intro to data science with probability & statistics

CSCI 3022

September 26, 2018

HW #2
Due Fri 5PM

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

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 $a_1, a_2, \dots, a_k$, or an infinite number of values $a_1, a_2, \dots$

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

$$f_X(a) = P(X = a)$$

- **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*.

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

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?

HW duration $(0, \infty)$

Frequencies of Sound

Temperatures in October $[t_{min}, \infty)$

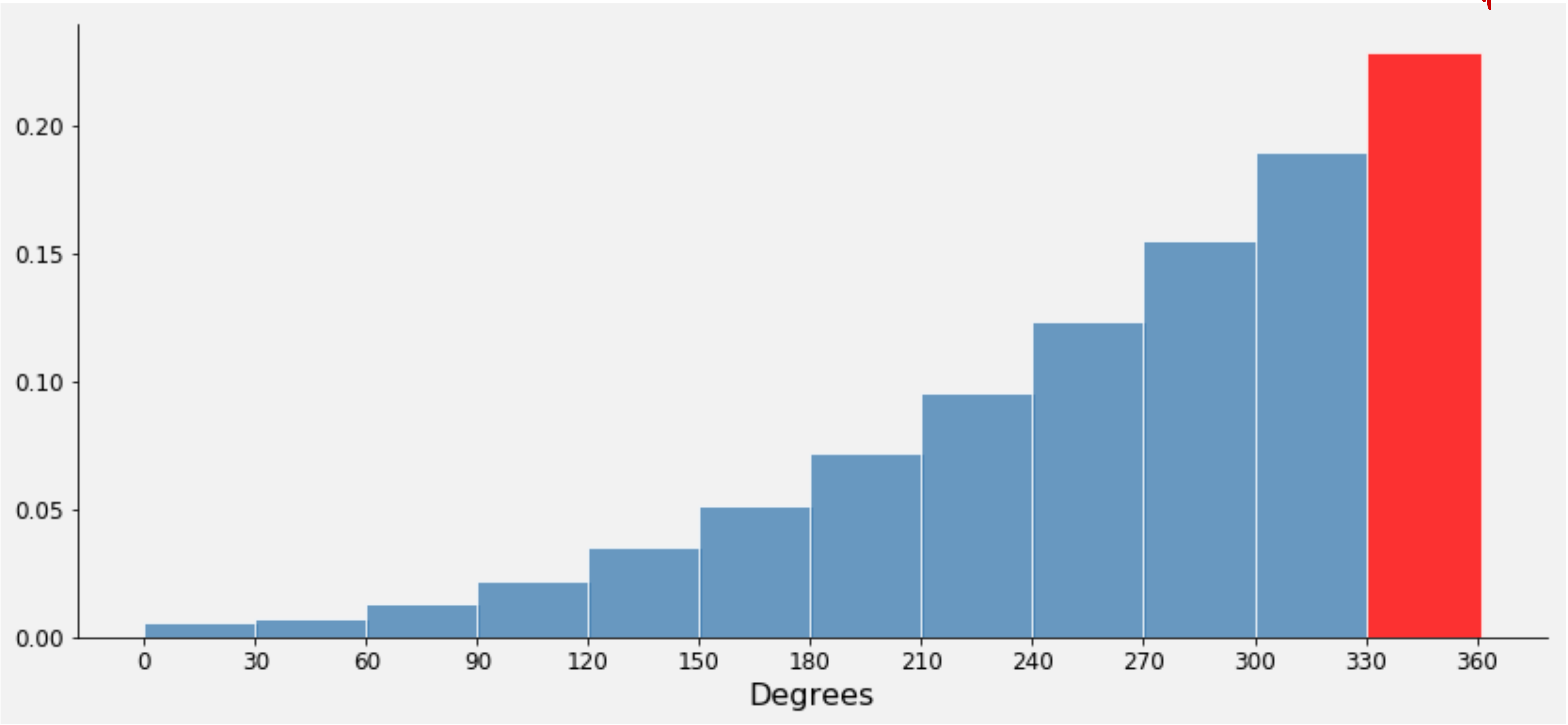
Spectrum of color

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.

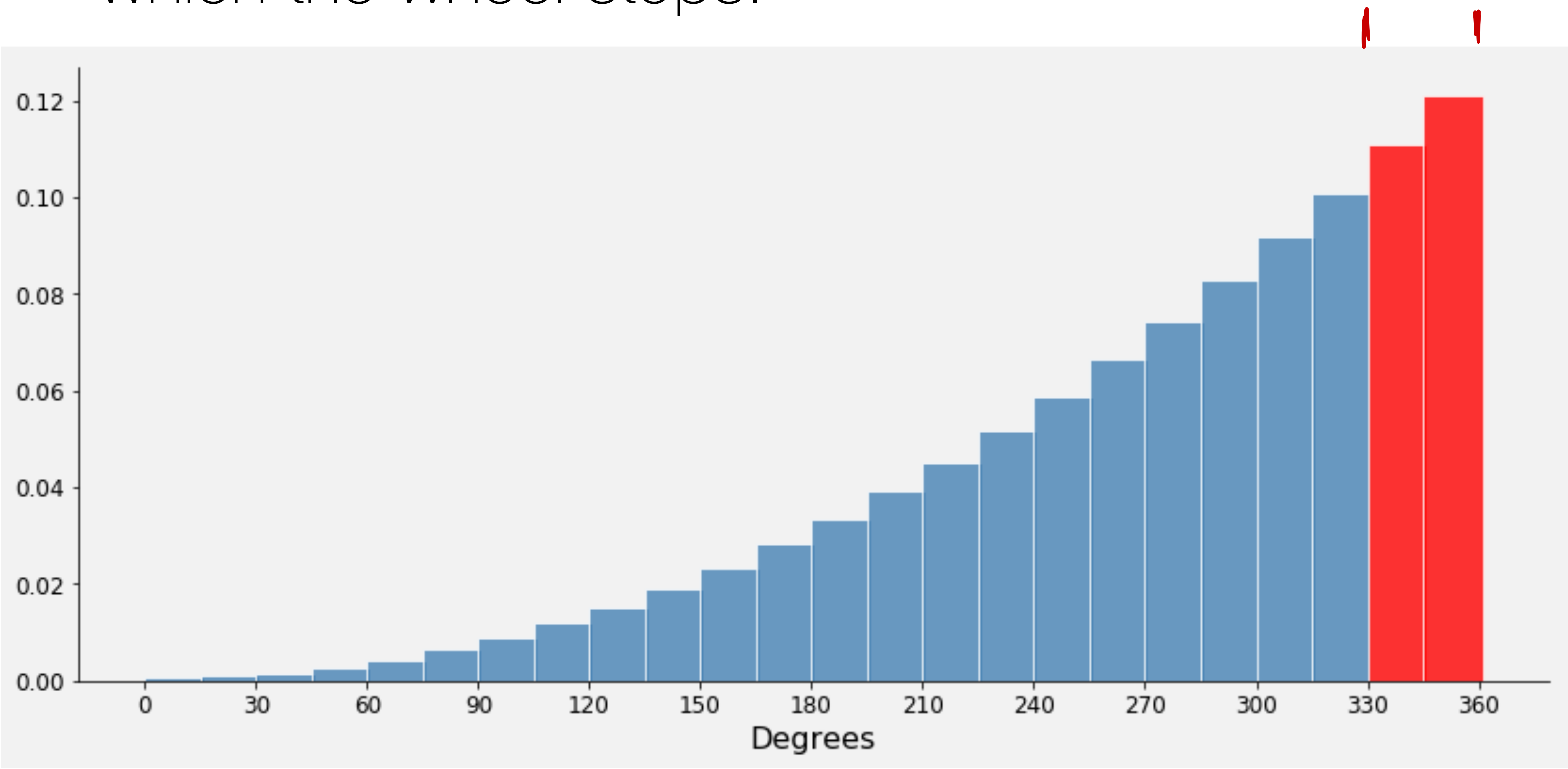
Intuition pump

- **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.



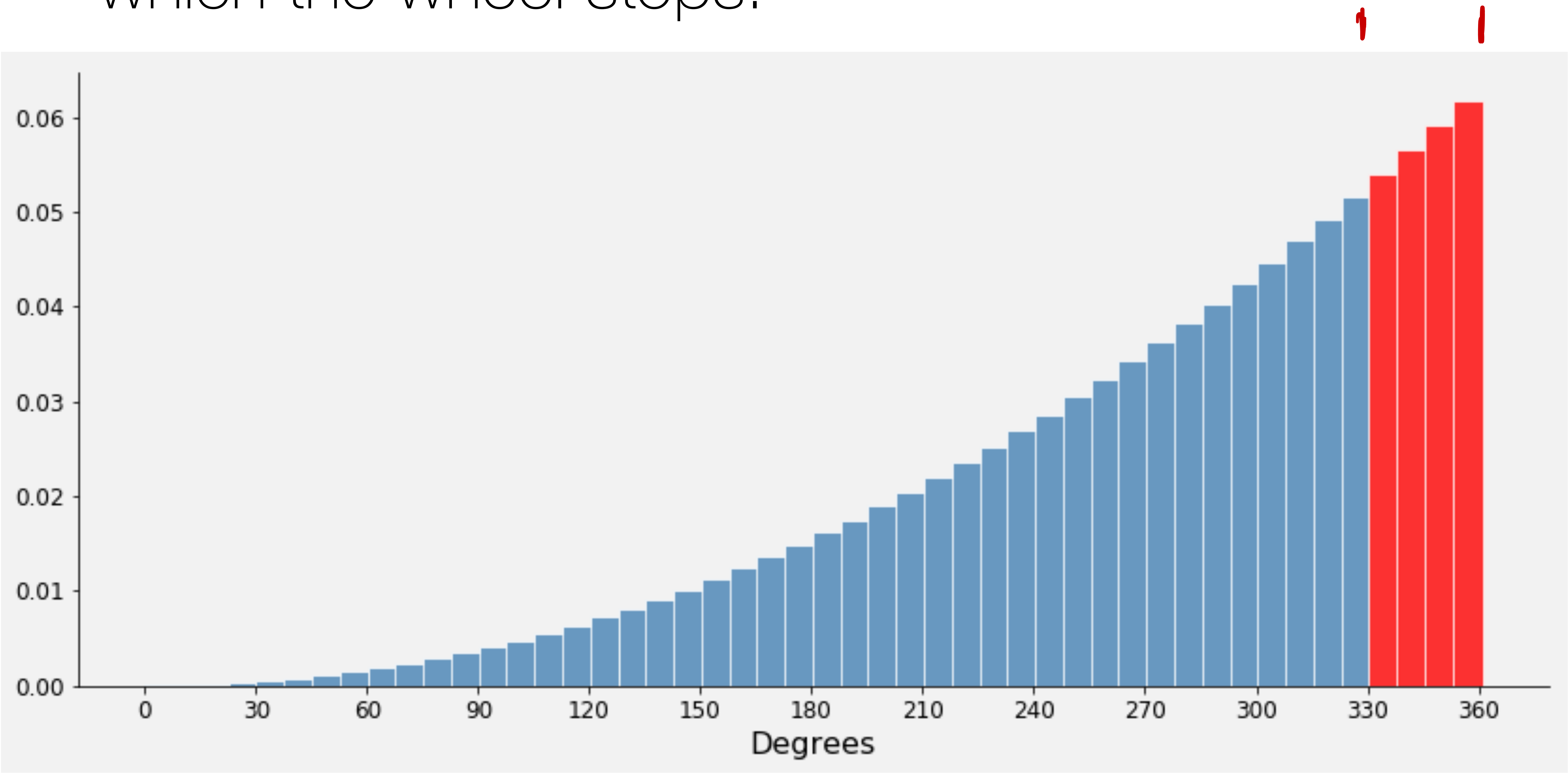
Intuition pump

- **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.



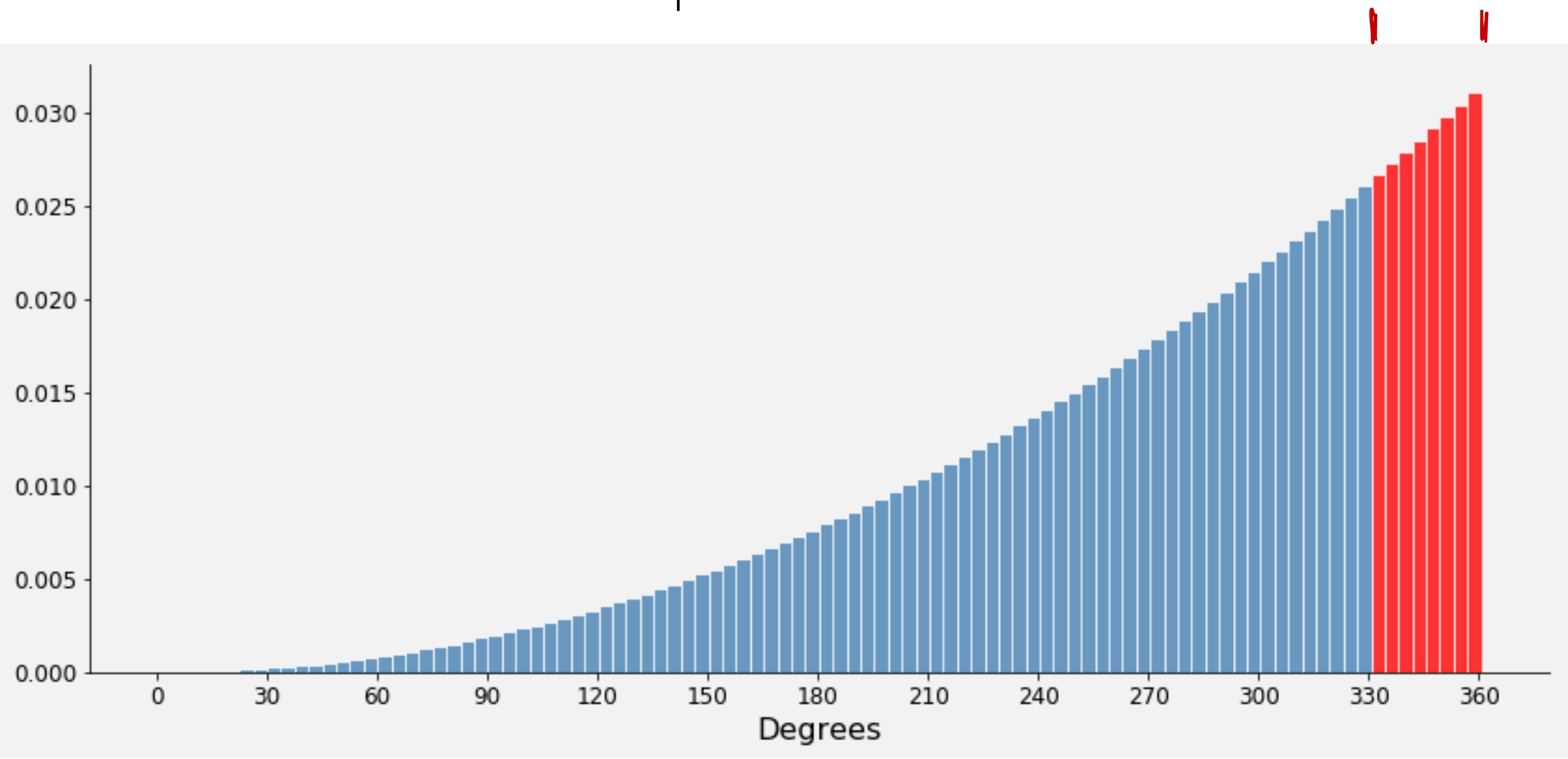
Intuition pump

- **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.



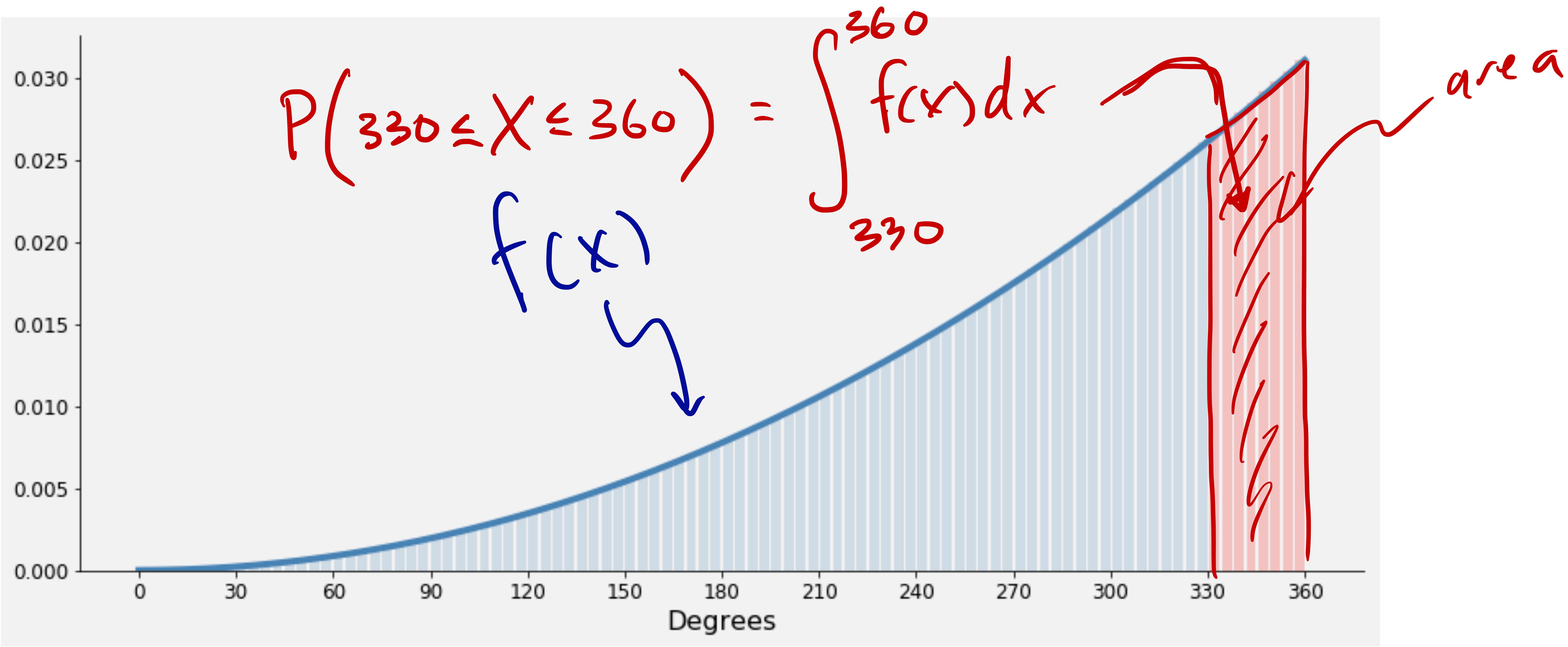
Intuition pump

- **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.

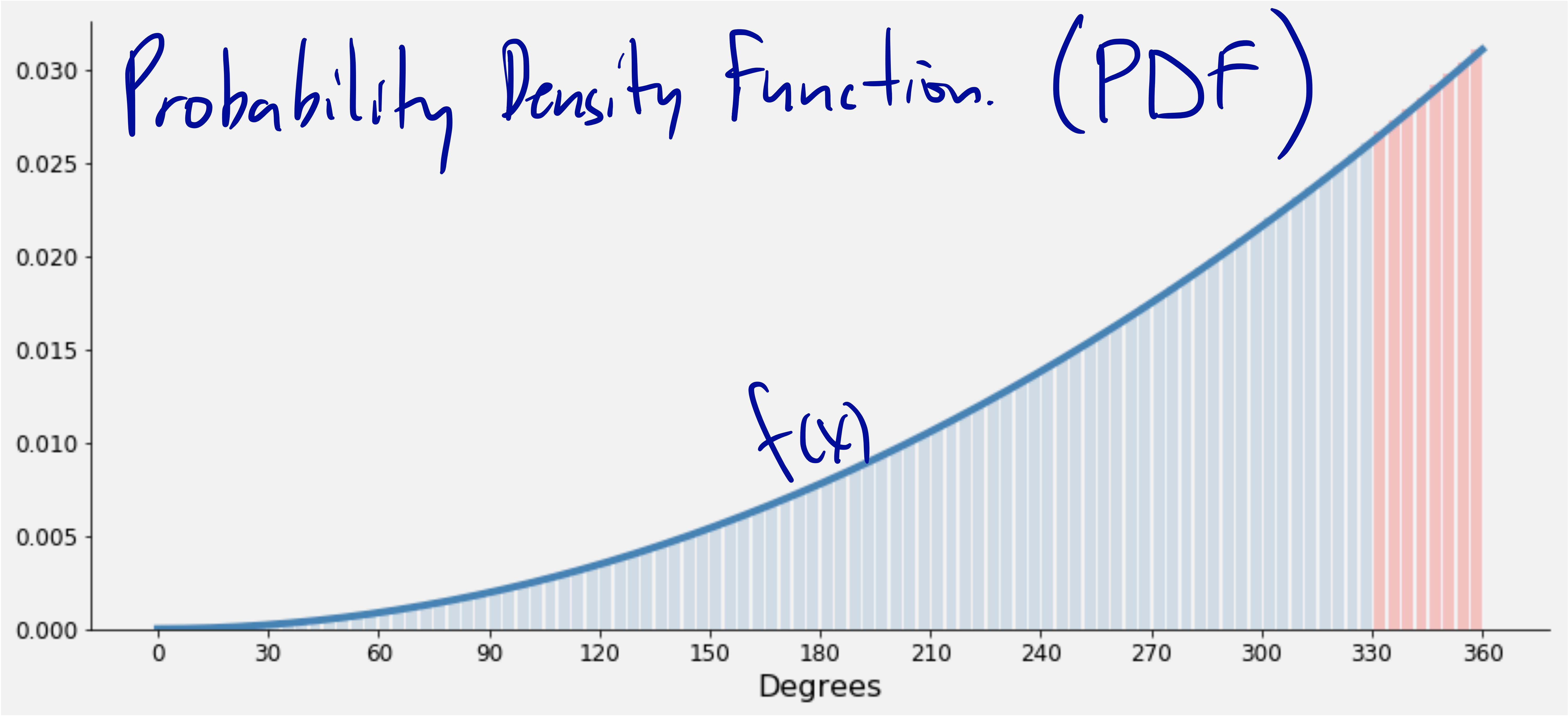


Intuition pump

- 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…



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

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

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

1. $f(x) \geq 0 \quad \forall x$ $\qquad \forall$ "for all" \forall

2. $P(\Omega) = 1$ i.e., $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Probability density function (PDF)

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

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

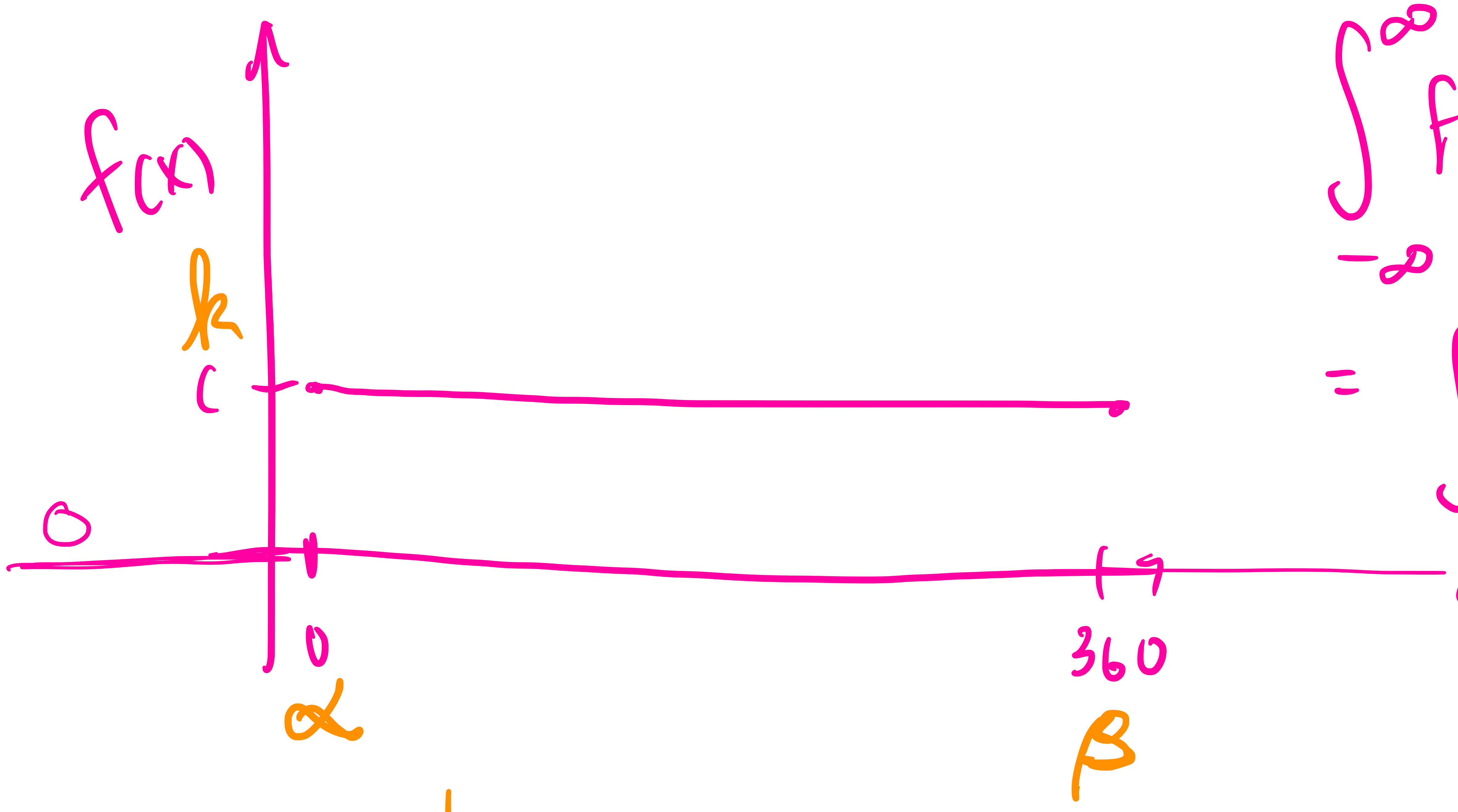
$$P(a-\varepsilon \leq X \leq a+\varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f(x)\,dx$$

Think about limit as $\varepsilon \to 0$

Continuous uniform distribution

PDF for unif$[\alpha,\beta]$: $f(x)=\frac{1}{\beta-\alpha}$

- **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?



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

$$= \int_0^{360} f(x)\,dx = 1$$

$$\int_0^{360} c\,dx = 1$$

$$c(360-0) = 1$$

$$\Rightarrow c = \frac{1}{360}$$

$$c = \frac{1}{\beta-\alpha}$$

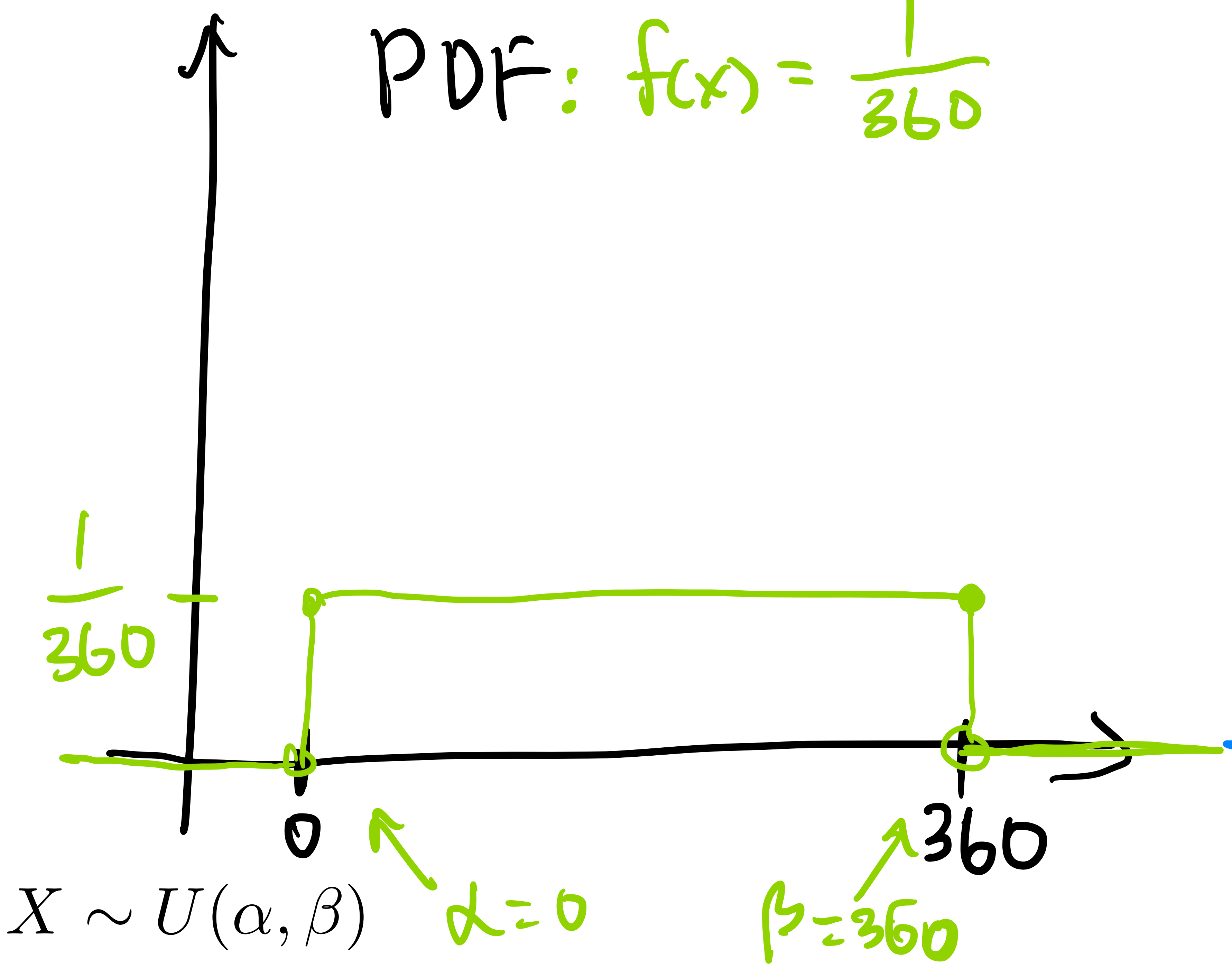
Continuous uniform distribution

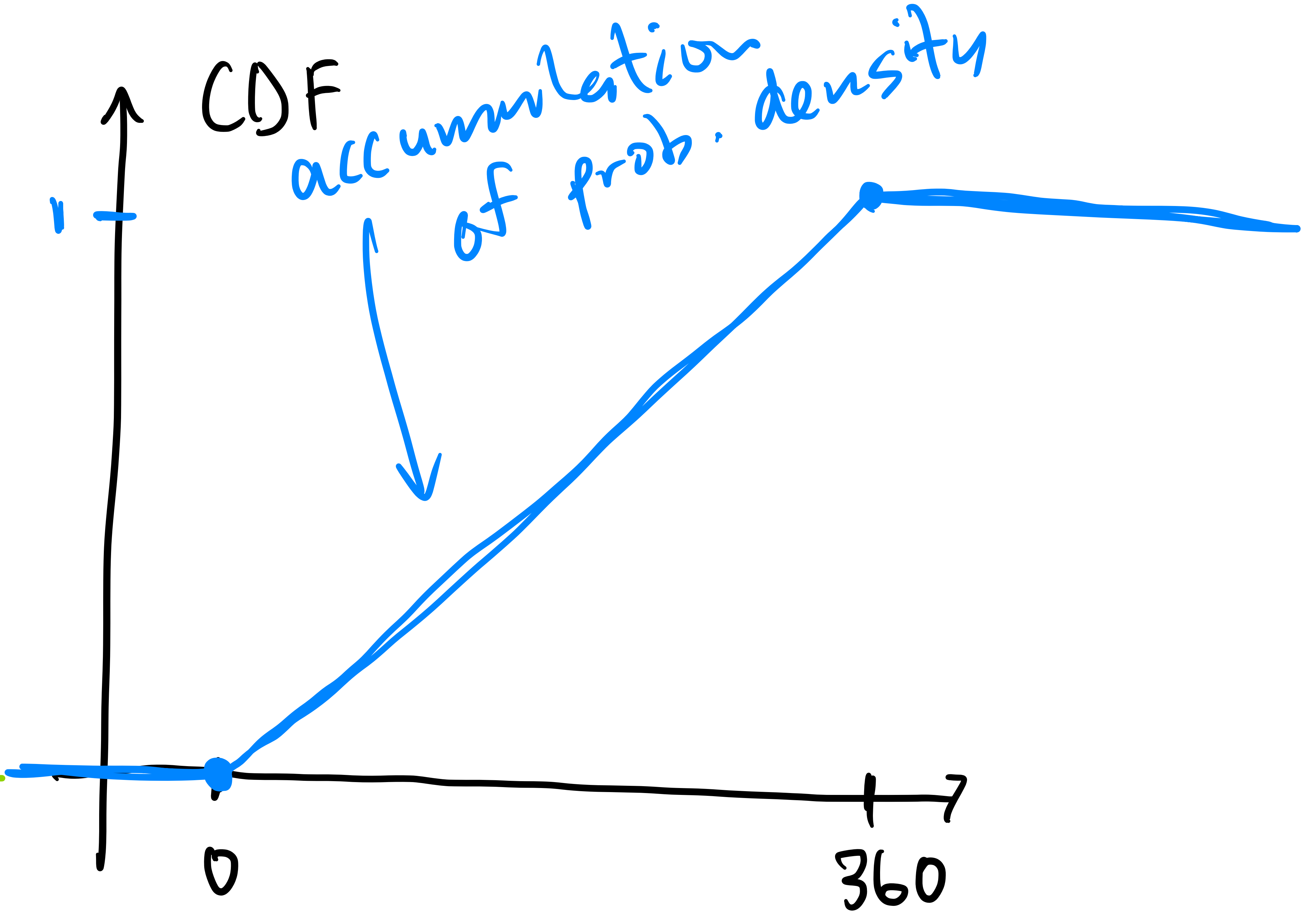
- **Def**: A continuous random variable has a uniform distribution on the interval [α, β] if its probability density function f(x) is

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

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

PDF: $f(x) = \frac{1}{360}$





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

Reiterate: density

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

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

- 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)

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

  PMF

- Recall: continuous RVs don't have a probability mass function.

  PDF

- And yet! Both have a CDF: $F_X(a) = P(X \leq a)$
- What is the CDF for a discrete RV?

$$F_X(a) = \sum_{x=-\infty}^{a} f(x)$$

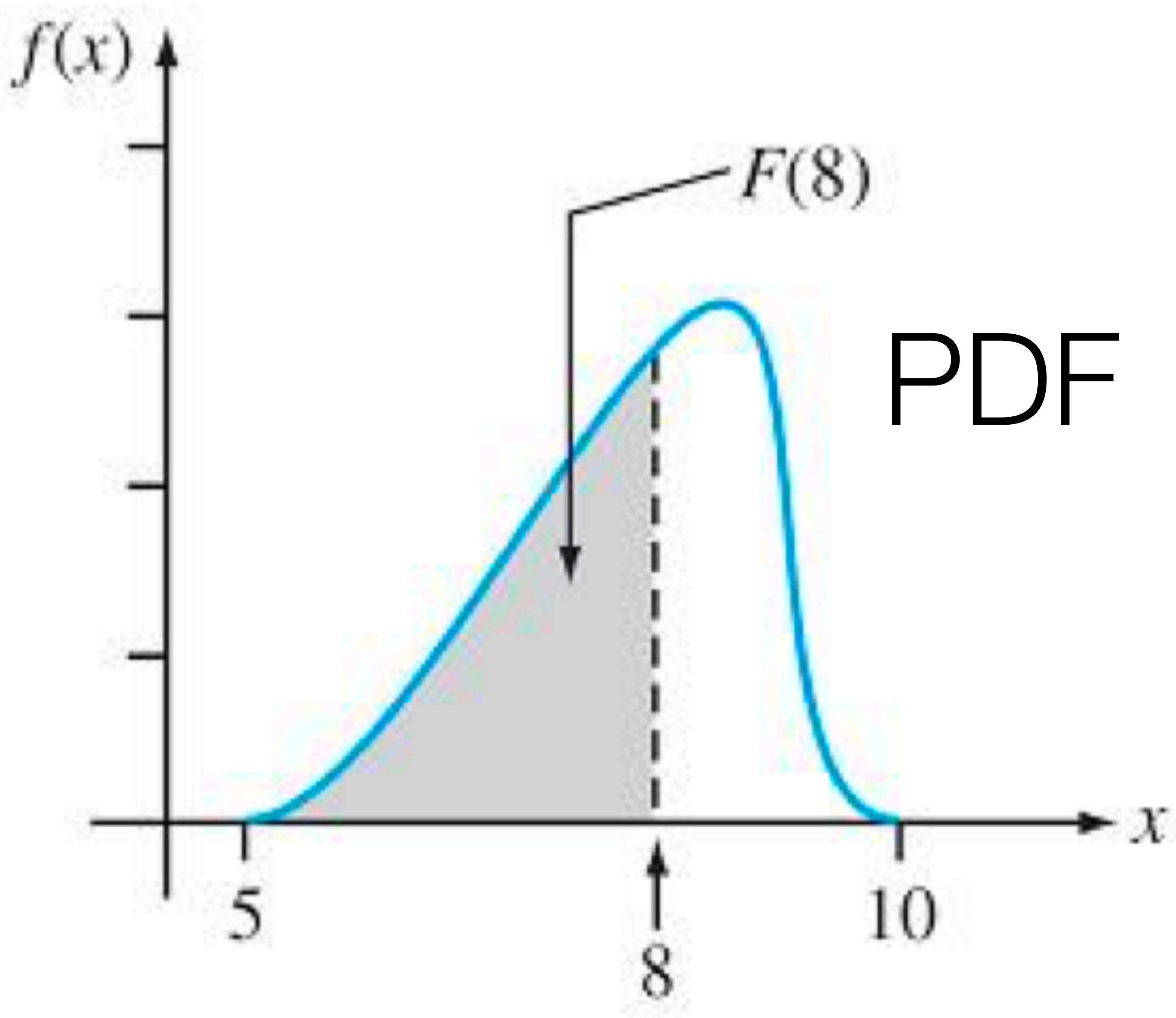
- What is the CDF for a continuous RV?

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

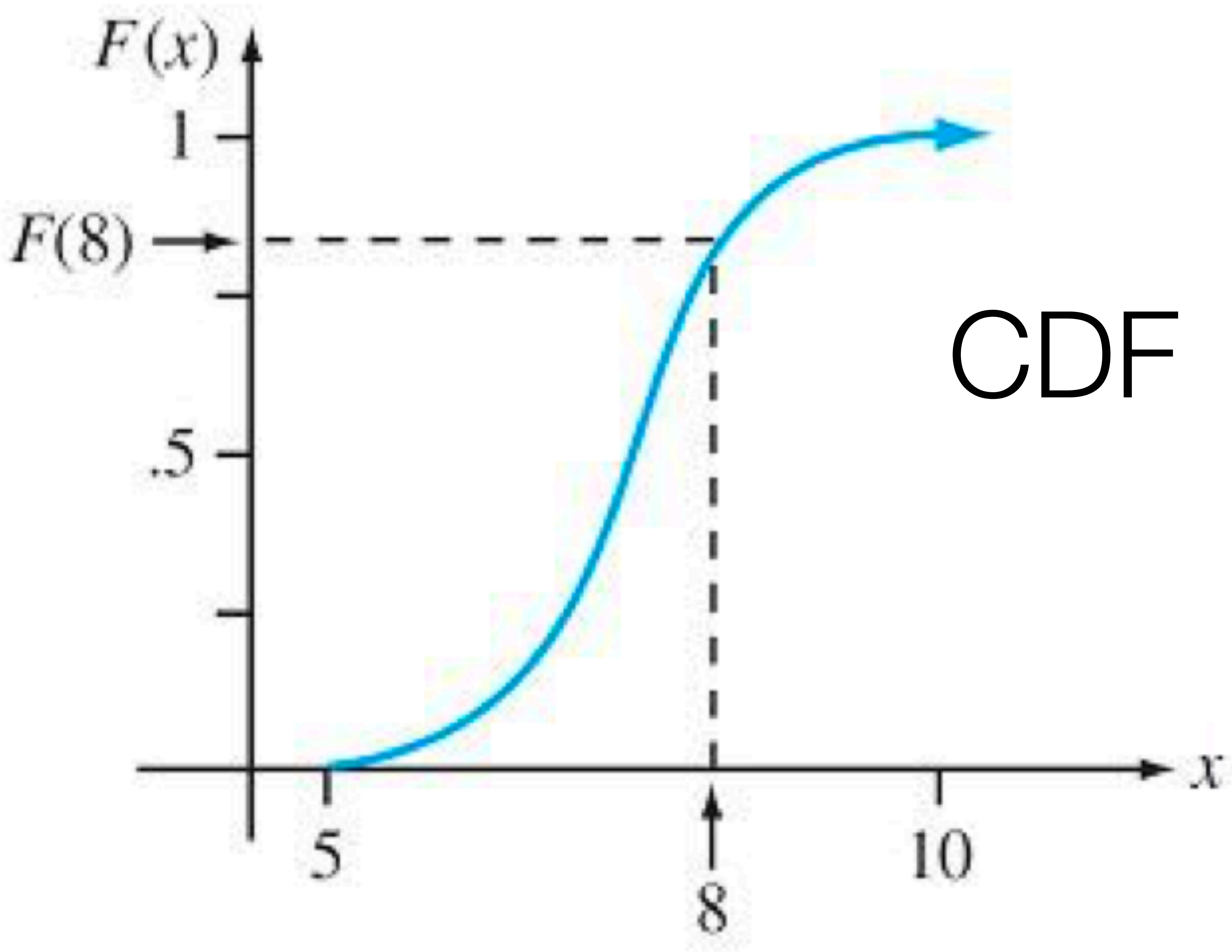
Cumulative distribution functions (CDFs)

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

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



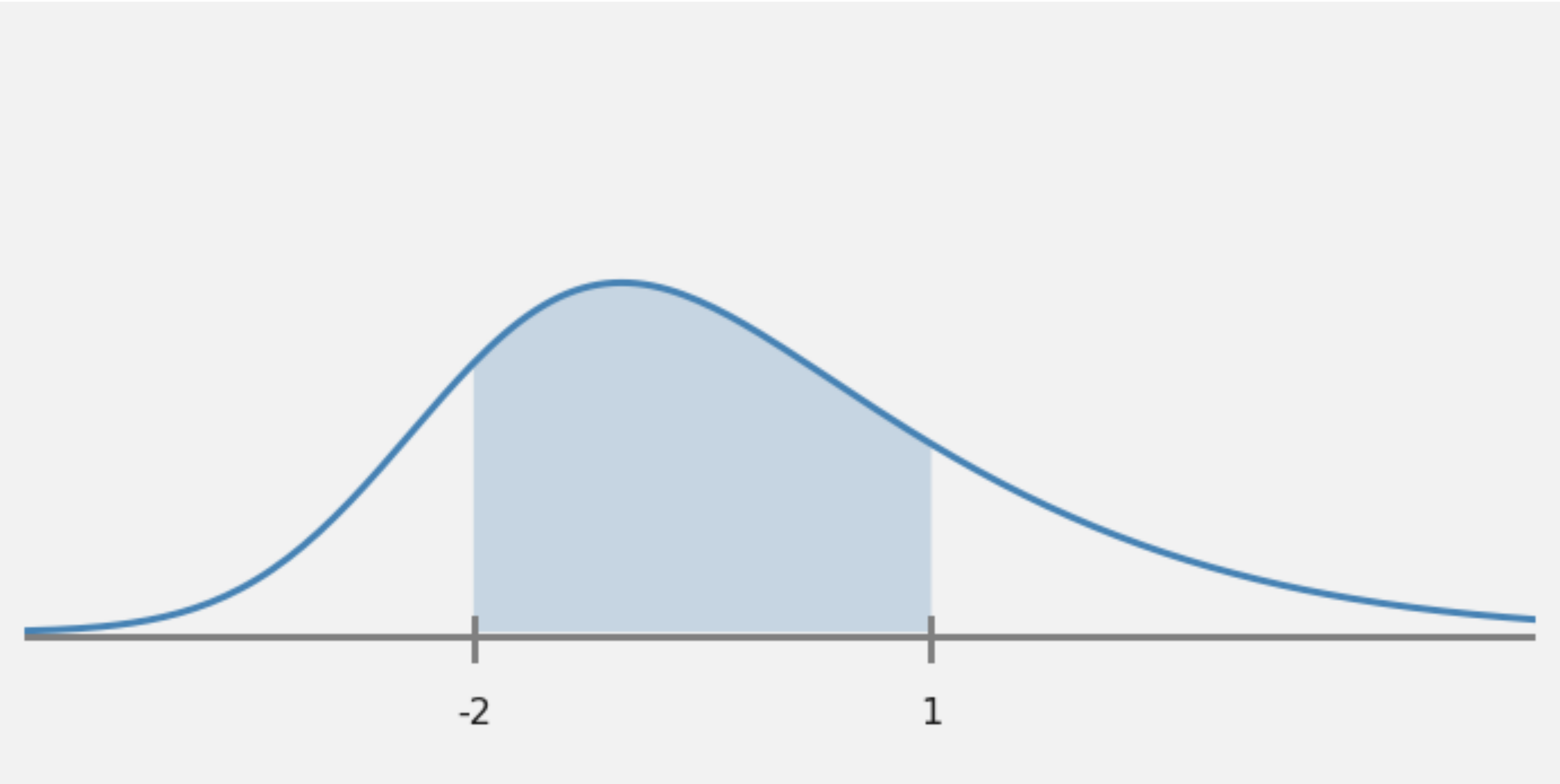
PDF



CDF

Cumulative distribution functions (CDFs)

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

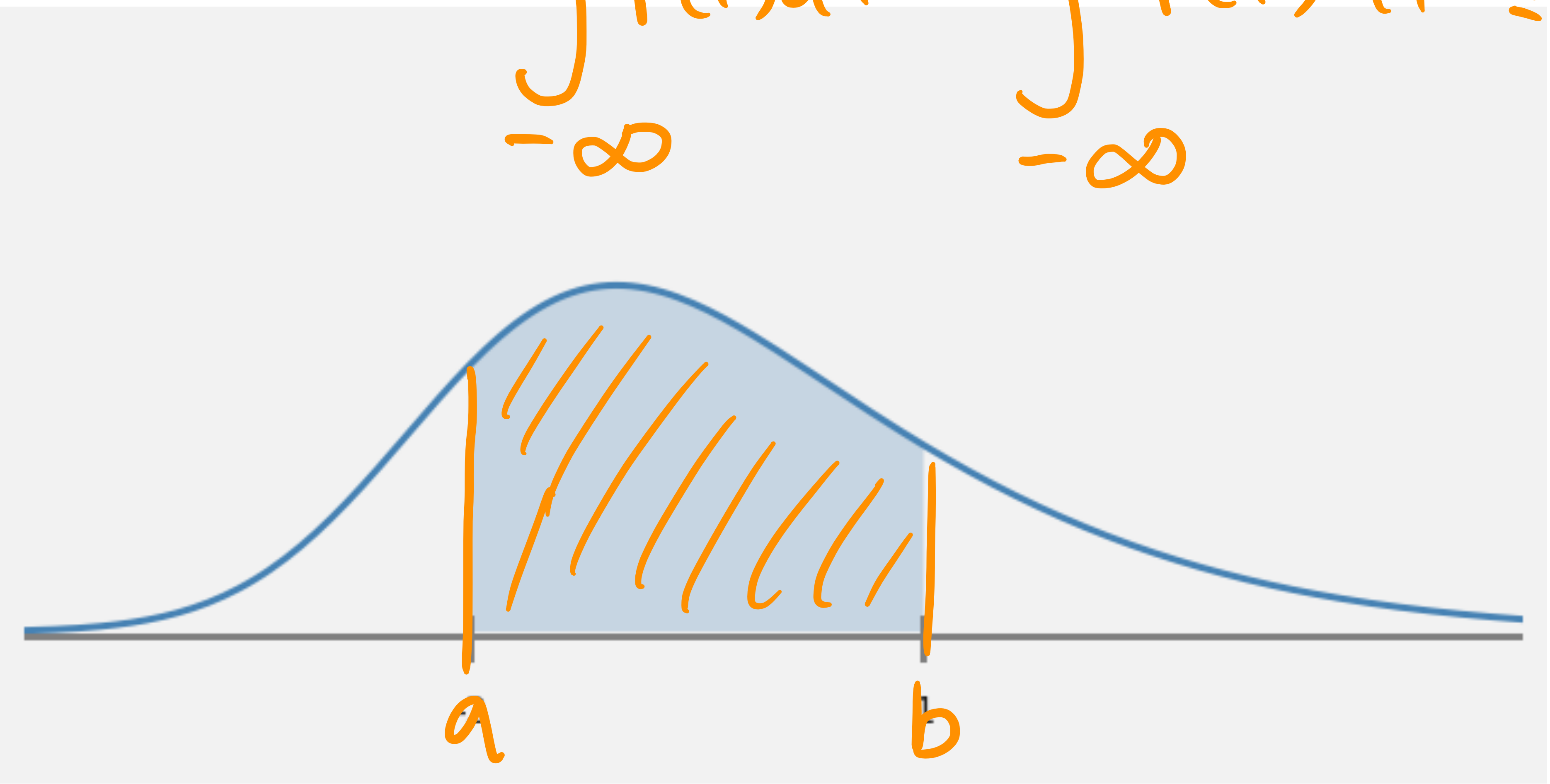


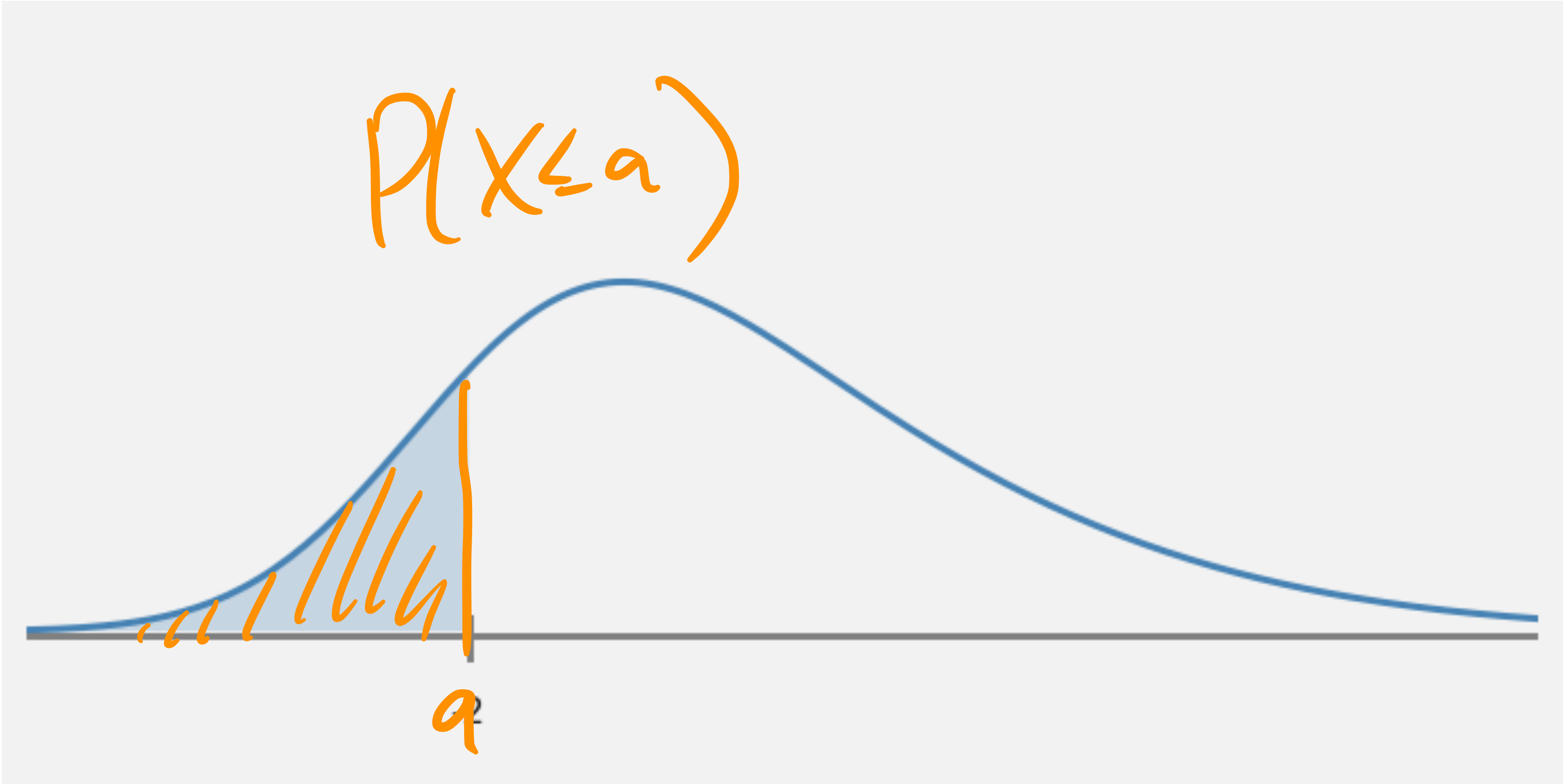
Cumulative distribution functions (CDFs)

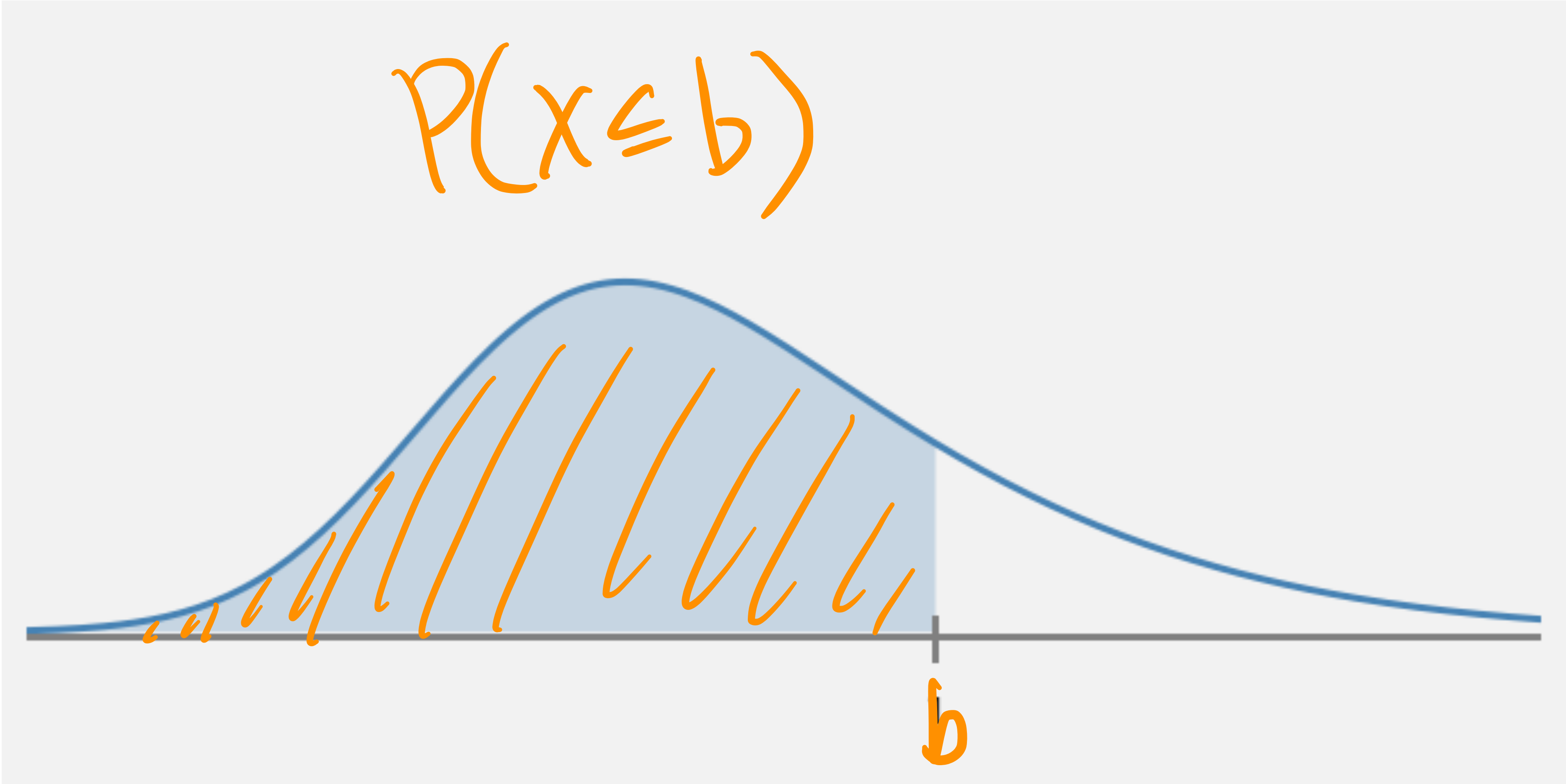
- What about things like $P(a \leq X \leq b)$? Like the probability of spinning red?

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

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







Hello, old friend! 😘

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

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

- Does this remind you of anything? F is the antiderivative of f

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

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

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

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

- 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??** $\in [0, \infty)$

How long until 1st text comes in?

Now it's t

Wait until $t + \Delta t$

Let X be wait time

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

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

$$= 1 - \text{Prob I get 0 texts in } \Delta t$$

From last time

Exponential Distribution!

- **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 zero texts in Δt window

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

$= 1 - \frac{1}{1} e^{-\lambda \Delta t}$ $\Rightarrow P(X \leq \Delta t) = F(\Delta t) = 1 - e^{-\lambda \Delta t}$

$F(t) = 1 - e^{-\lambda t}$

$\uparrow \frac{d}{dt}$

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

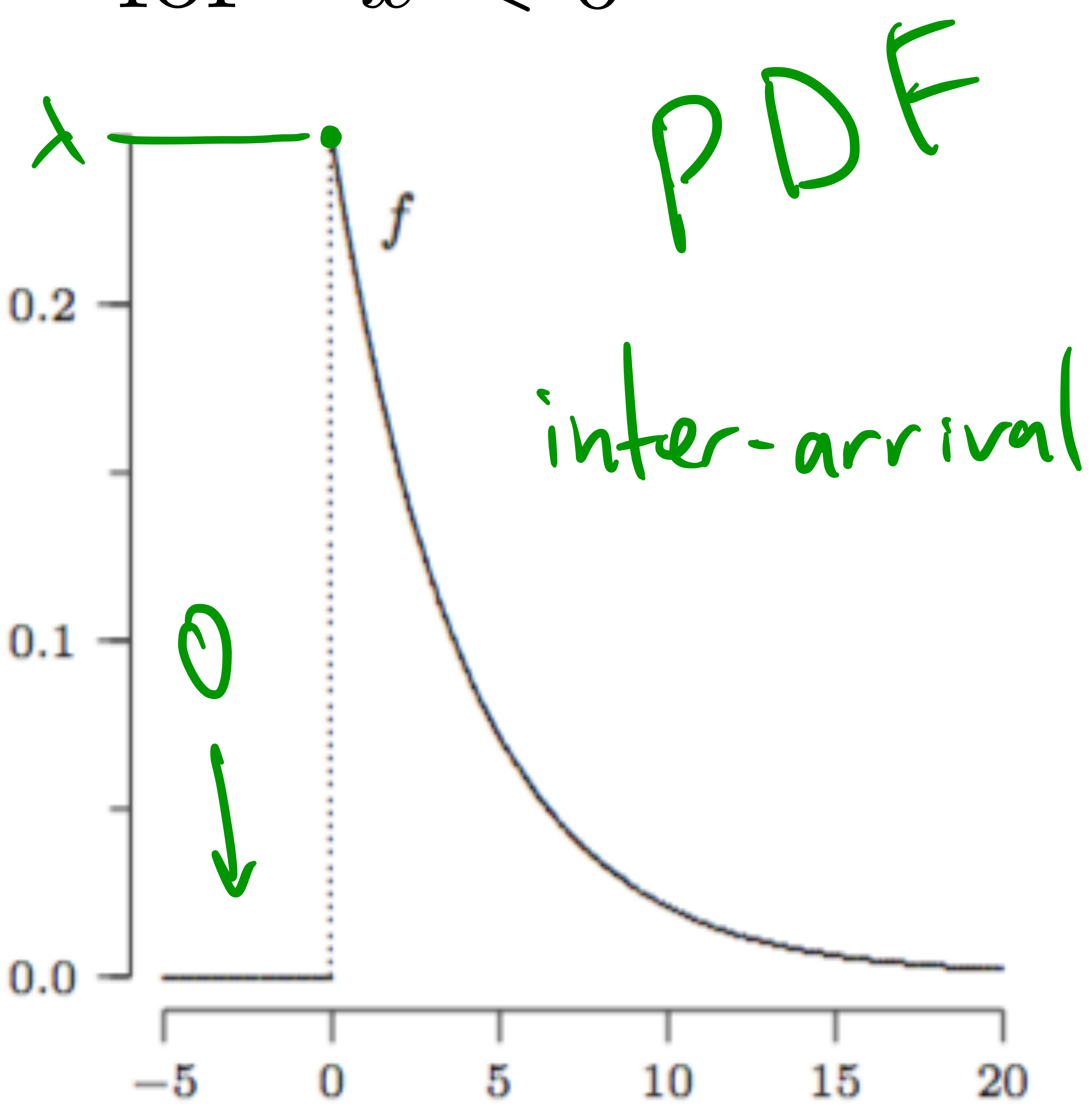
The exponential distribution

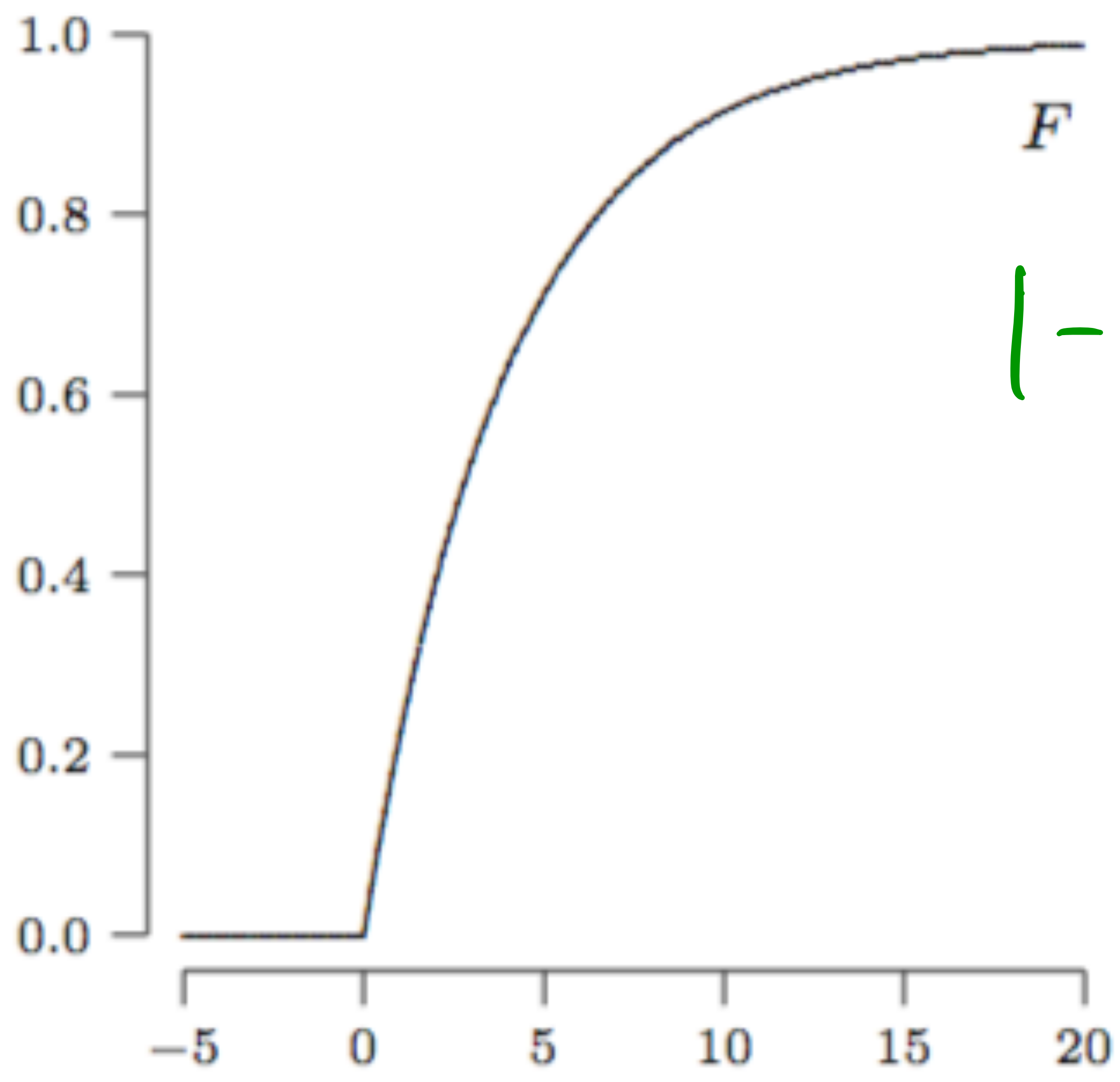
when $x=0$, $f(0)=\lambda$

- **Definition**: a continuous random variable has an *exponential distribution* with parameter λ if its probability density function f is given by

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad x \geq 0$$

and $f(x) = 0 \quad \text{for} \quad x < 0$





$CDF = P(X \leq a)$

wait time is atmost a,

$1 - CDF = P(X > a)$

$P(\text{wait time is} \geq a)$

PROB WARS VI: Return of the Quantiles

- In exploratory data analysis, Q1, median (Q2), and Q3 were values that divided a set of values evenly: the bottom 1/4, the middle, and the top 1/4.

- 🤔🤔🤔… use the CDF to write down the p^{th} quantile of a CRV X.

$$\int_{-\infty}^{Q_2} f(x)\,dx = \frac{1}{2} = \int_{Q_2}^{\infty} f(x)\,dx$$

$$\int_{-\infty}^{p^{th}\text{ quantile}} f(x)\,dx = p$$

$$F(Q_2) = \frac{1}{2}$$

$$F\left(p^{th}\text{ quantile}\right) = p$$