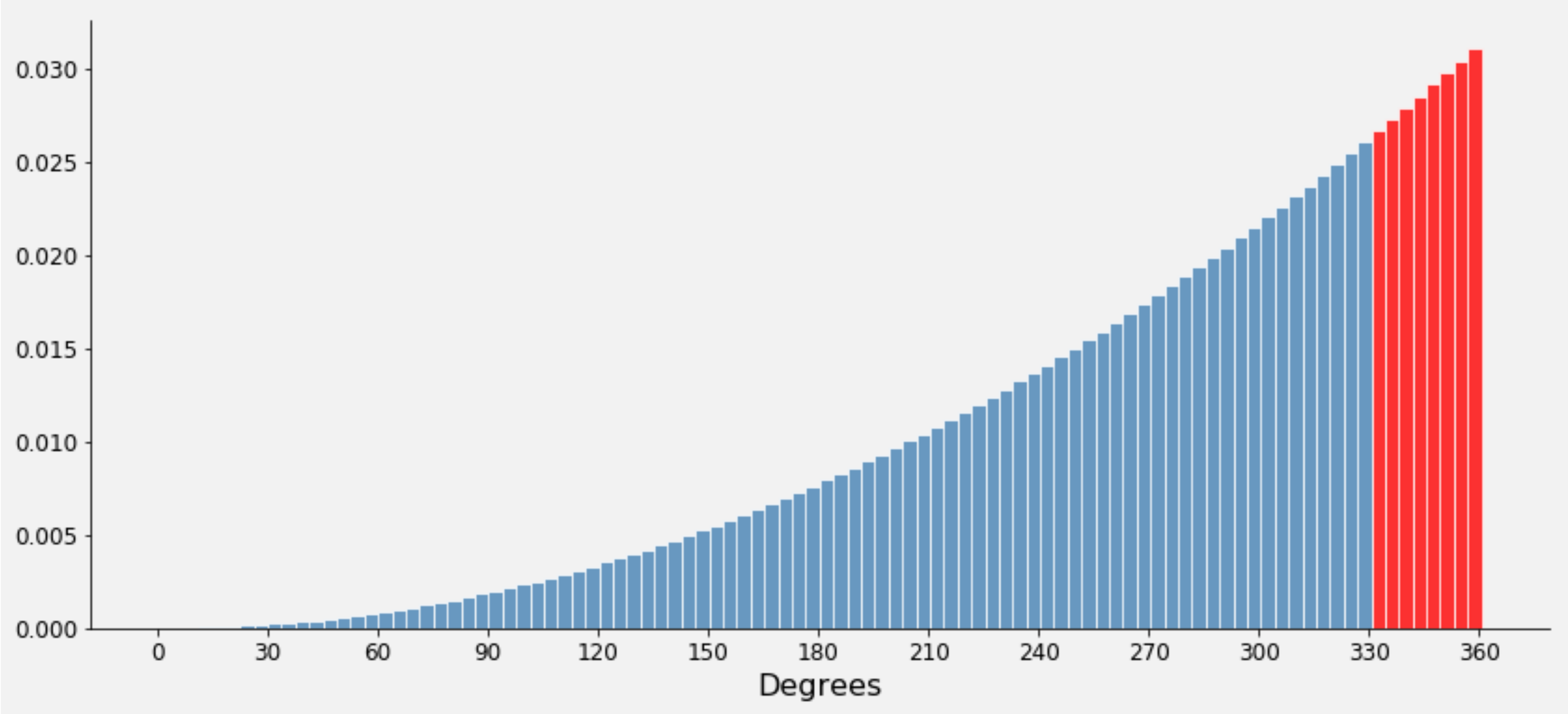
Intuition pump

- **Example**: Suppose you spin the wheel on a game show. Unfortunately the wheel is in **disrepair** and the closer it gets to 360 Degrees the more likely it is to stop! Let X be the random variable describing the angle in Degrees at which the wheel stops.



Intuition pump

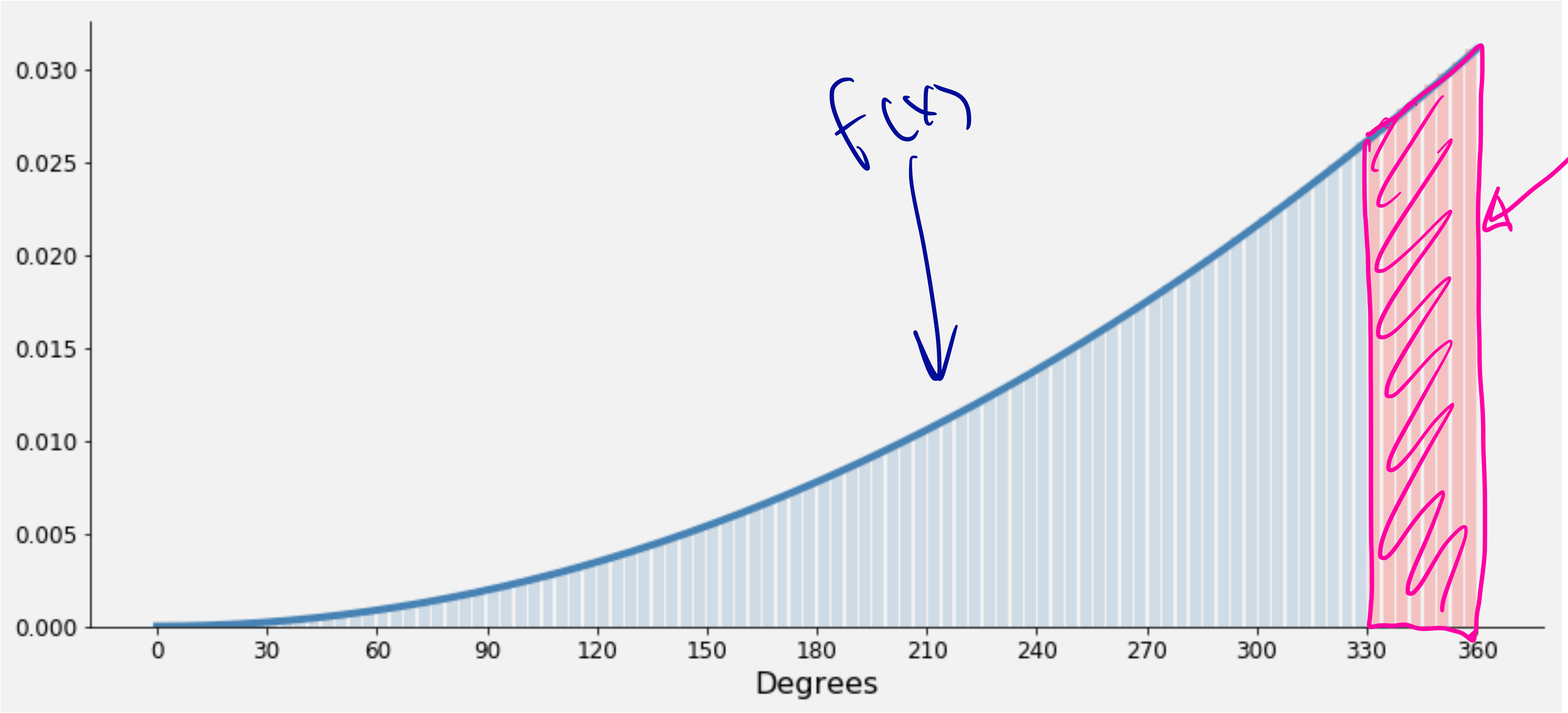
- **Example**: Suppose you spin the wheel on a game show. Unfortunately the wheel is in **disrepair** and the closer it gets to 360 Degrees the more likely it is to stop! Let X be the random variable describing the angle in Degrees at which the wheel stops.



Intuition pump

- **Example**: Suppose you spin the wheel on a game show. Unfortunately the wheel is in **disrepair** and the closer it gets to 360 Degrees the more likely it is to stop! Let X be the random variable describing the angle in Degrees at which the wheel stops.

$P(330 \le X \le 360) = \text{area}$



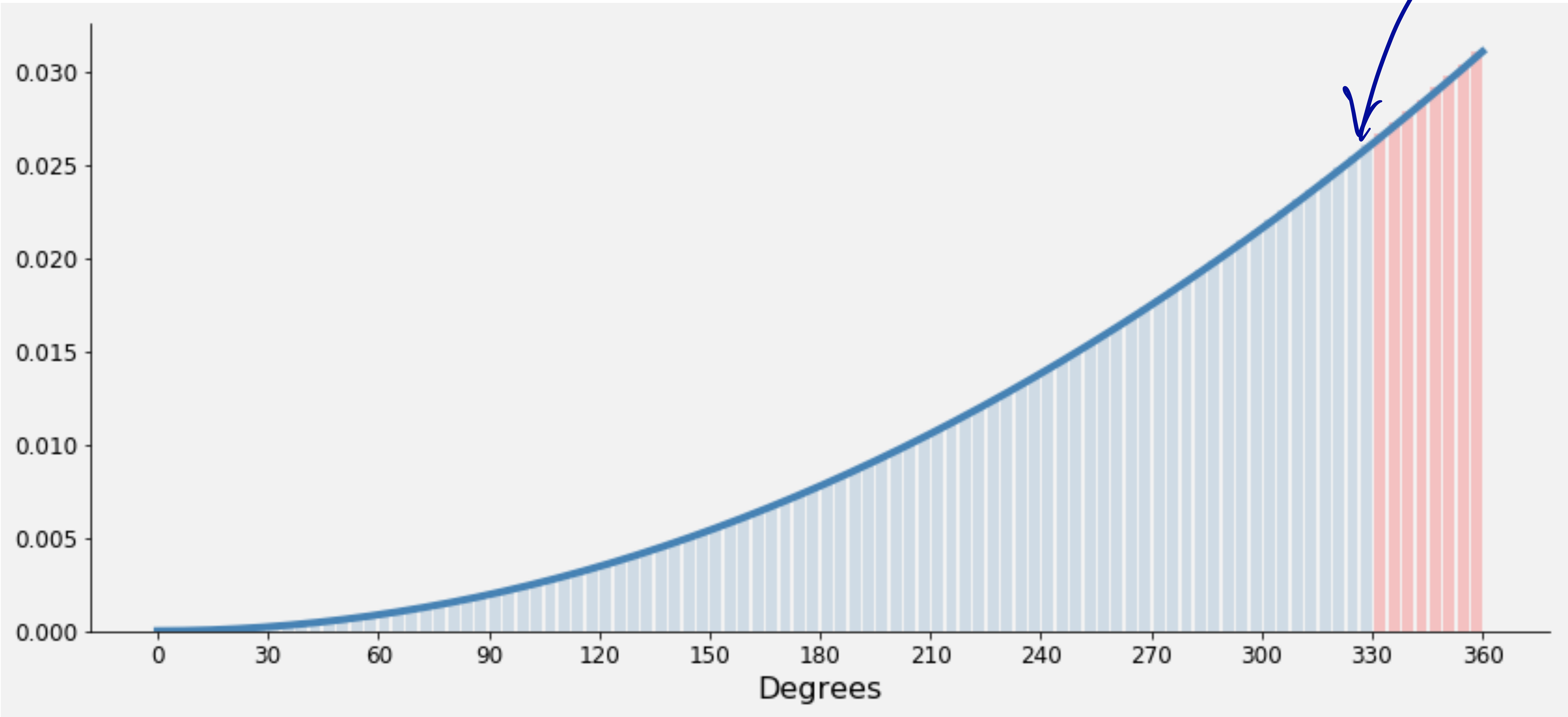
Intuition pump

- The probability looks like it's the area under a curve!

$$P(330 \leq X \leq 360) = \int_{330}^{360} f(x)dx$$

- Somehow, f(x) is counting up the amount of probability "mass" in each little segment dx…

f(x)
probability density function (PDF)



Probability density function (PDF)

- **Def**: A random variable *X* is continuous if for some function f(x) and for any numbers a and b, with a≤b

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

- We call f(x) the *probability density function* and it has two requirements:

1. $f(x) \geq 0 \quad \forall x \qquad$ \forall
2. $P(\Omega) = 1 \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$

Probability density function (PDF)

- **Back to the wheel**: suppose you spin the wheel. What's the probability that it stops at a particular value,

$$P(X = 30.57534) = 0$$

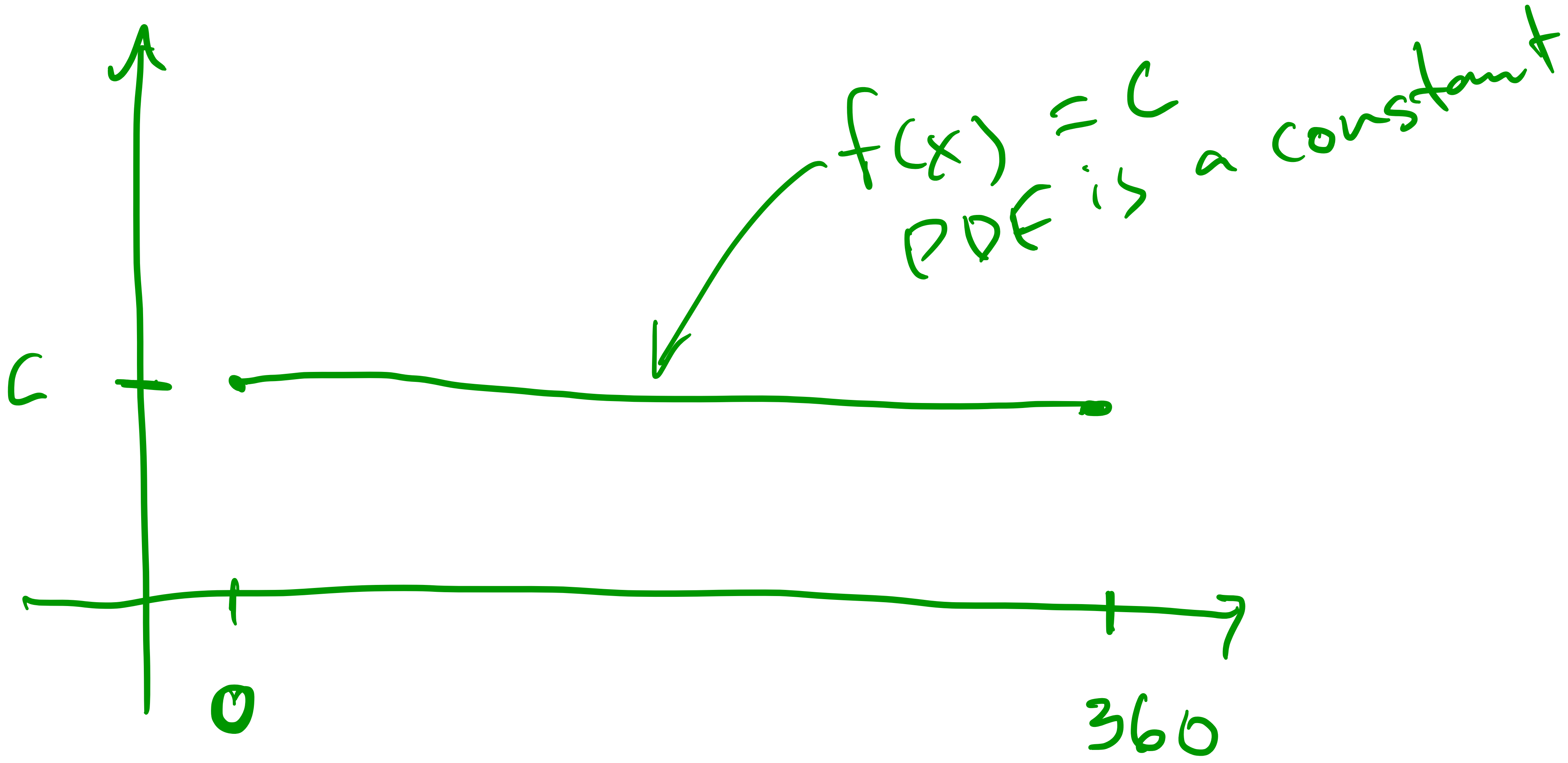
$$P(X = a) = 0$$

$$P(a - \varepsilon \leq X \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x)\,dx$$

Think about $\lim_{\varepsilon \to 0}$

Continuous uniform distribution

- **Fix that wheel**: you oil the wheel and now the probability that it stops on any particular angle is equally likely. What is the probability density function for the angle X?



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

$$\int_{0}^{360} f(x)\,dx = 1$$

$$c(360-0) = cx\Big|_{0}^{360} = \int_{0}^{360} c = 1$$

$$\boxed{c = \frac{1}{\beta-\alpha} = f(x)}$$

$$c = \frac{1}{360-0}$$

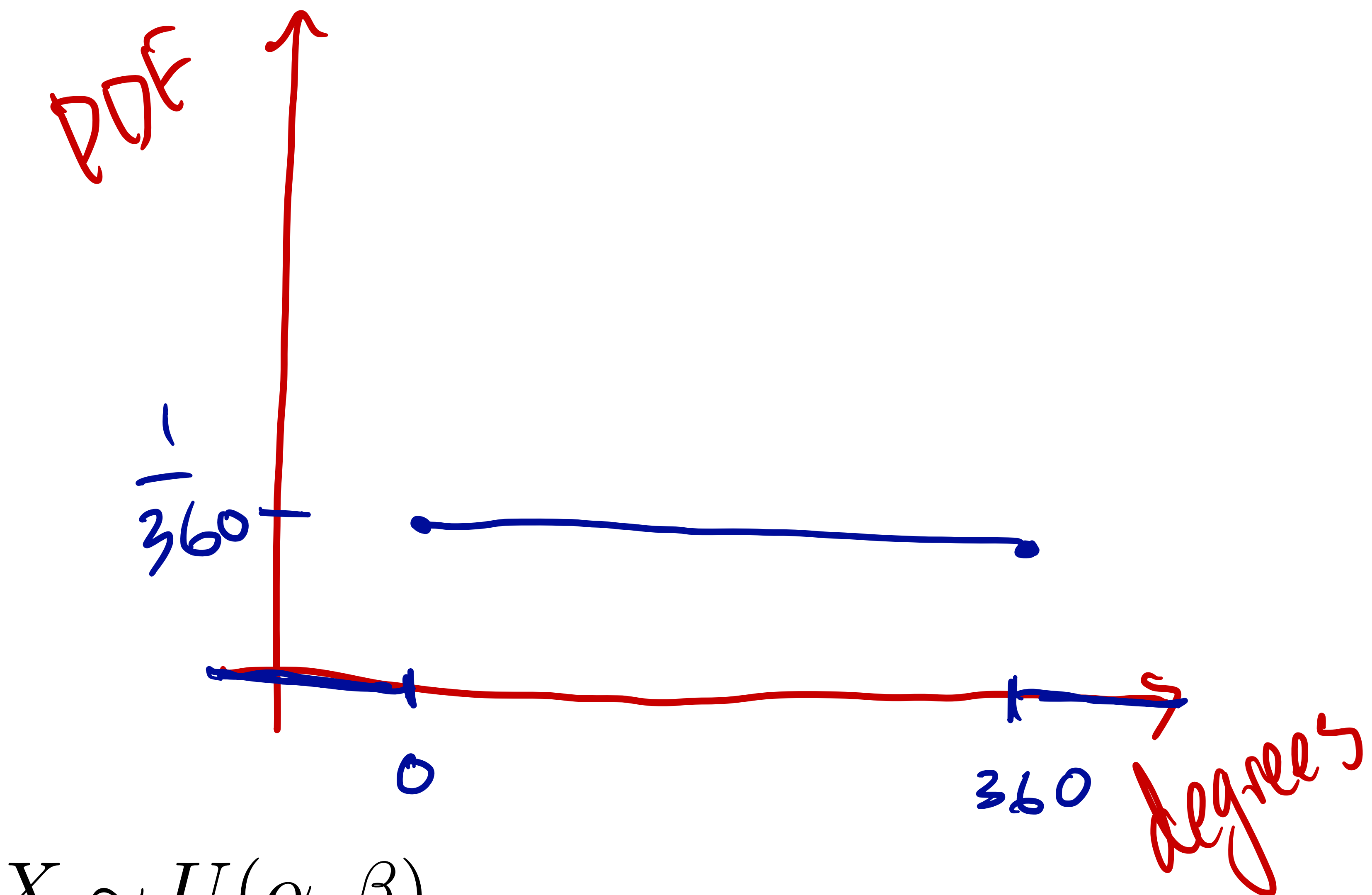
Continuous uniform distribution

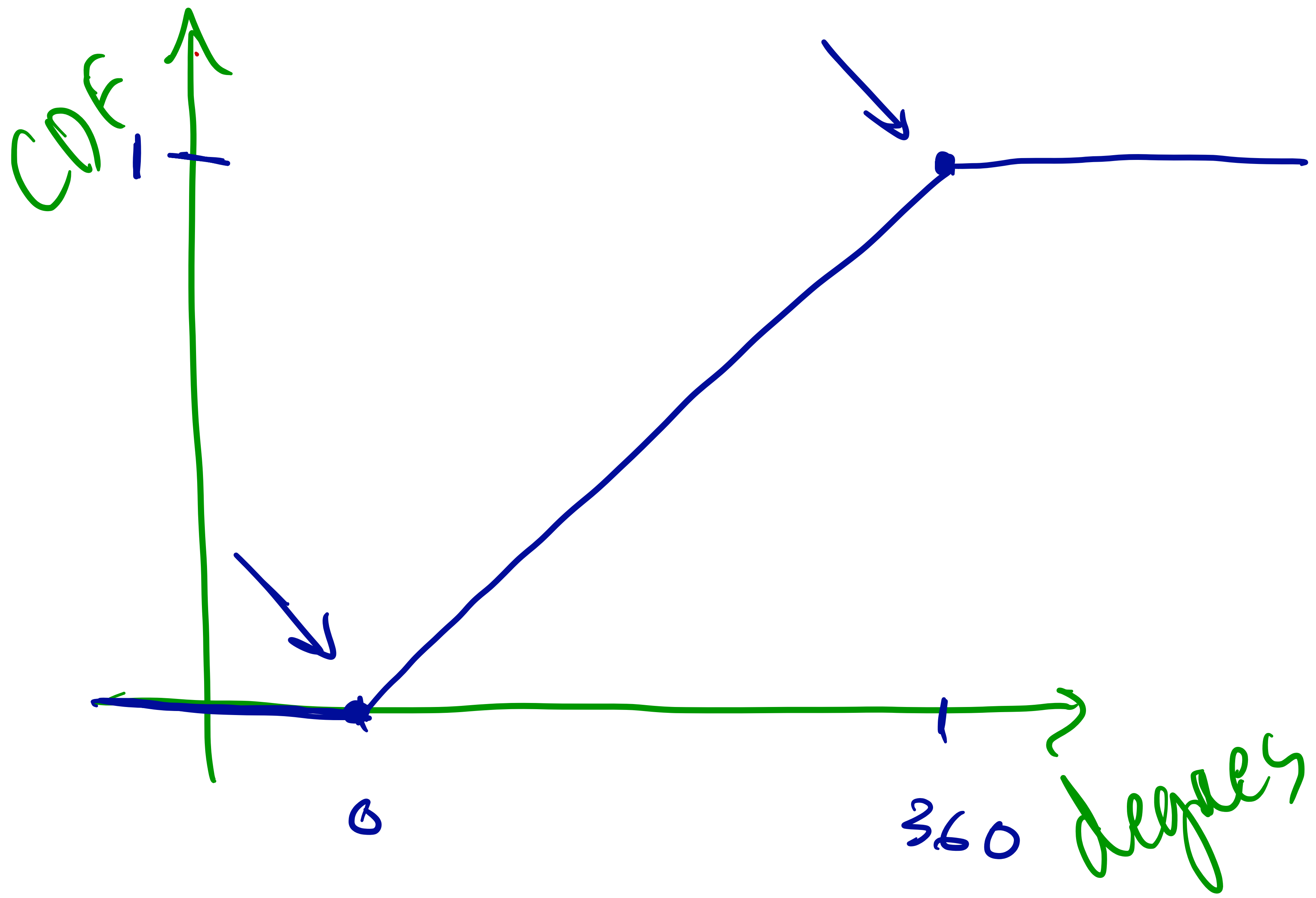
- **Def**: A continuous random variable has a uniform distribution on the interval [α, β] if its probability density function f(x) is

$$f(x) = \frac{1}{\beta - \alpha} \quad \text{for} \quad \alpha \le x \le \beta \quad \text{and} \quad 0 \quad \text{otherwise}$$

- **Challenge**: write the PDF for the wheel, and then plot the PDF & CDF:

The PDF is: $f(x) = \frac{1}{360-0} = \frac{1}{360}$





$X \sim U(\alpha, \beta)$

Reiterate: density

- We only end up with probability when we integrate the density over an interval. The probability of any particular value is zero.

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

look up a delta function δ

- If we send $\epsilon \to 0$ then $P(X = a) = 0$ for any *a* and for [almost] any *f*!

- Get loose:

$$P(a < x < b) = P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b)$$

$X \sim U(\alpha, \beta)$

Cumulative distribution functions (CDFs)

PMF

- Recall: discrete RVs don't have a probability density function.

PDF

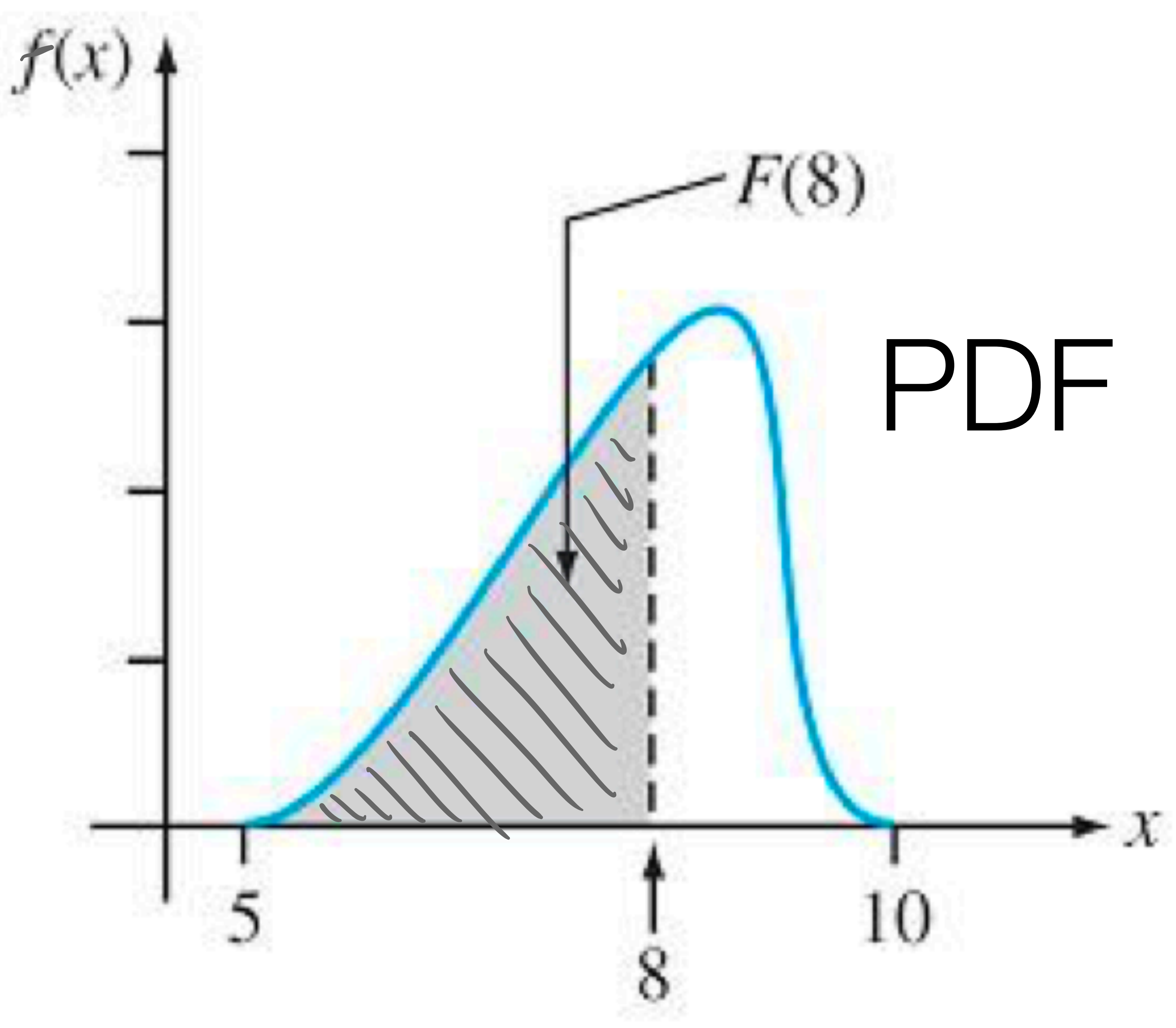
- Recall: continuous RVs don't have a probability mass function.
- And yet! Both have a CDF: $F_X(a) = \underline{P(X \leq a)}$
- What is the CDF for a discrete RV? $F_X(a) = \sum_{x \leq a} f(x)$
- What is the CDF for a continuous RV?

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

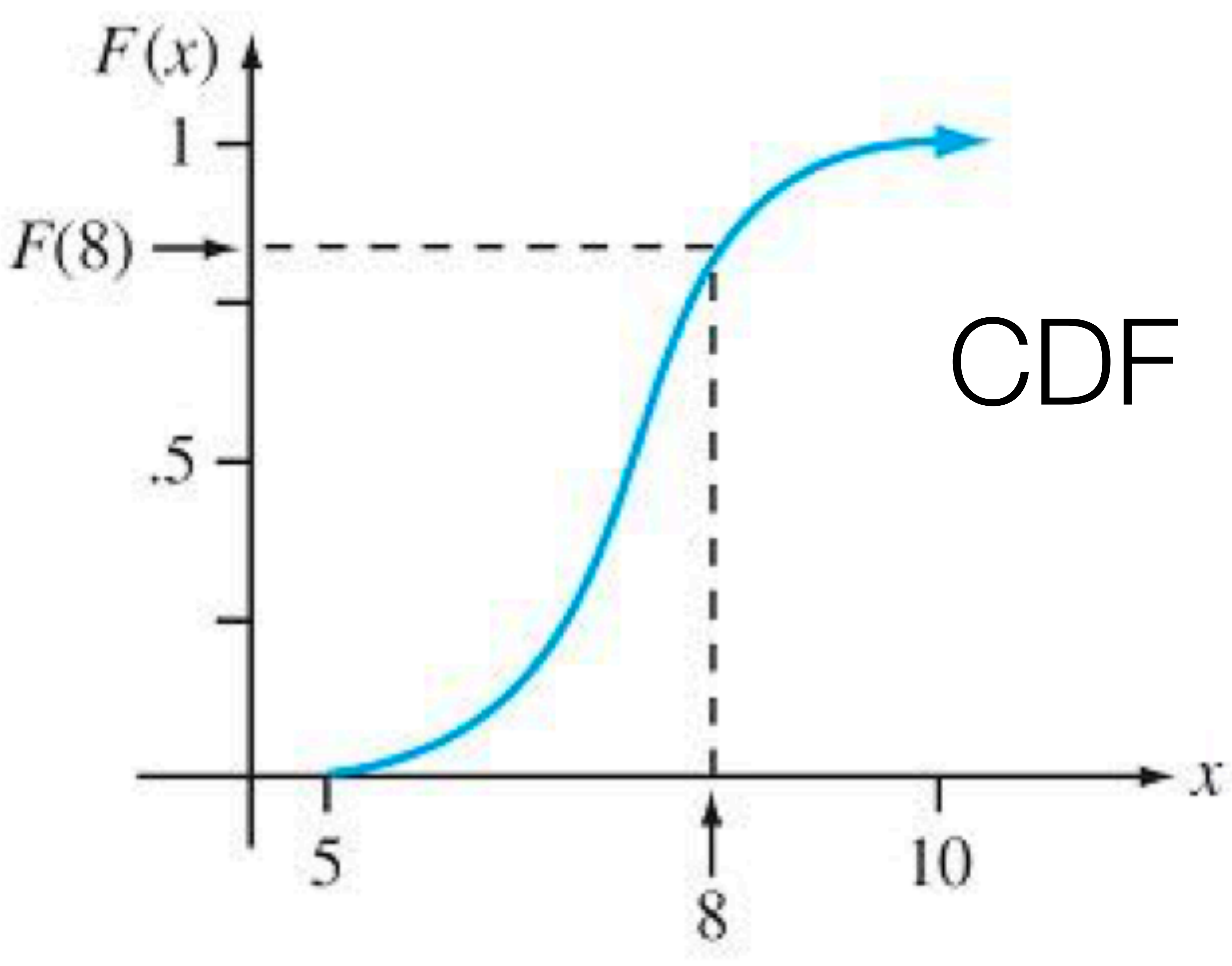
Cumulative distribution functions (CDFs)

- We can use the CDF to compute things like $P(X \leq a)$

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



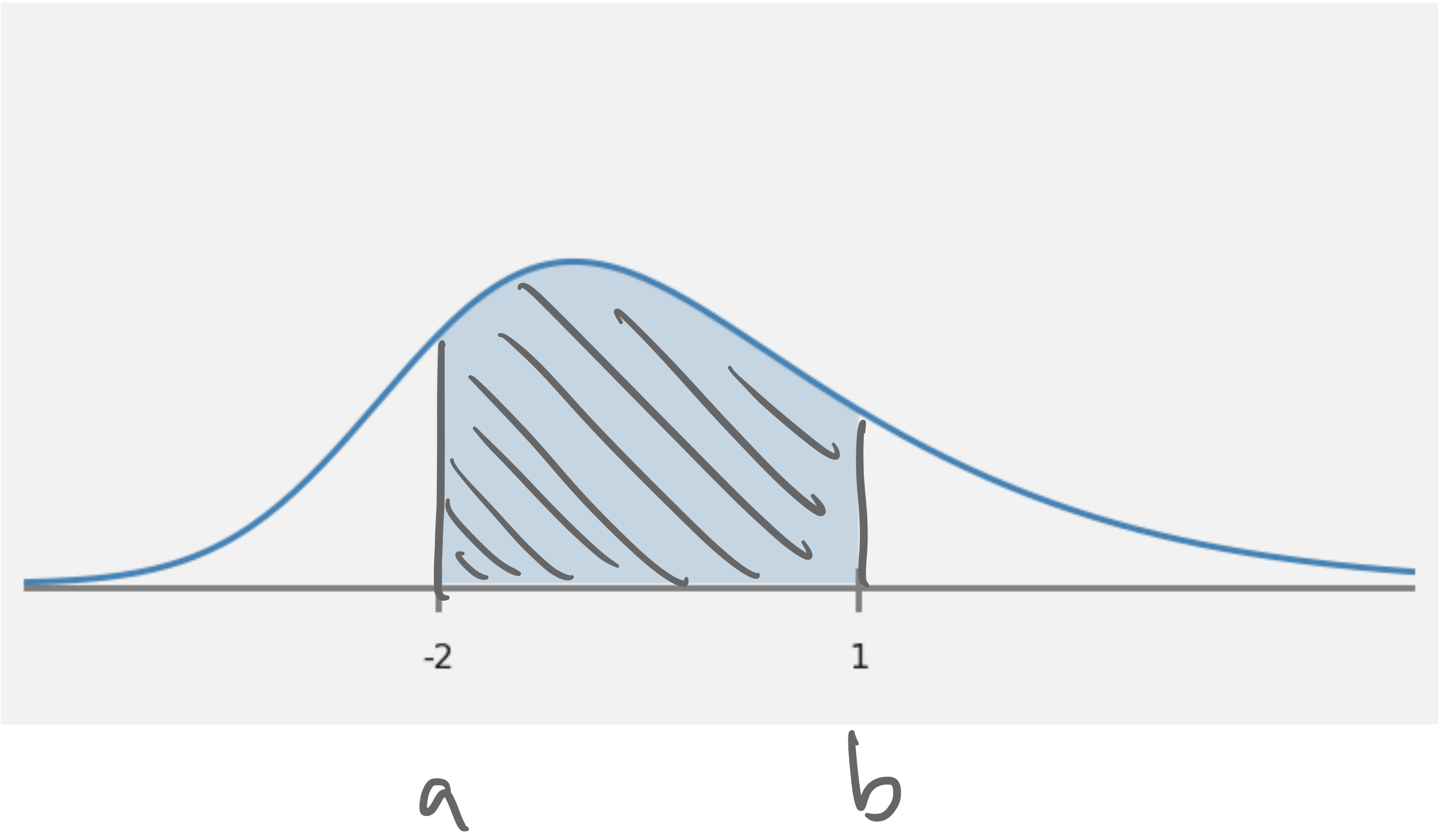
PDF



CDF

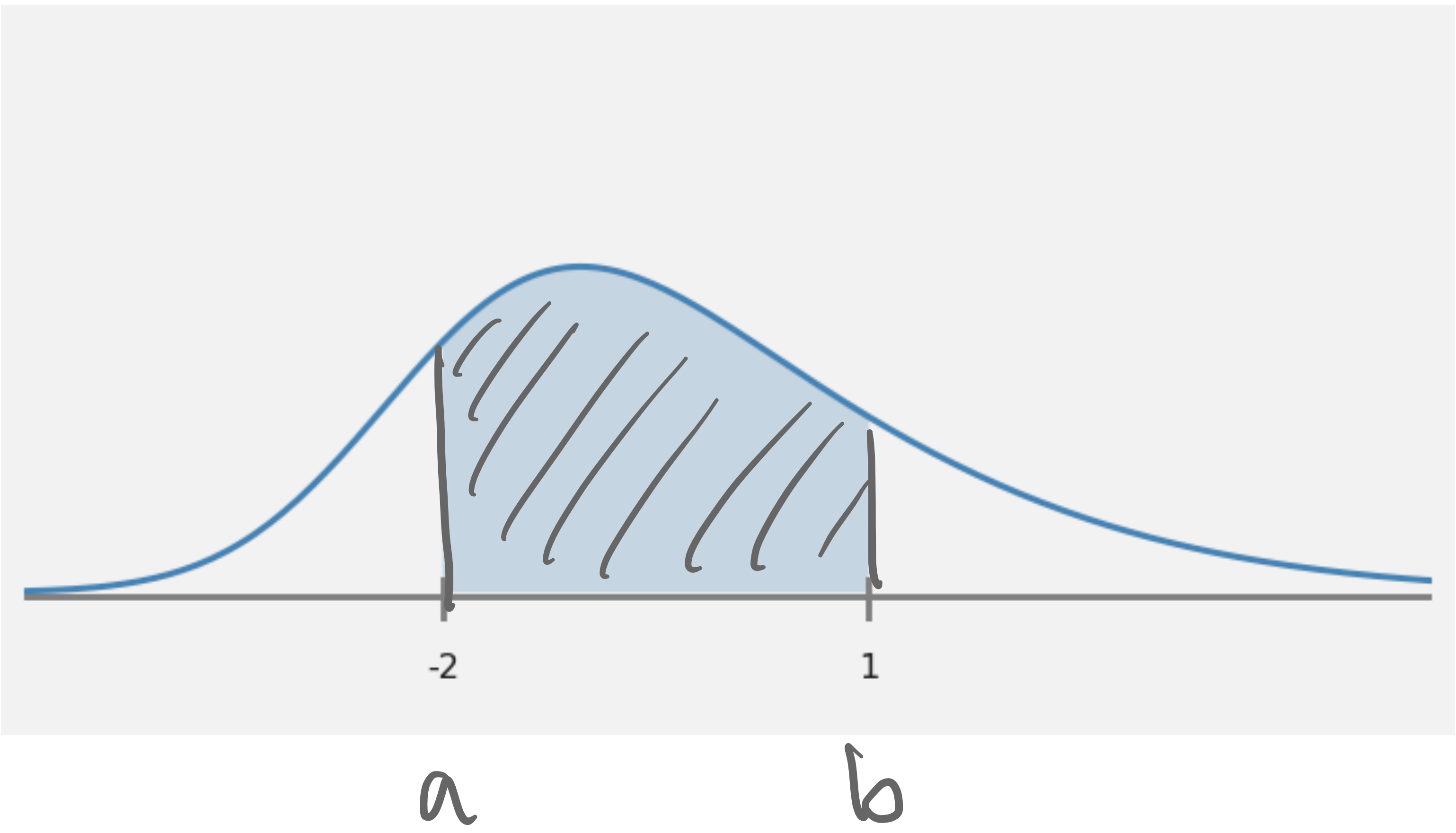
Cumulative distribution functions (CDFs)

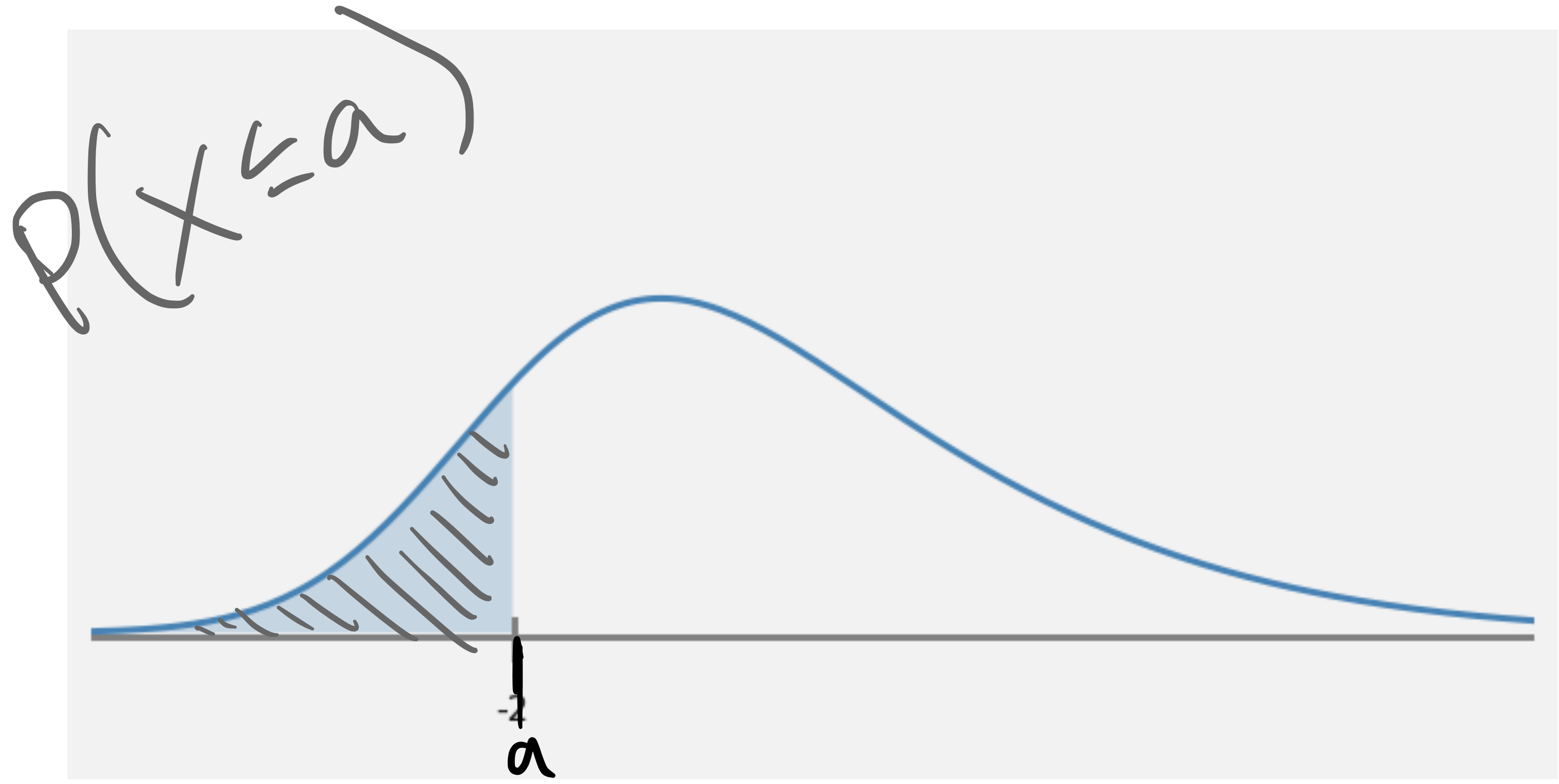
- What about things like $P(a \leq X \leq b)$? Like the probability of spinning red?

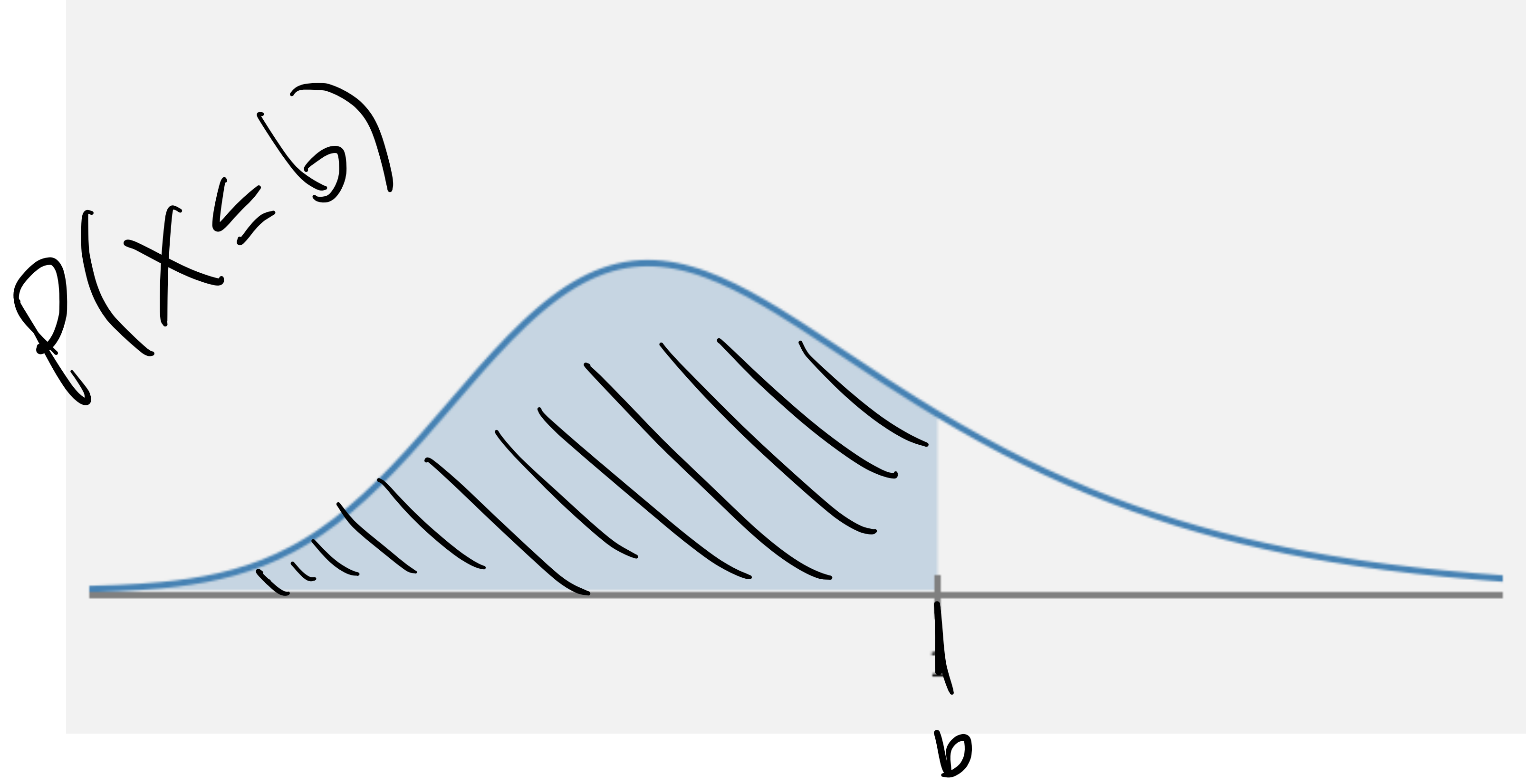


Cumulative distribution functions (CDFs)

- What about things like $P(a \leq X \leq b)$? Like the probability of spinning red?







$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

$$= \int_{-\infty}^{b} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt = \int_{a}^{b} f(t)\,dt$$

Hello, old friend! 😘

- The relationship between the PDF f(x), the CDF F(x), and probability:

$$P(a \leq X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$$ Calc I ^_^;"

- Does this remind you of anything?

F is the antiderivative of f

$$\underline{1 - P(X \leq a)} = 1 - F(a) = \int_{-\infty}^{\infty} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt = \int_{a}^{\infty} f(t)\,dt = \underline{P(X > a)}$$

Hello, old friend! 😘

- The relationship between the PDF f(x), the CDF F(x), and probability:

$$P(a \leq X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$$

- Does this remind you of anything?

$$f(x) = \frac{d}{dx}F(x)$$

at any point where F(x) is differentiable

- F and f contain the same information!

Hello, new friend... 😳

- **Definition**: a continuous random variable has a *normal distribution* with parameters μ and σ^2 if its probability density function f is given by

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

μ, σ^2
change these
see what happens

- Let's explore! https://academo.org/demos/gaussian-distribution/

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

Hello, new friend... 😳

- **Definition**: a continuous random variable has a *normal distribution* with parameters μ and σ^2 if its probability density function f is given by

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

- Let's explore! https://academo.org/demos/gaussian-distribution/

- No closed-form function for F(x). But… with a little magic, we can turn *any* normal distribution into N(0,1), which we call a *standard* normal distribution.

To be continued!

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

From last time

- **Question**: suppose you get texts during class at an average rate of 200 per hour lol. If every instance during class has the same probability of a text arriving, we learned that P(X=k) texts during class is X~Pois(200). *But now:* **What is the distribution of times t between text arrivals??** $\in [0, \infty)$

PDF → $\boxed{\text{CDF}}$ trick!

How long until the first text comes in?

Now it's t

Wait until $t + \Delta t$

$$P(X \leq \Delta t) = 1 - P(X > \Delta t)$$

$$= 1 - P(N_{\text{texts in } t+\Delta t} - N_{\text{texts } t} = 0)$$

$= 1 -$ Prob that I get 0 texts in Δt

From last time

- **Question**: suppose you get texts during class at an average rate of 200 per hour lol. If every instance during class has the same probability of a text arriving, we learned that P(X=k) texts during class is X~Pois(200). *But now:* **What is the distribution of times t between text arrivals??**

$P(X \leq \Delta t) = 1 -$ Prob 0 texts in Δt window

$$= 1 - \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t} \Big|_{k=0}$$

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

$$\boxed{P(k) = \frac{\lambda^k}{k!} e^{-\lambda}}$$

Poisson

derivative

$$\boxed{f(x) = \lambda e^{-\lambda x}}$$

exponential distribution.