Computer Science

CSC196: Analyzing Data

Some Continuous Probability Distributions

Jason Pacheco and Cesim Erten

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Outline

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Continuous Uniform Distribution

Example: Weather forecast.

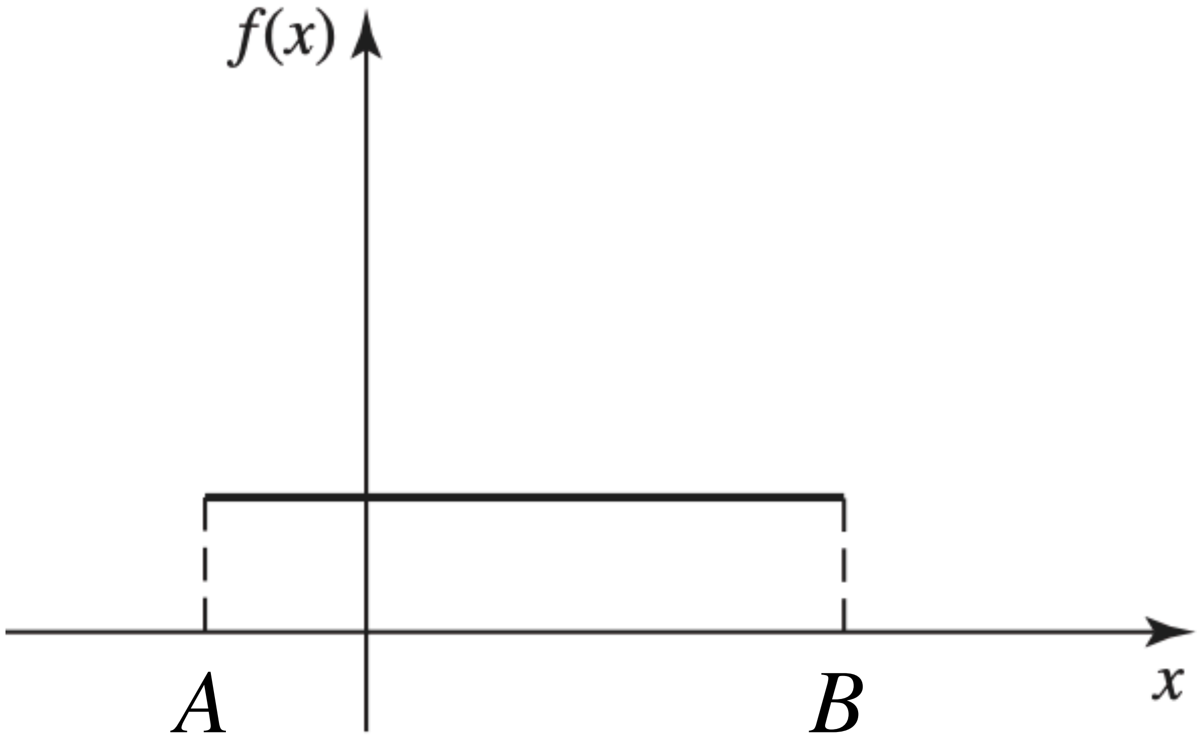
Rounded to the nearest integer.

We can assume

- The predicted temperature is equally likely any number in the interval from 55.5 and 56.5
- The temperature has continuous uniform distribution on $[55.5, 56.5]$.

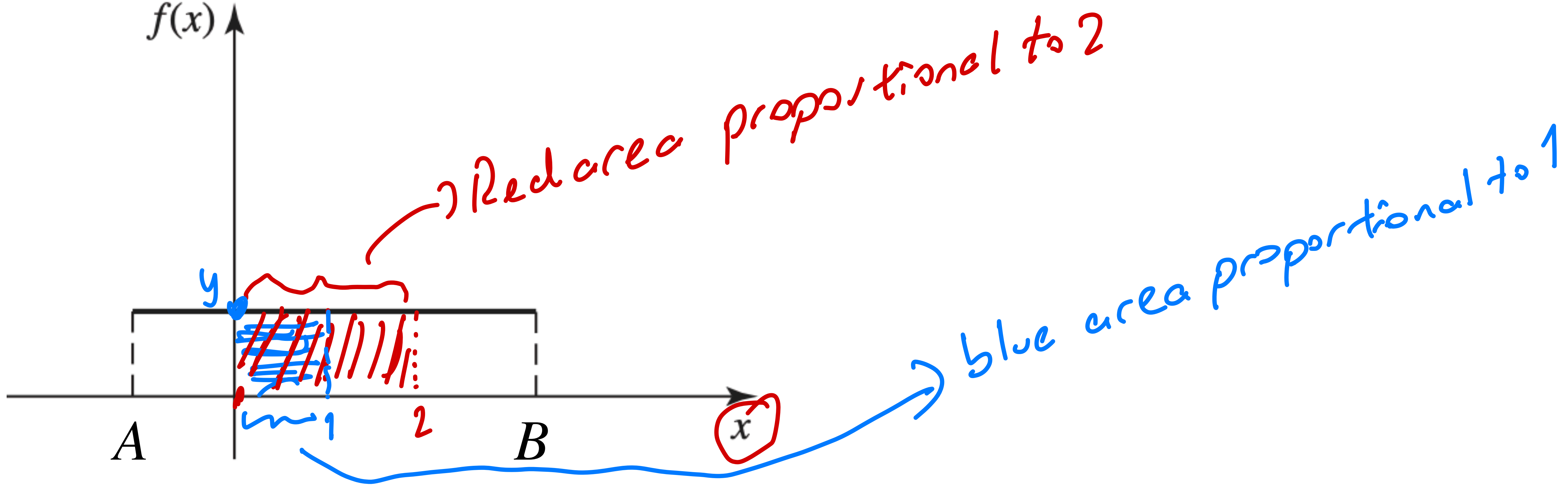
Continuous Uniform Distribution

Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



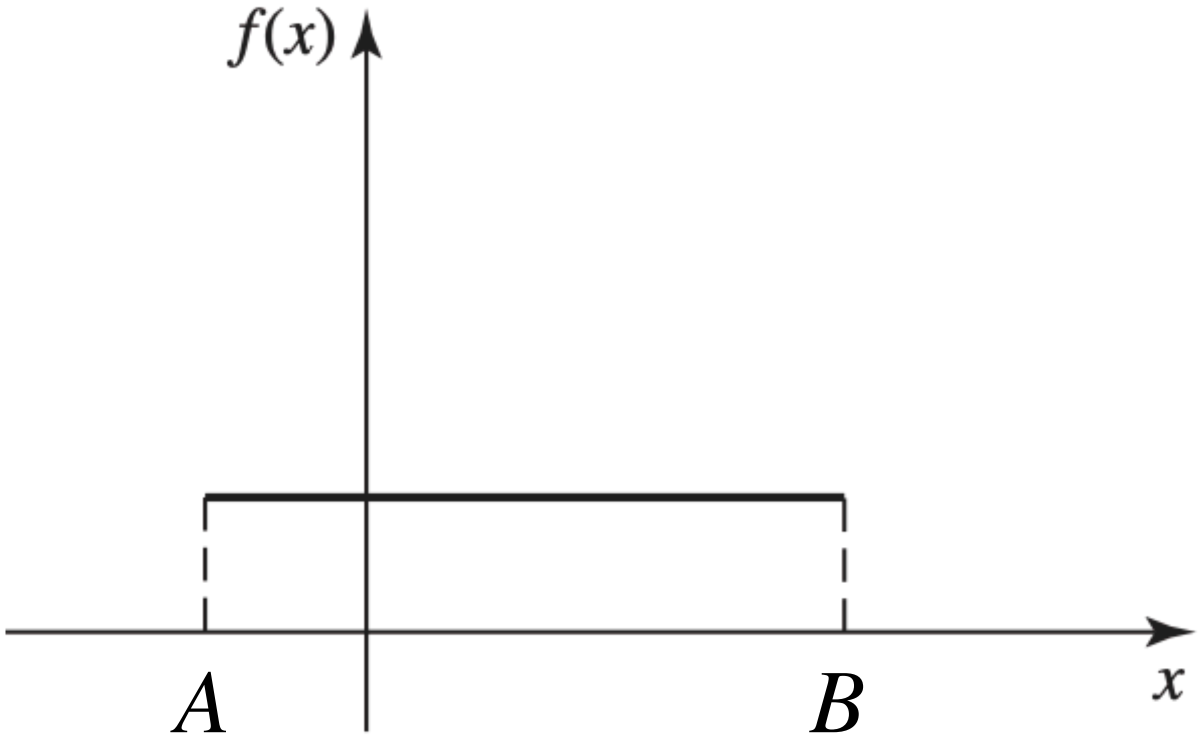
Continuous Uniform Distribution

Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



Continuous Uniform Distribution

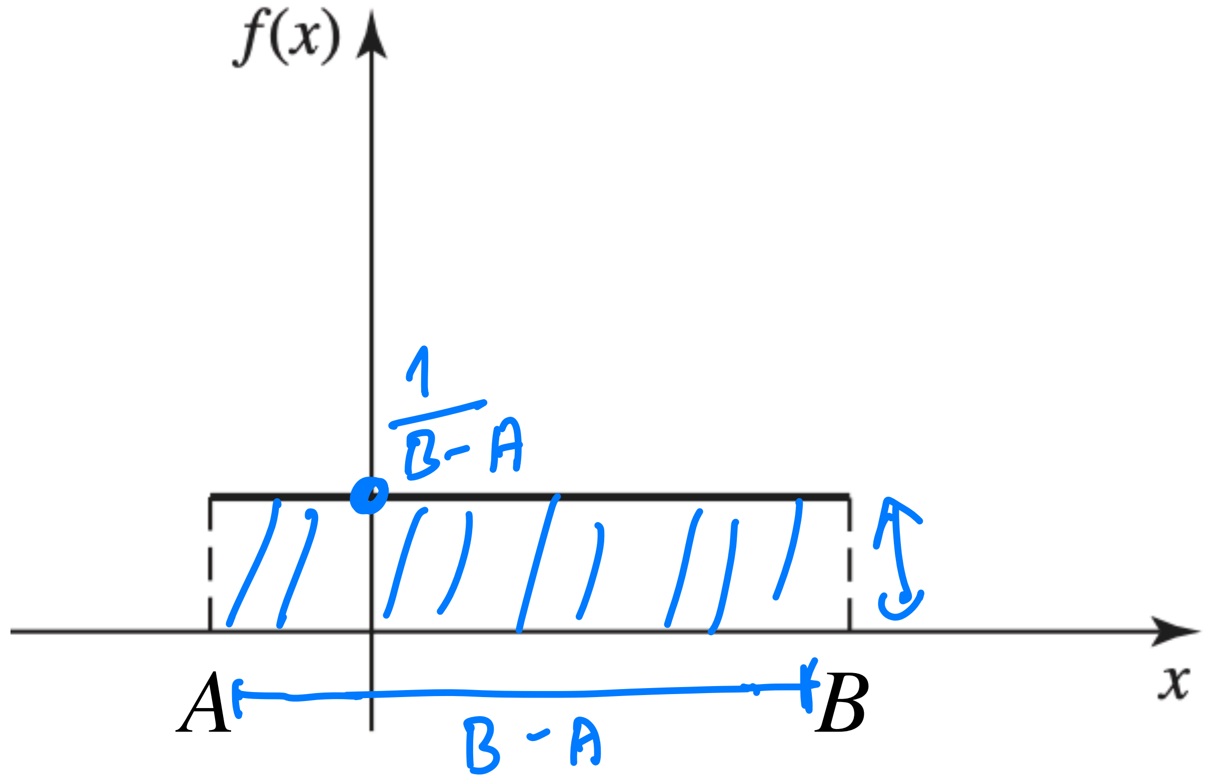
Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a **continuous uniform random variable**.



$$f(x) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

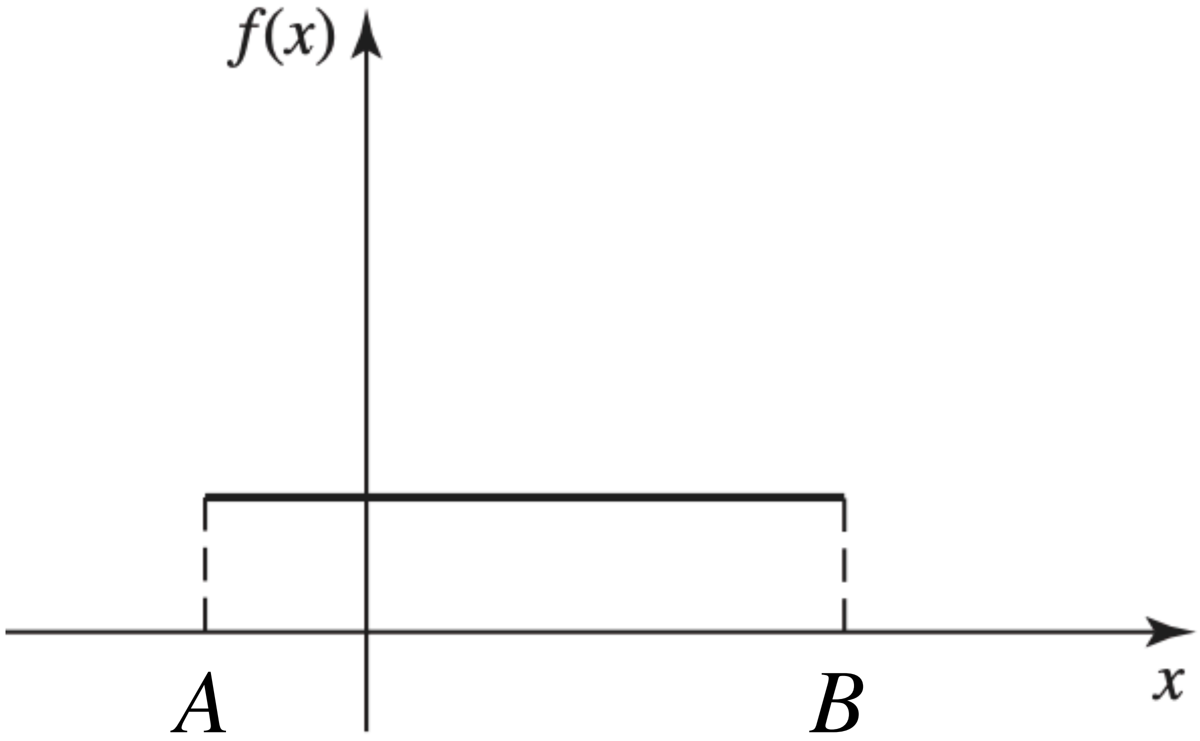
Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



$$f(x) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.

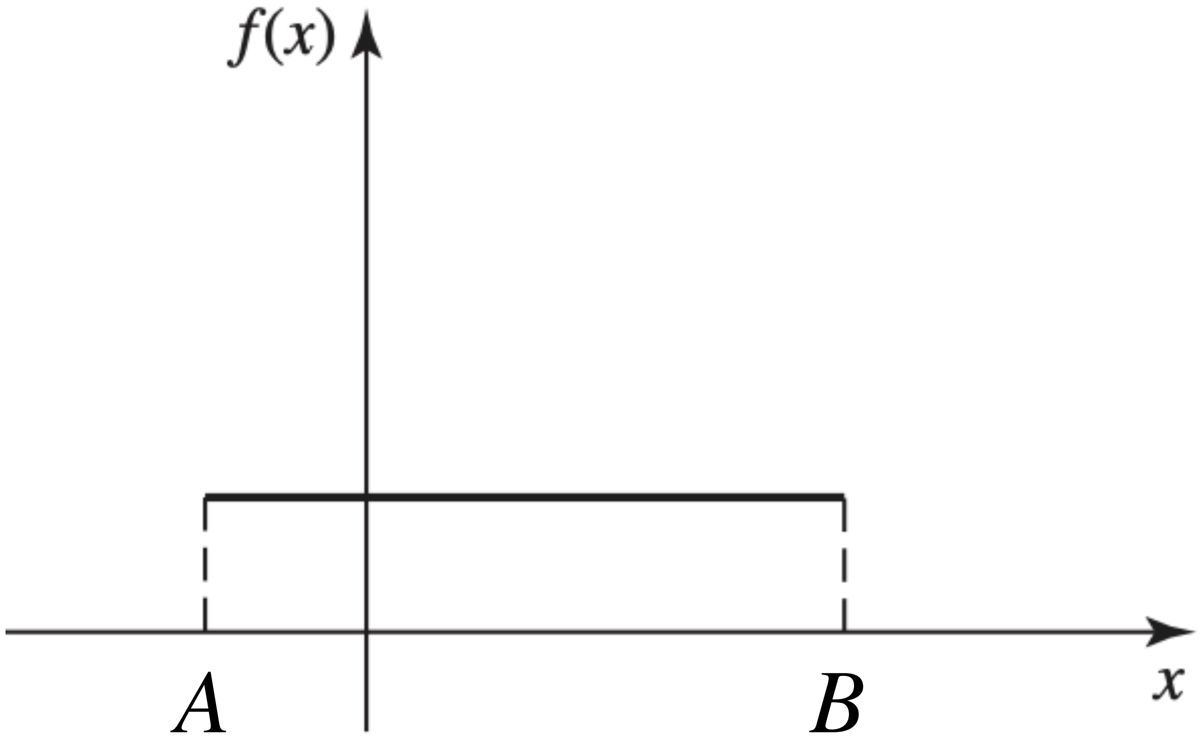


Can it be larger than 1?

$$f(x) = \begin{cases} \dfrac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Definition: If $A \leq X \leq B$ and for every subinterval of $[A, B]$ probability that X is in that subinterval is proportional to the length of that subinterval, then X is a continuous uniform random variable.



Is it a pdf: Always nonnegative?
Sum to 1?

$$f(x) = \begin{cases} \dfrac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Example: A conference room can be reserved for no more than 5 hours. Both long and short conferences occur often and it is assumed that the length X of a conference has a uniform distribution on the interval $[0, 5]$.

- The pdf?
- The probability that any given conference lasts at least 2 hours?

Continuous Uniform Distribution

Example: A conference room can be reserved for no more than 5 hours. Both long and short conferences occur often and it is assumed that the length X of a conference has a uniform distribution on the interval $[0, 5]$.

- The pdf?
- The probability that any given conference lasts at least 2 hours?

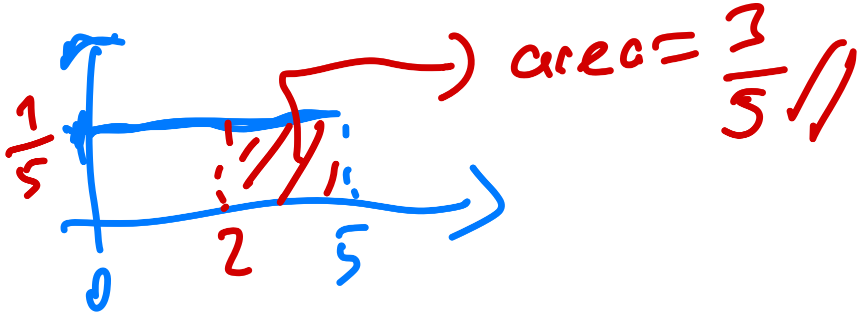
$$f(x) = \begin{cases} \frac{1}{5}, & 0 \le x \le 5 \\ 0, & \text{otherwise} \end{cases}$$

Continuous Uniform Distribution

Example: A conference room can be reserved for no more than 5 hours. Both long and short conferences occur often and it is assumed that the length X of a conference has a uniform distribution on the interval $[0, 5]$.

- The pdf?
- The probability that any given conference lasts at least 2 hours?

$$f(x) = \begin{cases} \frac{1}{5}, & 0 \le x \le 5 \\ 0, & \text{otherwise} \end{cases}$$

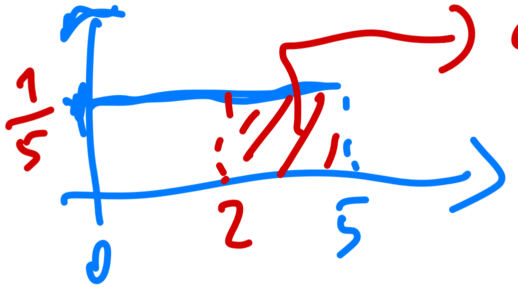


Continuous Uniform Distribution

Example: A conference room can be reserved for no more than 5 hours. Both long and short conferences occur often and it is assumed that the length X of a conference has a uniform distribution on the interval $[0, 5]$.

- The pdf?
- The probability that any given conference lasts at least 2 hours?

$$f(x) = \begin{cases} \frac{1}{5}, & 0 \le x \le 5 \\ 0, & \text{otherwise} \end{cases}$$



area $= \frac{3}{5}$ //

$$\int_2^5 f(x)\,dx = \int_2^5 \frac{1}{5}\,dx = \frac{x}{5}\Big|_2^5 = \frac{5}{5} - \frac{2}{5} = \frac{3}{5}$$ //

Continuous Uniform Distribution

Mean and Variance: It has the following mean and variance:

$$\mu = \frac{A+B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

Let’s verify the mean:

Continuous Uniform Distribution

Mean and Variance: It has the following mean and variance:

$$\mu = \frac{A+B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

Let's verify the mean: discrete $\sum x \cdot f(x)$, continuous $\int x \cdot f(x)\,dx$

Continuous Uniform Distribution

Mean and Variance: It has the following mean and variance:

$$\mu = \frac{A+B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

Let's verify the mean: discrete $\sum x \cdot f(x)$, continuous $\int x \cdot f(x) dx$

$$\int x \cdot f(x) dx = \int_A^B x \cdot \frac{1}{B-A} dx = \frac{x^2}{2 \cdot (B-A)} \Big|_A^B$$

Continuous Uniform Distribution

Mean and Variance: It has the following mean and variance:

$$\mu = \frac{A+B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

Let's verify the mean: discrete $\sum x \cdot f(x)$, continuous $\int x \cdot f(x)\,dx$

$$\int x \cdot f(x)\,dx = \int_A^B x \cdot \frac{1}{B-A}\,dx = \frac{x^2}{2 \cdot (B-A)}\Bigg|_A^B$$

$$= \frac{B^2}{2(B-A)} - \frac{A^2}{2 \cdot (B-A)} = \frac{(B-A)(B+A)}{2(B-A)} = \frac{B+A}{2}$$

Continuous Uniform Distribution

Mean and Variance: It has the following mean and variance:

$$\mu = \frac{A+B}{2} \quad \text{and} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

→ $E[x^2] - (E[x])^2$

(do it as an exercise)

Let's verify the mean: discrete $\sum x \cdot f(x)$, continuous $\int x \cdot f(x)\,dx$

$$\int x \cdot f(x)\,dx = \int_A^B x \cdot \frac{1}{B-A}\,dx = \left.\frac{x^2}{2 \cdot (B-A)}\right|_A^B$$

$$= \frac{B^2}{2(B-A)} - \frac{A^2}{2 \cdot (B-A)} = \frac{(B-A)(B+A)}{2(B-A)} = \frac{B+A}{2}$$ //

Outline

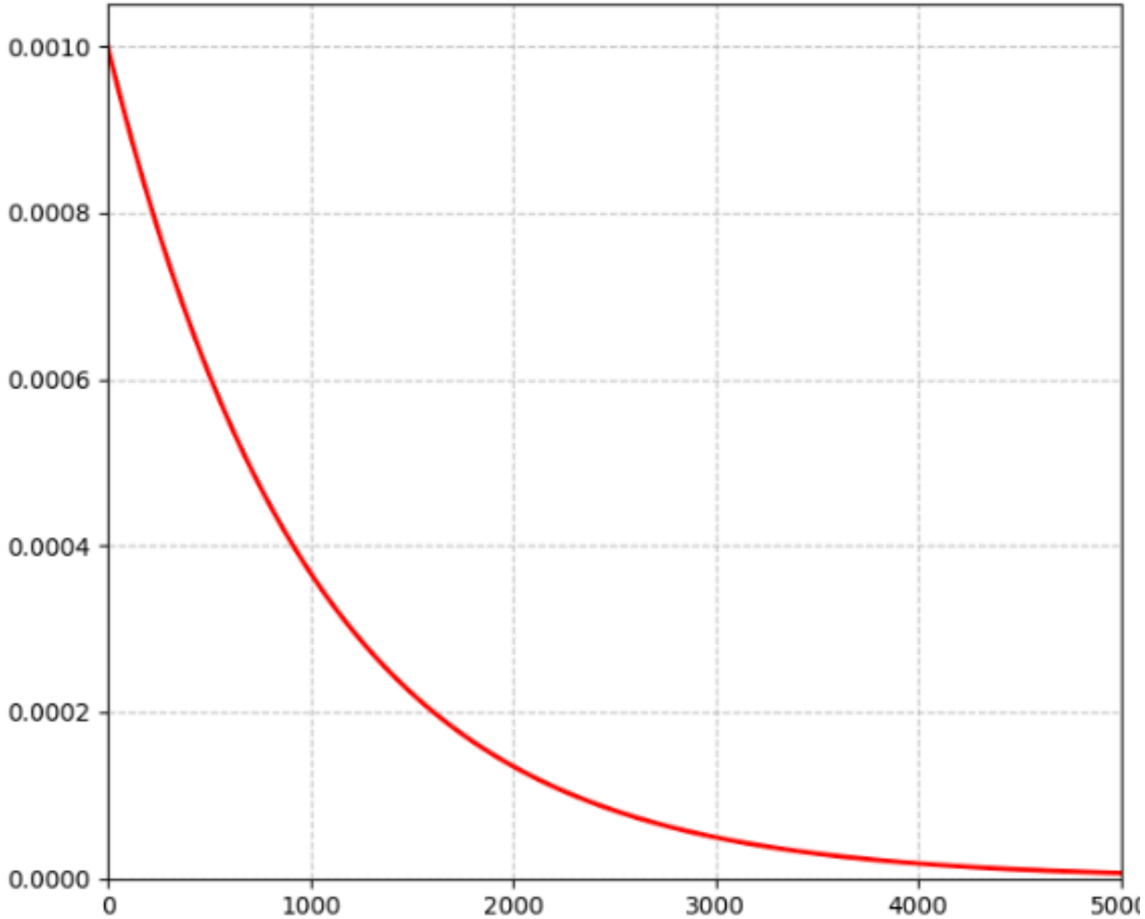
- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Normal Distribution
    - Area under the normal curve
    - Normal approximation of the binomial
- Lognormal Distribution

Exponential Distribution

Example Applications:

Time between arrivals at service facilities

Time to failure of component parts



Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

split 60 mins
into seconds:
3600 seconds.

x_i: student
arrives at
ith second.
(assume
$x_i = 0$ or $x_i = 1$)
no student student

Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

$x_1, x_2, x_3, \ldots \ldots x_{3600}$

$p = \frac{1}{1000}$ $p = \frac{1}{1000}$ . . . . $p = \frac{1}{1000}$

$\Downarrow$

p = probability that $x_1 = 1$

split 60 mins into seconds: 3600 seconds.

x_i: student arrives at ith second.

(assume $x_i = 0$ or $x_i = 1$)

$x_i = 0$: no student; $x_i = 1$: student

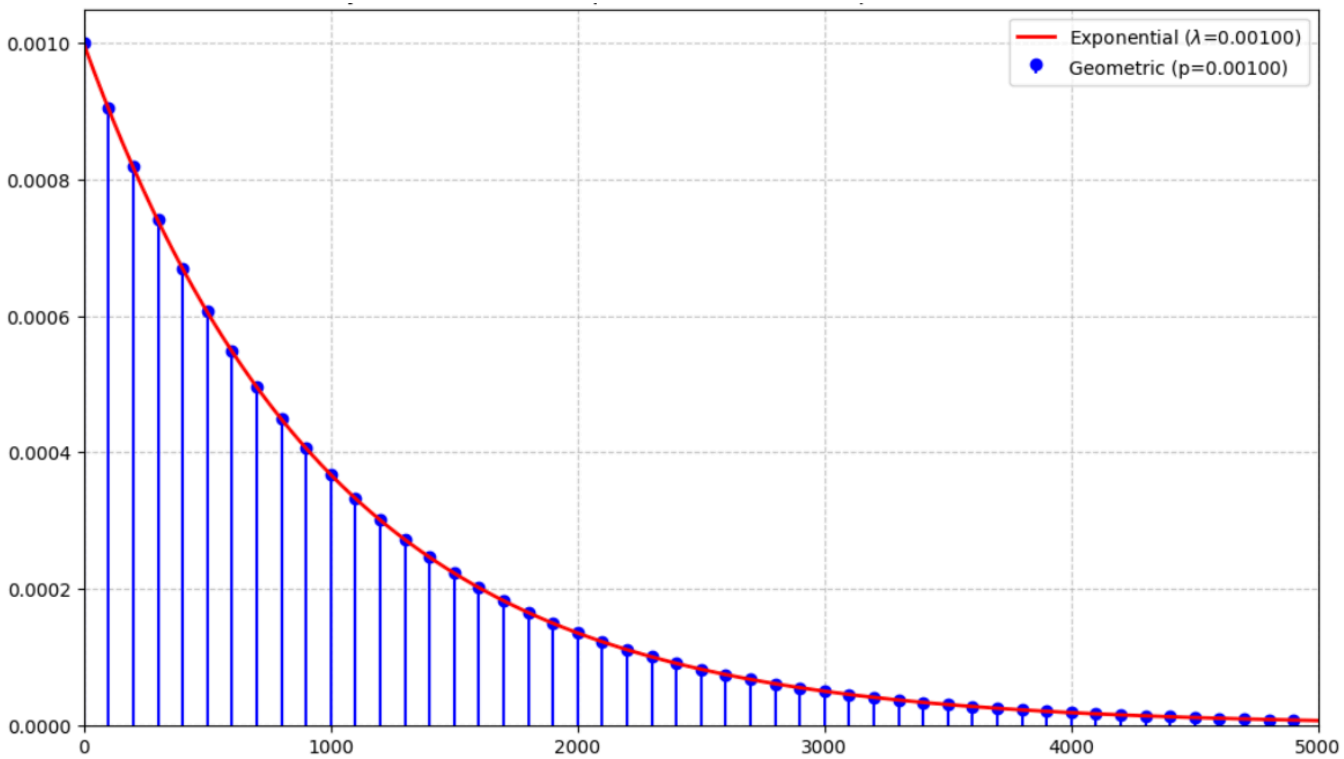
Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

Split time into minuscule units, say seconds: Each second 0 or 1 students.

Probability of x seconds until next arrival: $0.999^{x-1}0.001$



Exponential Distribution

Why does time until next arrival have exponential decay?

3.6 students per hour come to office hour. Time until next student arriving?

Split time into minuscule units, say seconds: Each second 0 or 1 students.

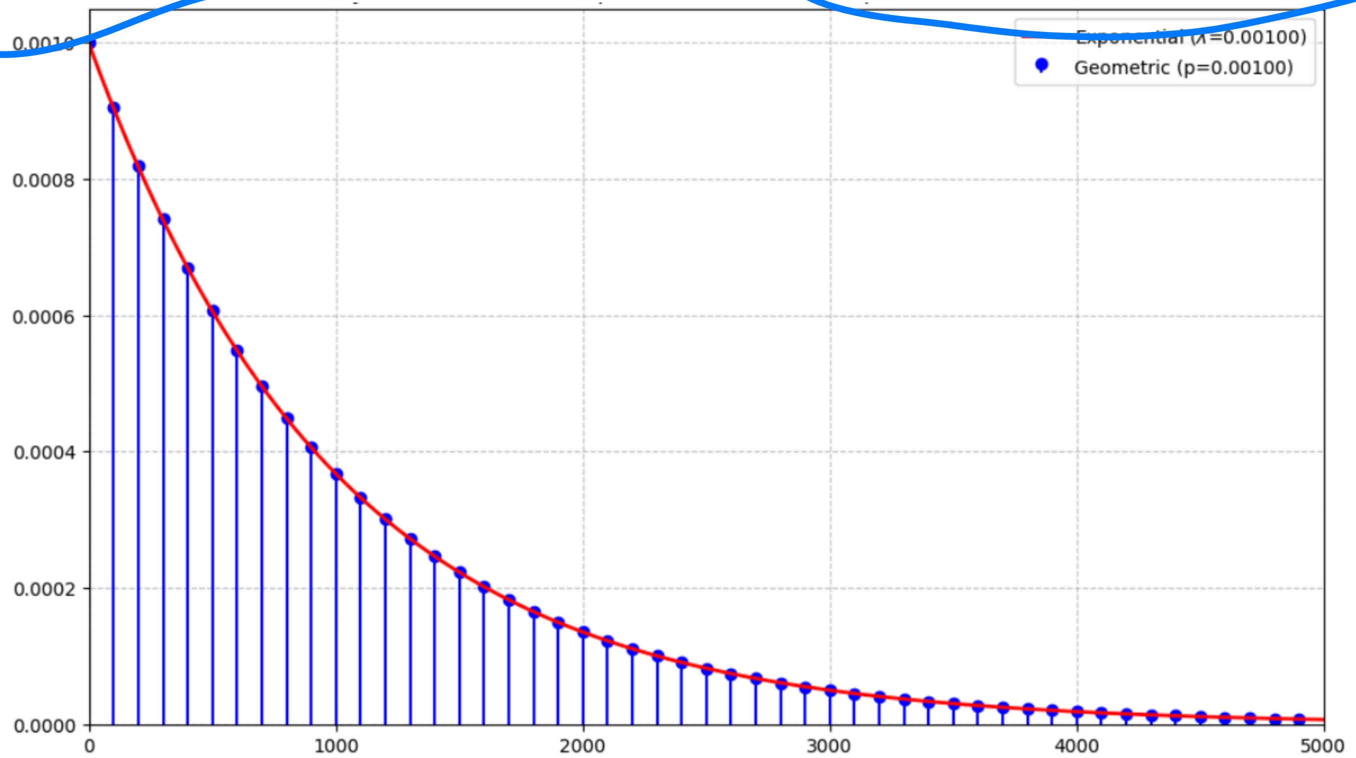
Probability of x seconds until next arrival: $0.999^{x-1}0.001$

geometric distribution

⇓

this is discrete

(an approximation of exponential distribution)



Exponential Distribution

Definition: X has exponential distribution with parameter λ, if

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

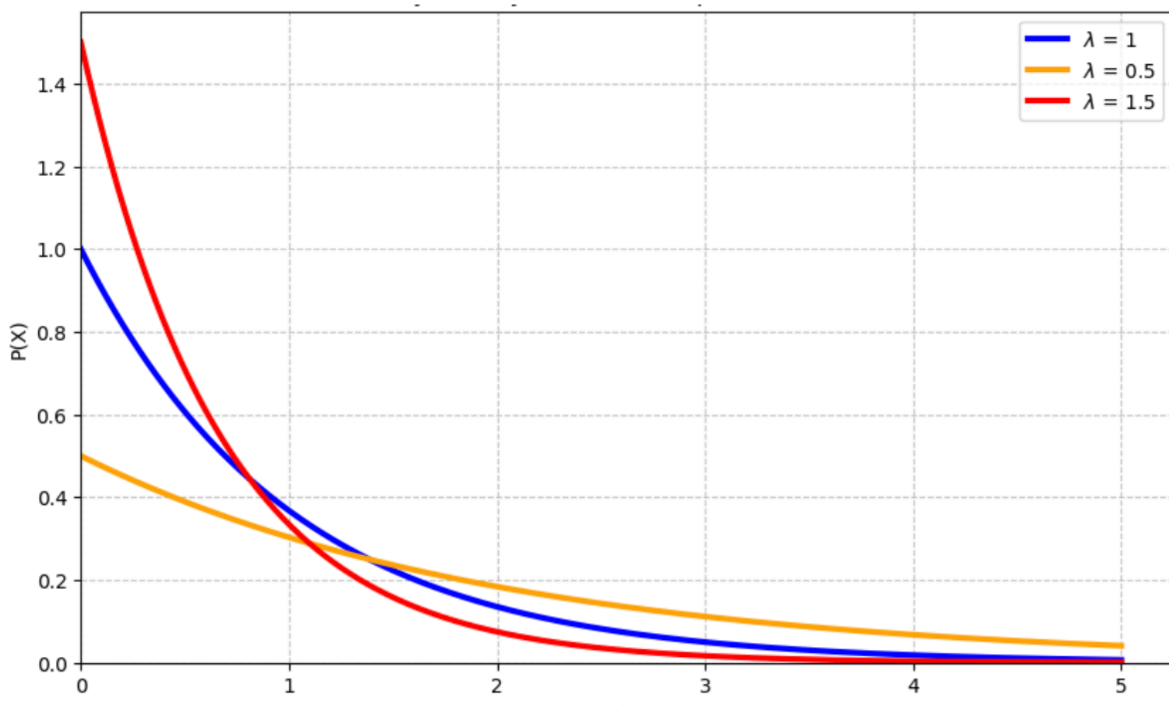
→ Time until next event nonnegative

Note: Textbook uses $1/\beta$ instead of λ.

Exponential Distribution

Definition: X has exponential distribution with parameter λ, if

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$



Note: Textbook uses $1/\beta$ instead of λ.

Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✔

(Note that $\int e^{ax} = \frac{1}{a} \cdot e^{ax}$ for constant a.)

Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$

What is the cdf?

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}$$

$0 - (-e^{\lambda x}) = e^{-\lambda x}$

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

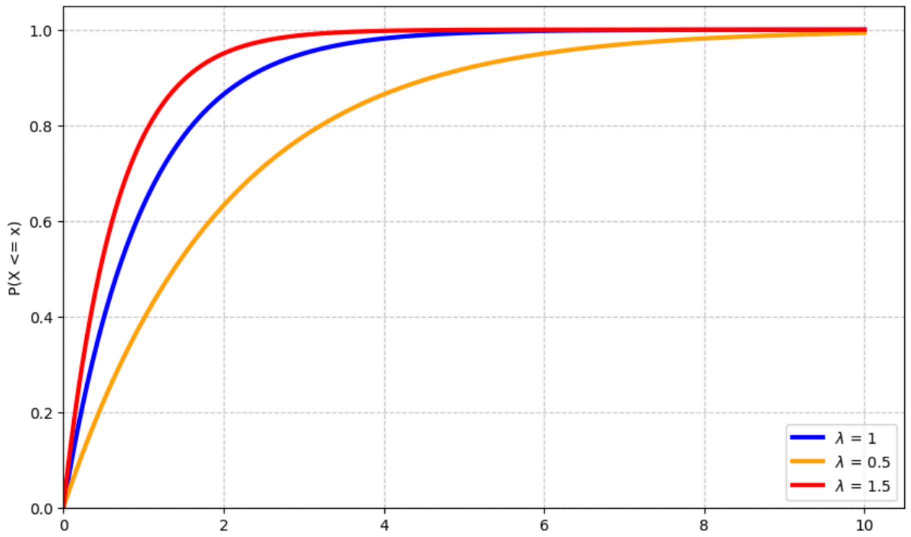
Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✔

What is the cdf?

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}$$

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



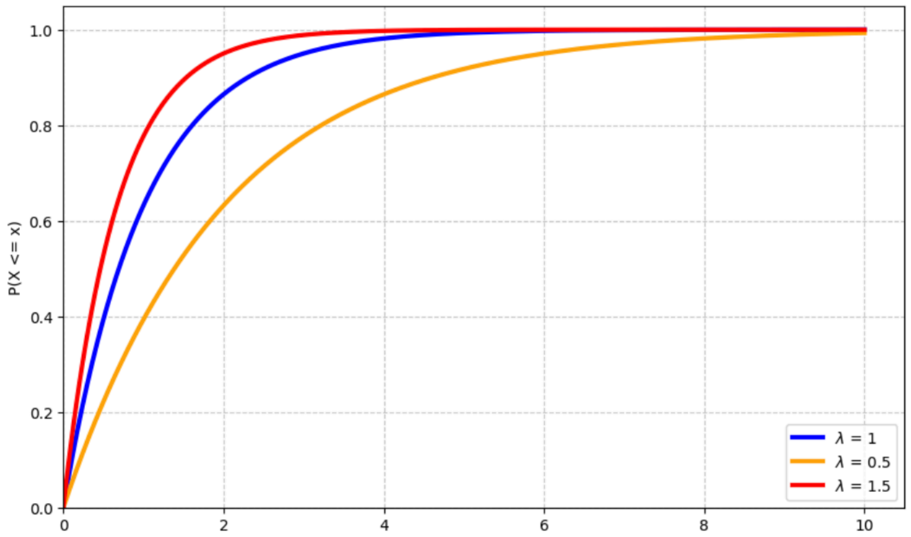
Exponential Distribution

Does it add to 1? $\int_0^\infty \lambda e^{-\lambda x} = -e^{-\lambda x}\Big|_0^\infty = 1$ ✔

What is the cdf?

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_x^\infty = e^{-\lambda x}$$

$$\Rightarrow P(X \le x) = 1 - P(X > x) = 1 - e^{-\lambda x}$$



Mean and Variance: $\mu = \frac{1}{\lambda},\ \sigma^2 = \frac{1}{\lambda^2}$

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Number of random arrivals during a period of time $\longrightarrow$ Poisson distribution

Time until occurrence of next Poisson event $\longrightarrow$ Exponential distribution

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Recall with Poisson random variable X: $$Pois(X, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$\Rightarrow \quad \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$ is the probability that x events occur in t units of time.

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Recall with Poisson random variable X: $Pois(X, \lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

$\Rightarrow \dfrac{(\lambda t)^x e^{-\lambda t}}{x!}$ is the probability that x events occur in t units of time.

Y: time to first occurrence.

$P(Y > y) = e^{-\lambda y}$ (probability of 0 events up to time y)

$P(0 \leq Y \leq y) = 1 - e^{-\lambda y}$ is cdf of Y. Then $f(y) = \lambda e^{-\lambda y}$

Compare with pdf of Exponential.

Exponential Distribution and Poisson

Relationship between Exponential and Poisson

Note that λ is the mean of Poisson and $\dfrac{1}{\lambda}$ is the mean of Exponential.

If on average 4 events occur in 1 hour, time for first occurrence of the event on average is $1/4$ hours.

Exponential Distribution is Memoryless

$P(X \geq i + a \mid X \geq i)$ is the same as $P(X \geq a)$.

Example:

Arrival of your bus due to a Poisson process, rather than a fixed schedule.

Say you already waited for $i = 10$ mins. Probability that you wait ≥ 5 more mins is the same as the probability of you originally waiting for ≥ 5 mins (Consider a new passenger who comes after you waited for 10 mins.)

Note: This doesn't apply to cases where wear and tear is a factor.

Exponential Distribution is Memoryless

$P(X \geq i + a \mid X \geq i)$ is the same as $P(X \geq a)$.

Proof:

Exponential Distribution is Memoryless

$P(X \geq i + a \mid X \geq i)$ is the same as $P(X \geq a)$.

Proof:

$$P(x \geq i+a \mid x \geq i) = \frac{P(x \geq i+a,\ x \geq i)}{P(x \geq i)}$$

$$= \frac{P(x \geq i+a)}{P(x \geq i)}$$

$$= \frac{e^{-\lambda(i+a)}}{e^{-\lambda i}} = \frac{e^{-\lambda i} \times e^{-\lambda a}}{e^{-\lambda i}} = e^{-\lambda a}$$