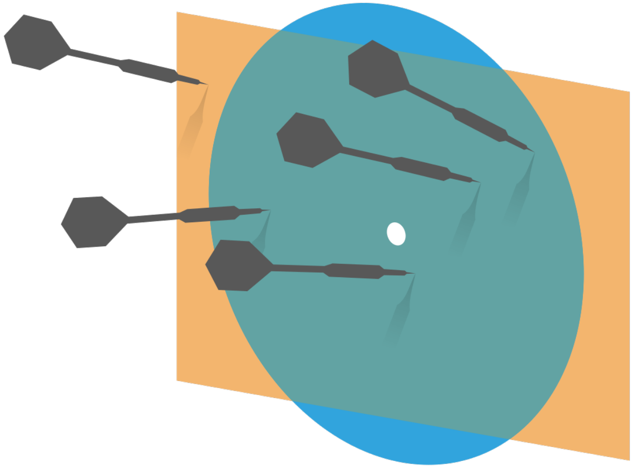
Probability and Statistics for Computer Science



Credit: wikipedia

Can we call e the exciting e ?

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

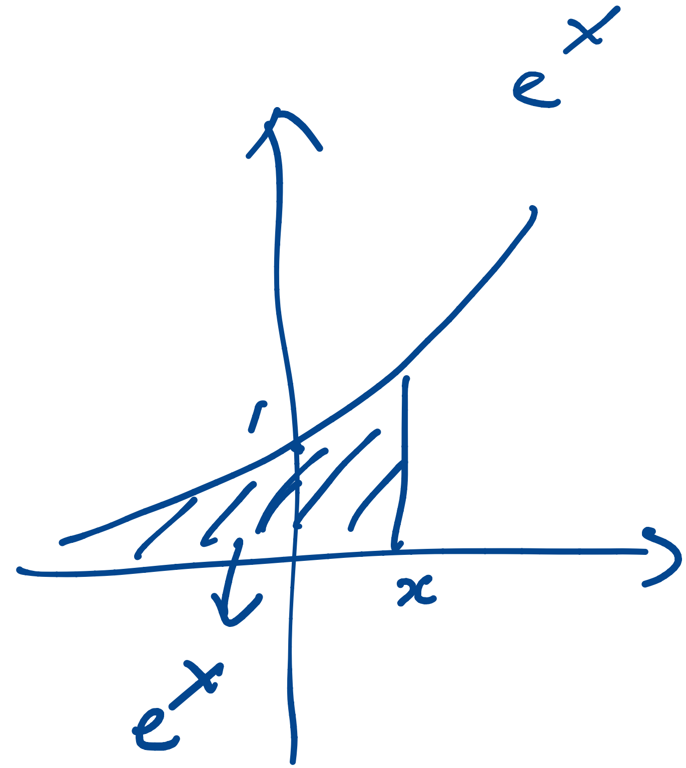
Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 9.29.2020

what is the number ?

$$e^x = \sum_{k=0}^{\infty} a_k x^k$$

$$a_k = ? \qquad \frac{1}{k!}$$

$$(e^x)' = e^x$$



How many empty slots?

Hashing N items to k slots, $(N \geq k)$ collisions are allowed, and will be handled by linked list. What is the expected number of empty slots?

$$X_i = \begin{cases} 1 & \text{slot } i \text{ remains empty} \\ 0 & \text{otherwise} \end{cases}$$

$$P(\text{slot } i \text{ remains empty}) = \left(1 - \frac{1}{k}\right)^N$$

$$E[X_i] = \sum x P(x) = 1 \cdot P(\text{slot } i \text{ empty}) + \cancel{0 \cdot (1-p)} = \left(1 - \frac{1}{k}\right)^N$$

$$E[X] = E\left[\sum X_i\right] = k \cdot \underbrace{\left(1 - \frac{1}{k}\right)^N}_{\frac{1}{e}}$$

$N = k$

$N \uparrow$

Last time

Bernoulli Distribution

Binomial Distribution

Geometric Distribution

Bernoulli trials

Objectives

Poisson Distribution

Continuous Random Variable

Probability Density Function

Exponential Distribution

Motivation for Poisson Distr.

COVID incidences in a time interval.

→ all these rate data in a counting process.

Motivation for a model called Poisson Distribution

- What's the probability of the **number of incoming customers (k)** in an hour?
- It's widely applicable in physics and engineering both for modeling of time and space.

Simeon D. Poisson Credit: wikipedia
(1781-1840)

Degroot
Pg 287
-288

Poisson Distribution

- A discrete random variable X is called **Poisson** with intensity **λ** (λ>0) if

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

for integer $k \geq 0$

Simeon D. Poisson
(1781-1840)

λ *is the* **average rate** *of the event's occurrence*

Poisson Distribution

- **Poisson** distribution is a valid pdf for

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{\lambda}$$

$$\sum_k p(X=k) = \sum_k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_k \frac{\lambda^k}{k!}$$

$$p(\Omega)=1$$

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\cdot e^{\lambda} = e^0 = 1$$

for integer $k \geq 0$


Simeon D. Poisson
(1781-1840)

λ *is the* **average rate** *of the event's occurrence*

Expectations of Poisson Distribution

- The expected value and the variance are wonderfully the same! That is **λ**

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$ for integer $k \geq 0$

$E[X]$

$= \sum x\, p(x)$

$= \sum \cancel{k}\, \frac{e^{-\lambda}\lambda^k}{(k-1)!}$

$= \sum \frac{e^{-\lambda}\lambda \cdot \lambda^{k-1}}{(k-1)!}$

$= \lambda \sum \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}$

$= \lambda$

$var[X]$

$$\boxed{\begin{aligned} E[X] &= \lambda \\ var[X] &= \lambda \end{aligned}}$$

Simeon D. Poisson
(1781-1840)

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

$$\operatorname{var}[X] = E[X^2] - \underline{E^2[X]}$$

$$= \underline{\sum \frac{k^2 e^{-\lambda} \lambda^k}{k!}} - \lambda^2$$

$$= \lambda$$

Examples of Poisson Distribution

- How many calls does a call center get in an hour?
- How many mutations occur per 100k nucleotides in an DNA strand?
- How many **independent** incidents occur in an interval?

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

for integer $k \geq 0$

Poisson Distribution: call center

- If a call center receives 10 calls per hour on average, what is the probability that it receives 15 calls in a given hour?

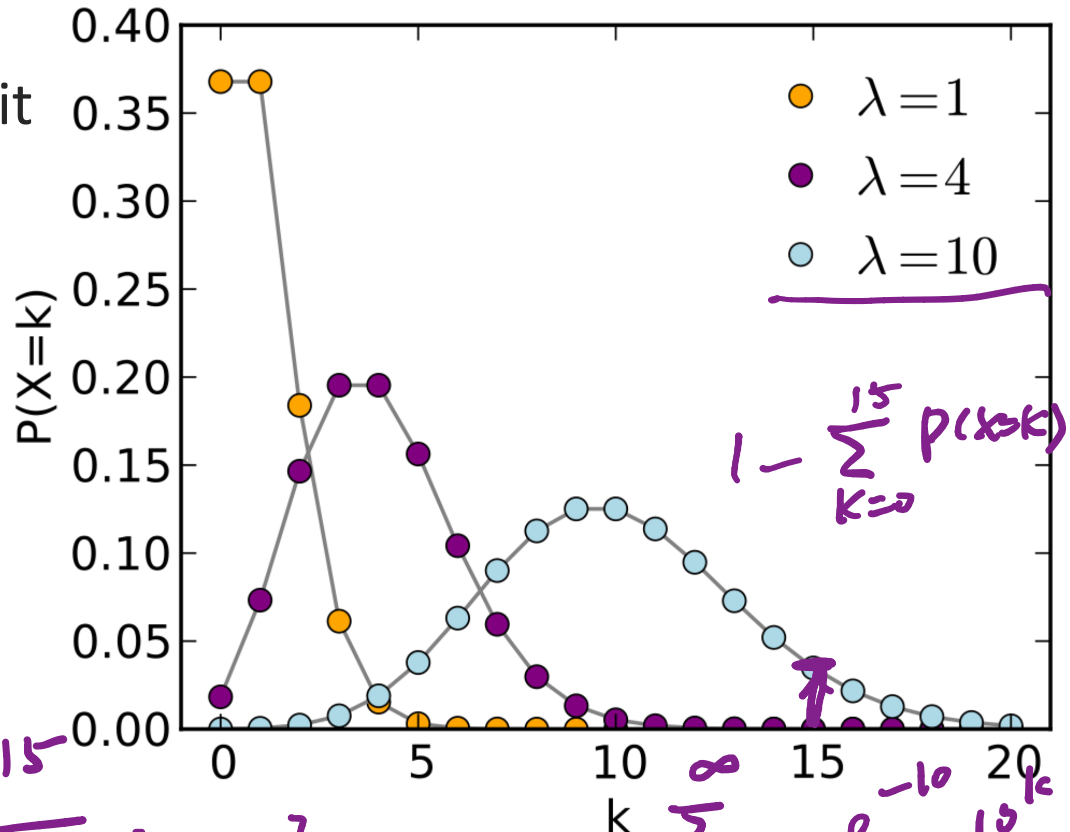
- What is λ here? $= 10$

- What is P(k=15)?

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

$k=15$

$$= \frac{e^{-10}\cdot 10^{15}}{15!} \doteq 0.03$$



$$1 - \sum_{k=0}^{15} P(X=k)$$

$$\sum_{k=16}^{\infty} \frac{e^{-10}\,10^{k}}{k!}$$

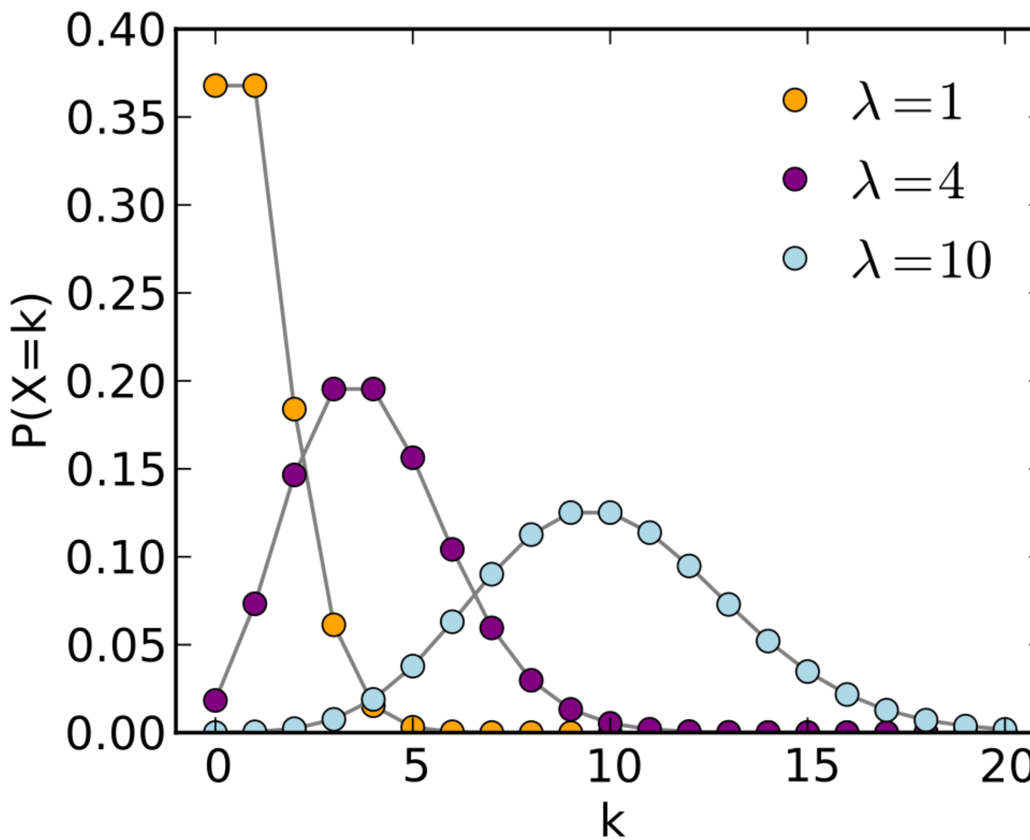
Credit: wikipedia

Q. Poisson Distribution: call center

If a call center receives 4 calls per hour on average.

What is intensity **λ** here for an hour?

A. 1
B. 4
C. 8



Credit: wikipedia

Q. Poisson Distribution: call center

If a call center receives 4 calls per hour on average.

What is probability the center receives 0 calls in an hour?

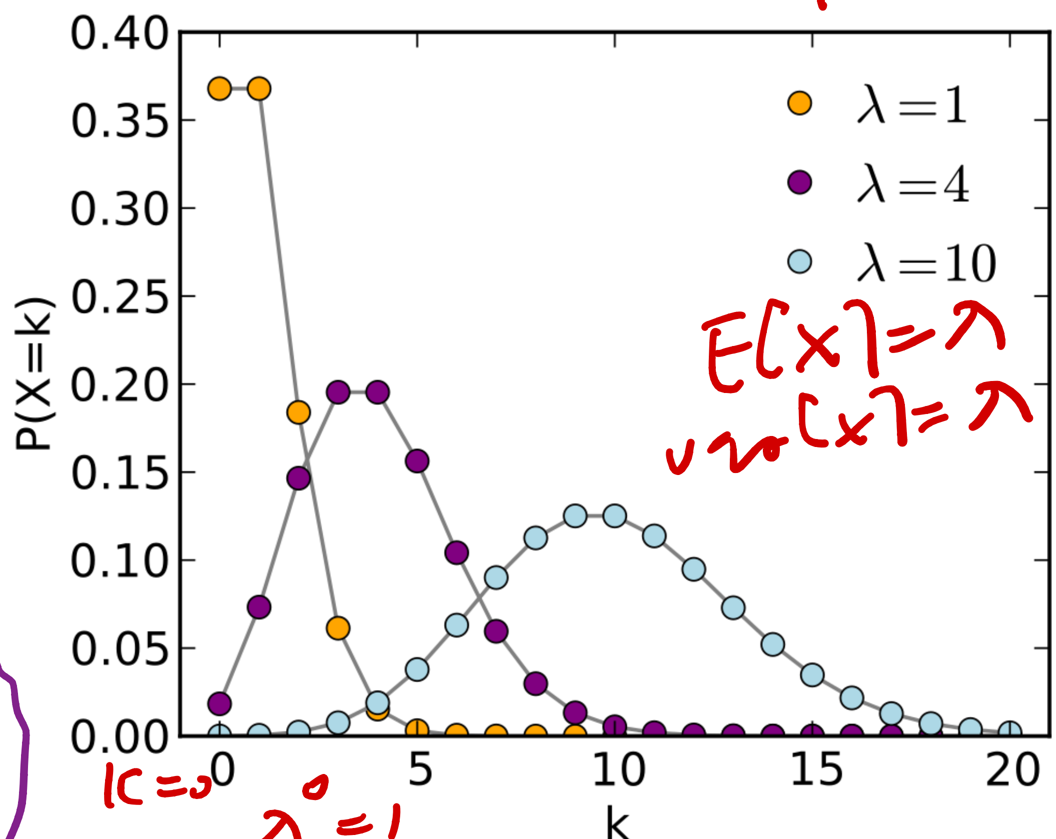
A. $\boxed{e^{-4}}$

B. 0.5

C. 0.05

$\lambda = 4$

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



λ

$E[X] = \lambda$

$\mathrm{var}[X] = \lambda$

$k=0$

$\lambda^0 = 1$

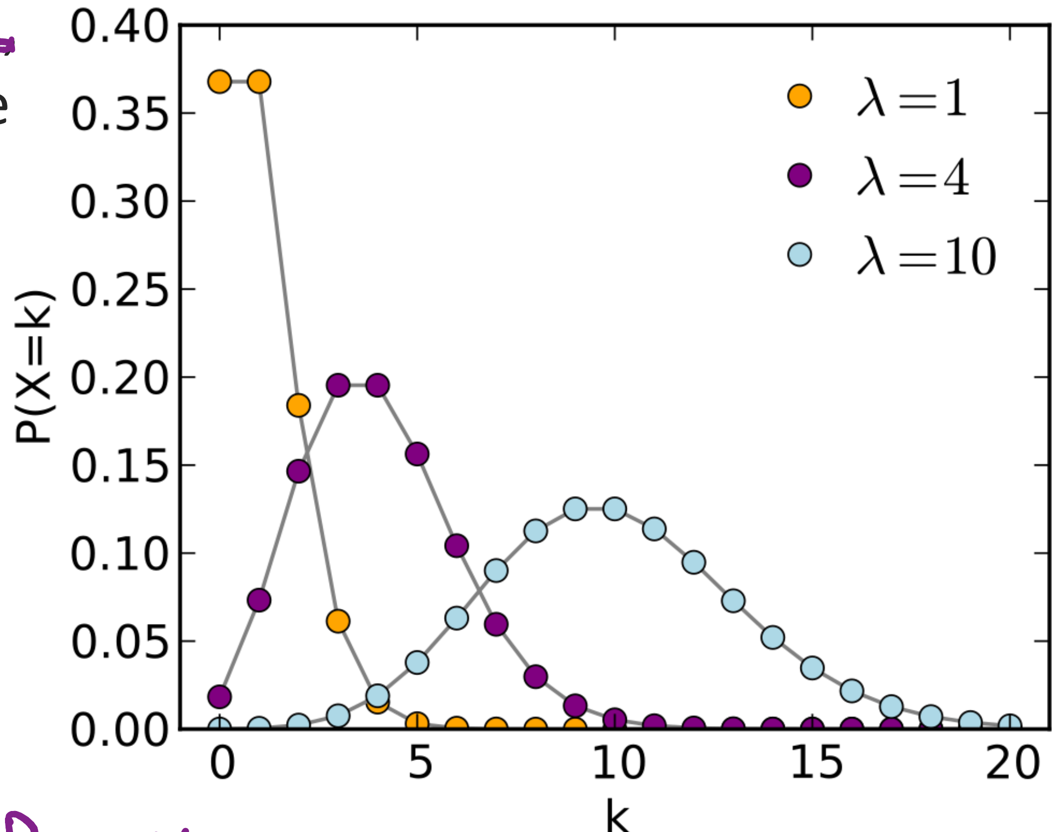
$k=0 \quad k! = 0! = 1$

Credit: wikipedia

Q. Poisson Distribution: call center

- Given a call center receives 10 calls per hour on average, what is the intensity **λ** of the distribution for calls in **Two** hours?

$\lambda = 20$

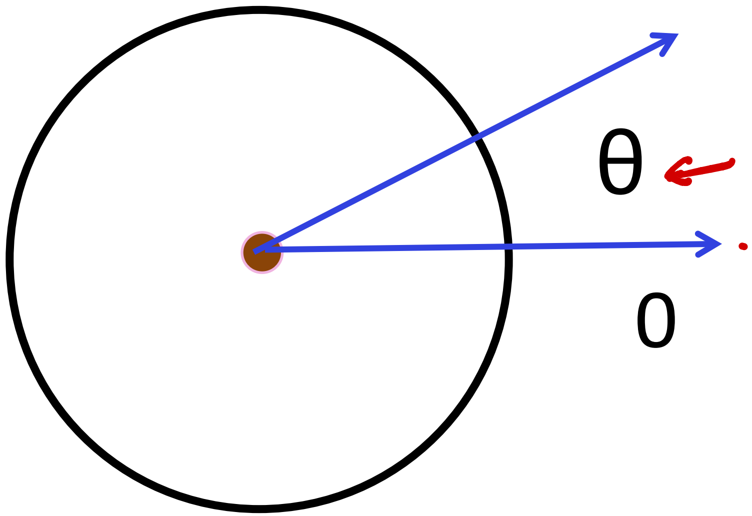


Proof: Degroot
pg 290

Credit: wikipedia

Example of a continuous random variable

- The spinner

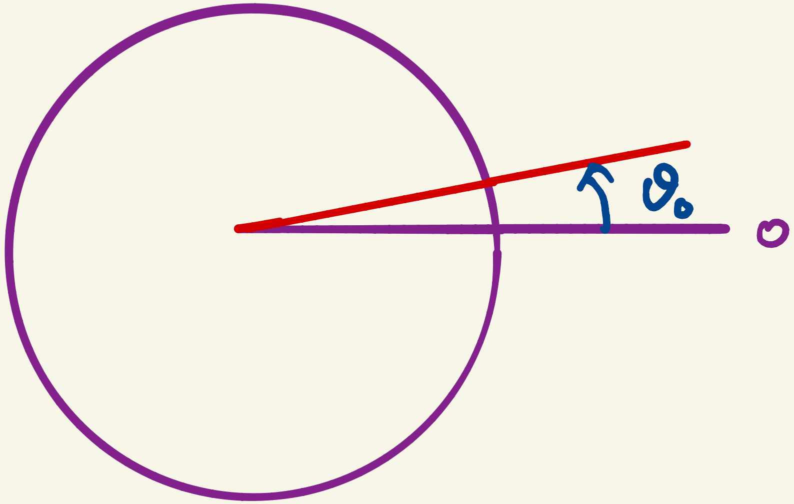


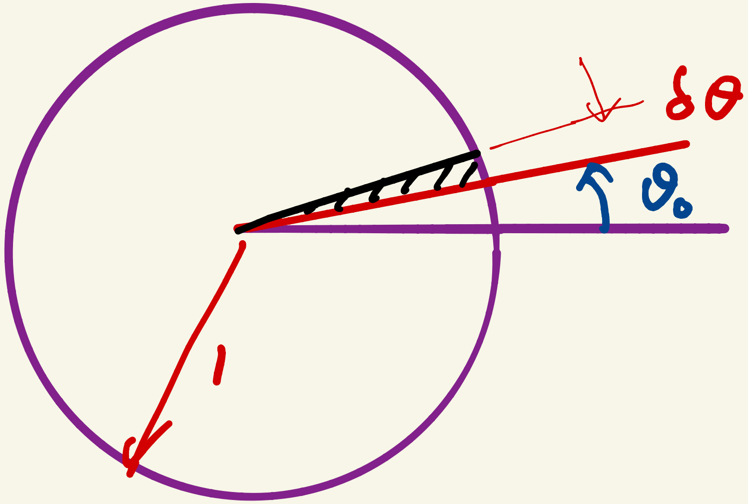
$$\theta \in (0, 2\pi]$$

- The sample space for all outcomes is not countable

* What is the probability of

$P(\theta = \theta_0)$? θ_0 is a constant in $(0, 2\pi]$





* What is the probability of

$P(\theta_0 < \theta < \theta_0 + \delta\theta)$?

$r=1$

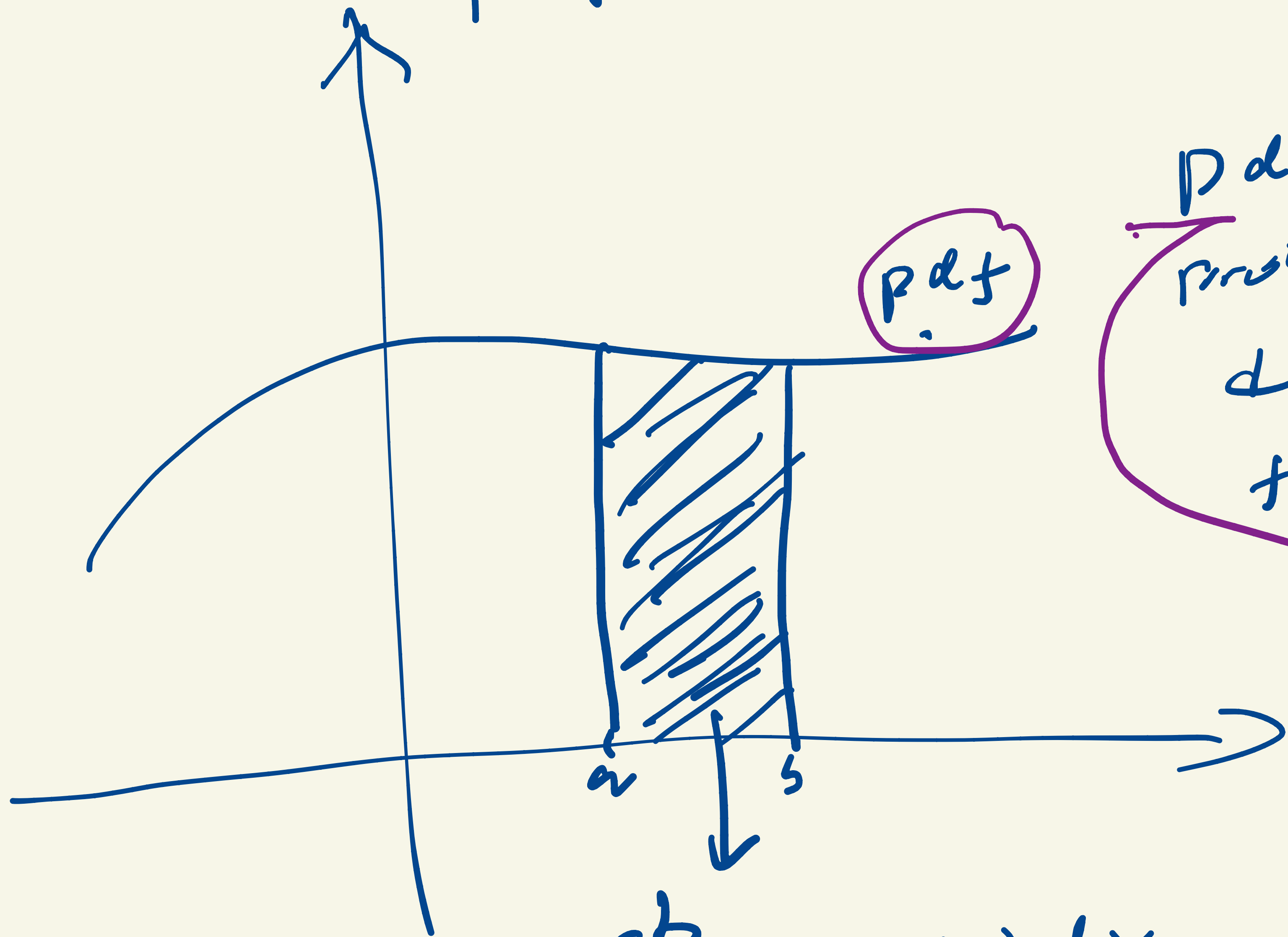
$$\frac{\frac{1}{2} \cdot r^2 \cdot \delta\theta}{\pi r^2} = \frac{\delta\theta}{2\pi}$$

$$\lim_{\delta\theta \to 0} \frac{\frac{\delta\theta}{2\pi}}{\delta\theta} = \frac{1}{2\pi}$$

Probability density function (pdf)

- For a continuous random variable X, the probability that X=x is essentially zero for all (or most) x, so we can't define $P(X = x)$
- Instead, we define the **probability density function** (pdf) over an infinitesimally small interval dx, $p(x)dx = P(X \in [x, x + dx])$
- For $a < b$ $$\int_a^b p(x)dx = P(X \in [a, b])$$

$pdf(X = x)$



$X = x$

pdf
probability
density
func.

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

$$\int_a^b pdf(x)\,dx$$

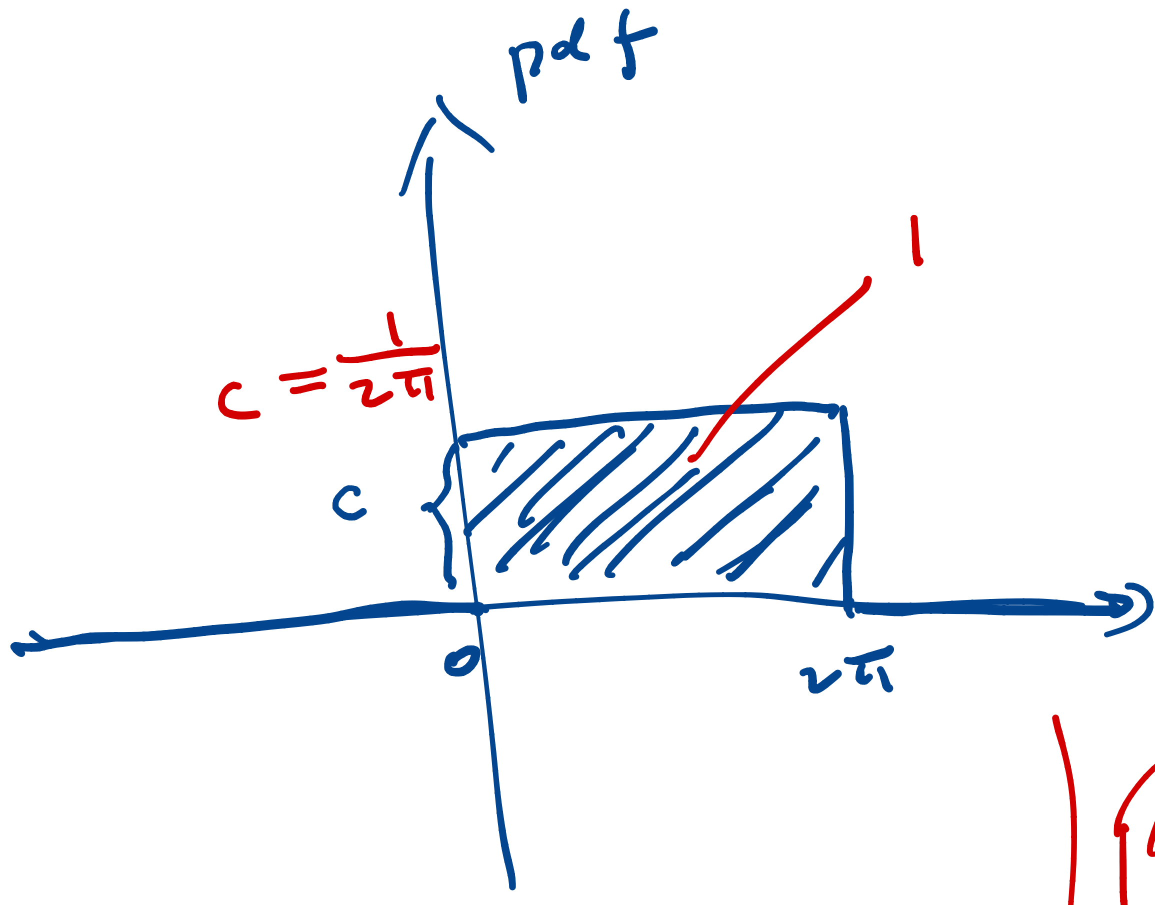
Properties of the probability density function

- $p(x)$ **resembles** the probability function of discrete random variables in that
  - $p(x) \geq 0$ for all x
  - The probability of X taking all possible values is 1.

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

$P(\Omega) = 1$

Area under the pdf curve



θ

$[0, 2\pi)$

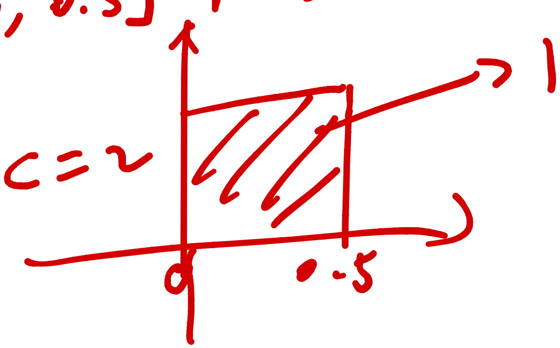
Properties of the probability density function

- $p(x)$ **differs** from the probability distribution function for a discrete random variable in that
  - $p(x)$ is not the probability that $X = x$
  - $p(x)$ can exceed 1

$$p(x) = \begin{cases} 2 & x \in [0, 0.5] \\ 0 & \text{o.w.} \end{cases}$$

pdf

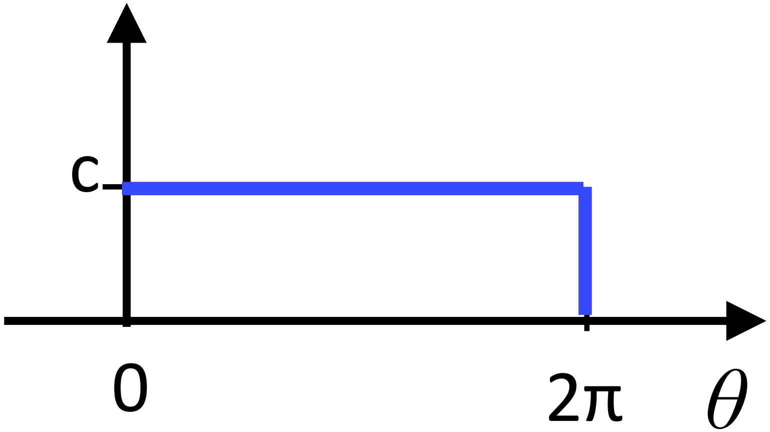
$c = 2$



Probability density function: spinner

- Suppose the spinner has equal chance stopping at any position. What's the pdf of the angle θ of the spin position?

$$p(\theta) = \begin{cases} c & if\ \theta \in (0, 2\pi] \\ 0 & otherwise \end{cases}$$



- For this function to be a pdf,

Then $$\underline{\int_{-\infty}^{\infty} p(\theta) d\theta = 1}$$

$c = \frac{1}{2\pi}$

Probability density function: spinner

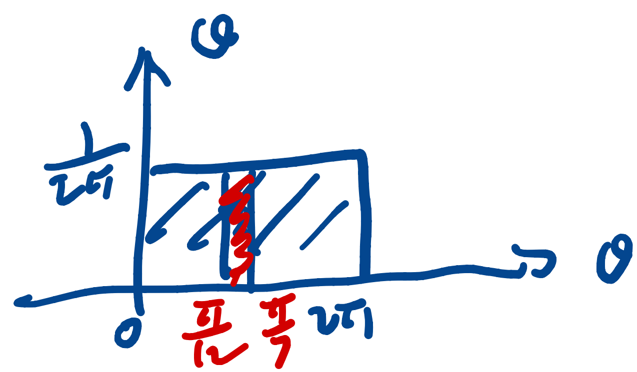
- What the probability that the spin angle θ is within $[\frac{\pi}{12}, \frac{\pi}{7}]$?

$$p\left(\theta \in \left[\frac{\pi}{12}, \frac{\pi}{7}\right]\right)$$

$$= \int_{\frac{\pi}{12}}^{\frac{\pi}{7}} p(\theta)\, d\theta$$

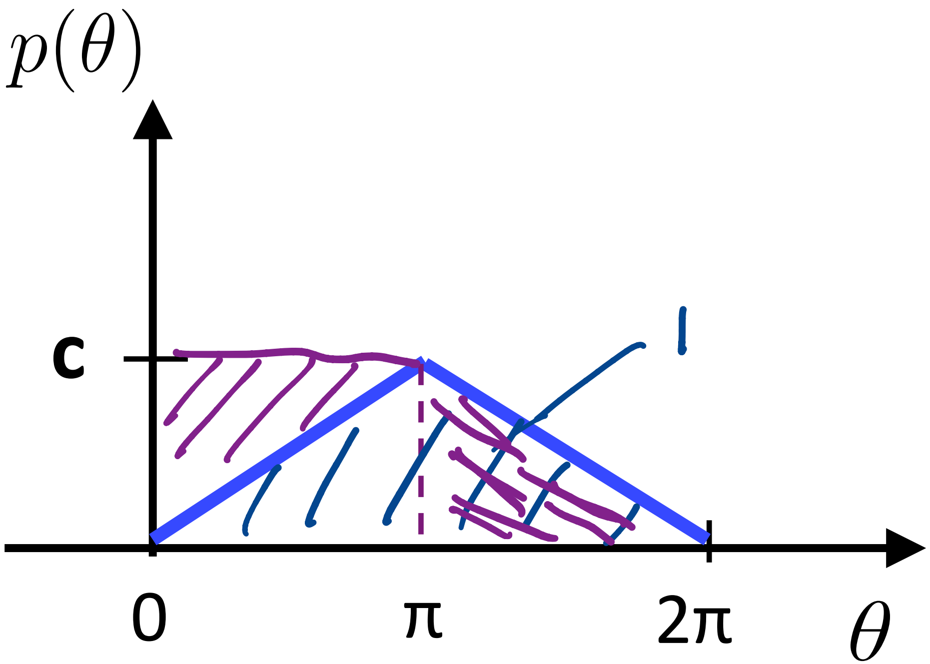
$$=$$

$$p(\theta) = \frac{1}{2\pi} \text{ for } \theta \in [0, 2\pi)$$



Q: Probability density function: spinner

- What is the constant **c** given the spin angle θ has the following pdf?



A. 1
B. 1/π
C. 2/π
D. 4/π
E. 1/2π

Expectation of continuous variables

- Expected value of a continuous random variable X

$\int$

S

$$E[X] = \int_{-\infty}^{\infty} x\boxed{p(x)dx}$$

weight

- Expected value of function of continuous random variable $Y = f(X)$

$$E[Y] = E[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)dx$$

Probability density function: spinner

- Given the probability density of the spin angle θ

$$p(\theta) = \begin{cases} \frac{1}{2\pi} & if\ \theta \in (0, 2\pi] \\ 0 & otherwise \end{cases}$$

- The expected value of spin angle is

$$E[\theta] = \int_{-\infty}^{\infty} \theta p(\theta) d\theta = \int_0^{2\pi} \theta \frac{1}{2\pi} d\theta$$

$$= \frac{1}{2\pi} \frac{\theta^2}{2} \Big|_0^{2\pi} = \frac{4\pi^2}{4\pi}$$

$$= \pi$$

Properties of expectation of continuous random variables

- The linearity of expected value is true for continuous random variables.

$$\sum \longrightarrow \int$$

- And the other properties that we derived for variance and covariance also hold for continuous random variable do at home

Q.

- Suppose a continuous variable has pdf

$$p(x) = \begin{cases} 2(1-x) & x \in [0,1] \\ 0 & otherwise \end{cases}$$

What is E[X]?

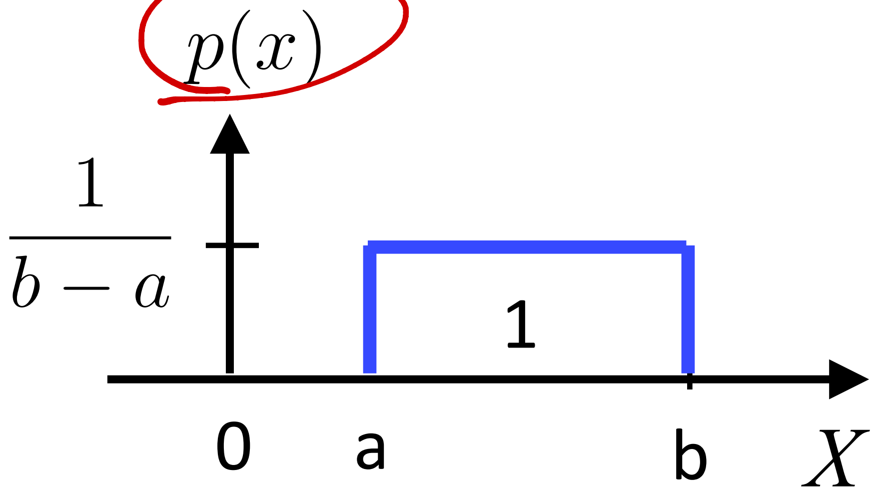
A. 1/2 B. 1/3 C. 1/4

D. 1 E. 2/3

$$E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

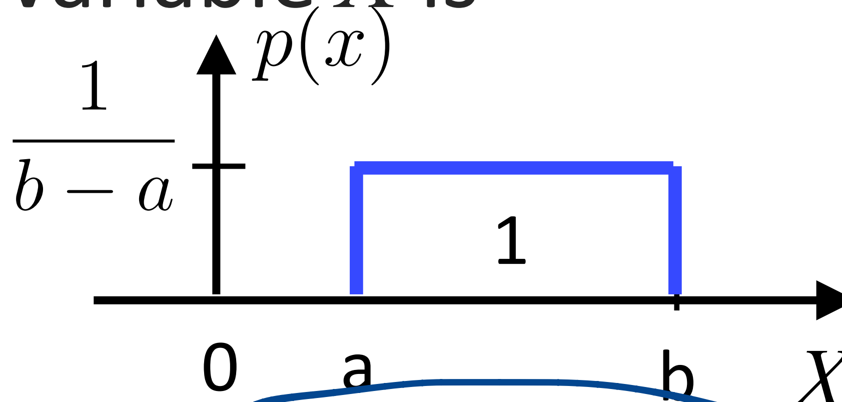
Continuous uniform distribution

- A continuous random variable X is uniform if



Continuous uniform distribution

- A continuous random variable X is uniform if



$$p(x) = \begin{cases} \frac{1}{b-a} & for\ x \in [a, b] \\ 0 & otherwise \end{cases}$$

$$E[X] = \frac{a+b}{2} \quad \& \quad var[X] = \frac{(b-a)^2}{12}$$

$$\boxed{\frac{1}{b-a}\frac{x^3}{3}\Big|_a^b - \left(\frac{a+b}{2}\right)^2}$$

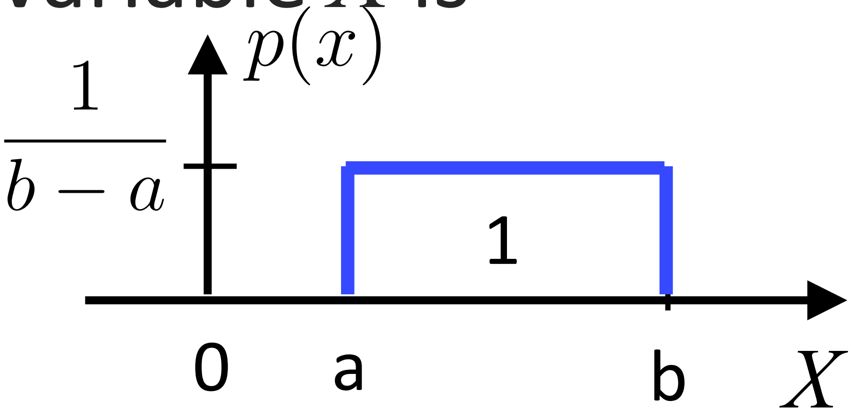
$$\underline{E[X^2]} - E^2[X]$$

$$\int_a^b x^2 \cdot p(x)dx = \int_a^b x^2 \cdot \frac{1}{b-a}dx$$

$$\frac{1}{b-a}\frac{x^3}{3}\Big|_a^b$$

Continuous uniform distribution

- A continuous random variable X is uniform if



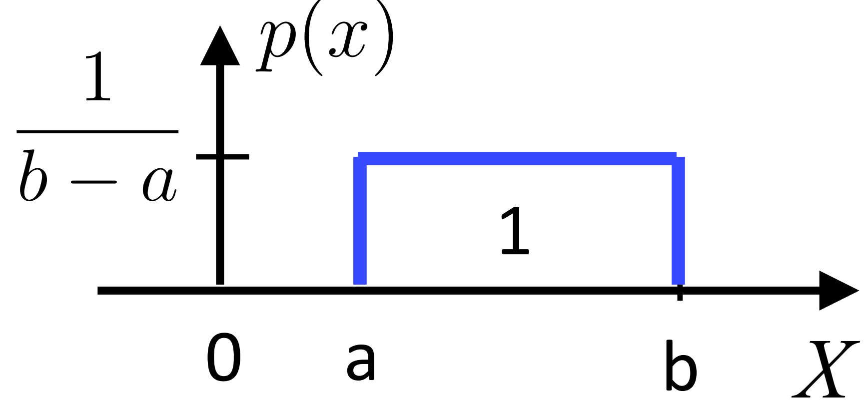
$$p(x) = \begin{cases} \frac{1}{b-a} & for\ x \in [a,b] \\ 0 & otherwise \end{cases}$$

$$E[X] = \frac{a+b}{2} \quad \& \quad var[X] = \frac{(b-a)^2}{12}$$

- Examples: 1) A dart's position thrown on the target

Continuous uniform distribution

- A continuous random variable X is uniform if



$$p(x) = \begin{cases} \frac{1}{b-a} & for\ x \in [a,b] \\ 0 & otherwise \end{cases}$$

$$E[X] = \frac{a+b}{2} \quad \& \quad var[X] = \frac{(b-a)^2}{12}$$

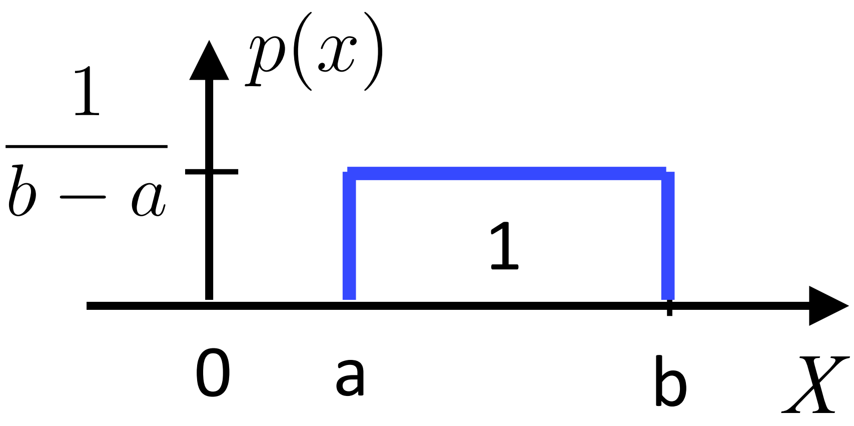
- Examples: 1) A dart's position thrown on the target 2) Often associated with **random sampling**

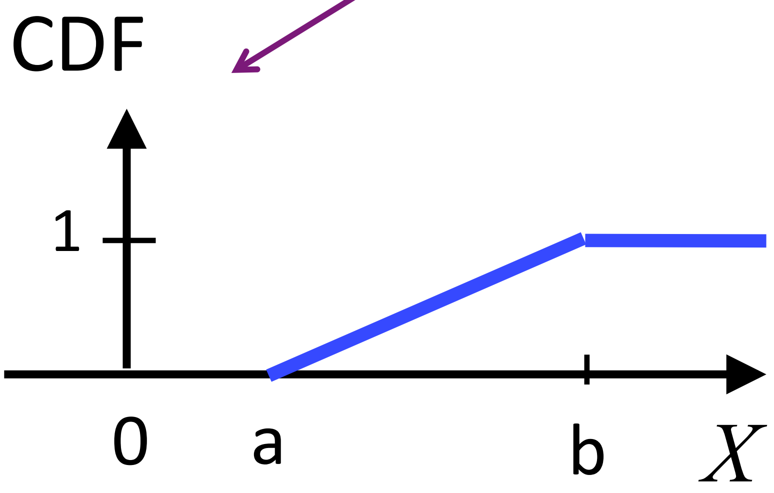
Cumulative distribution of continuous uniform distribution

- Cumulative distribution function (CDF)

$$P(X \leq x) = \int_{-\infty}^{x} p(x)dx$$

of a uniform random variable X is:





Additional References

- Charles M. Grinstead and J. Laurie Snell "Introduction to Probability"
- Morris H. Degroot and Mark J. Schervish "Probability and Statistics"

Qs for discrete distributions

Q.

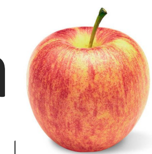
- A store staff mixed their fuji  and gala  apples and they were individually wrapped, so they are indistinguishable. Given there are 70% of fuji, if I want to know what is the probability I get 7 fuji in 20 apples? What is the distribution I should use?

A. Bernoulli
B. Binomial
C. Geometric
D. Poisson
E. Uniform

Q.

- A store staff mixed their fuji  and gala  apples and they were individually wrapped, so they are indistinguishable. Given there are 70% of fuji, if I want to know what is the probability I get 7 fuji in 20 apples? What is the distribution I should use? **What is the probability**?

Q.

- A store staff mixed their fuji  and gala  apples and they were individually wrapped, so they are indistinguishable. Given there are 70% of fuji, if I want to know the probability of picking the first gala on the 7th time (I can put back after each pick). What is the distribution I should use?

A. Bernoulli

B. Binomial

C. Geometric

D. Poisson

E. Uniform

Q.

- A store staff mixed their fuji  and gala  apples and they were individually wrapped, so they are indistinguishable. Given there are 70% of fuji, if I want to know the probability of picking the first gala on the 7^{th} time (I can put back after one pick). **What's the probability**?

Q.

- A store staff mixed their fuji  and gala  apples and they were individually wrapped, so they are indistinguishable. Given there are 70% of fuji, **what's the average times of picking to get the first gala?**

See you next time

See You!