Mathematics IV
(Probability and Statistics)

MATH 403

Lecture 6

Dr. Phoebe Edward Nashed

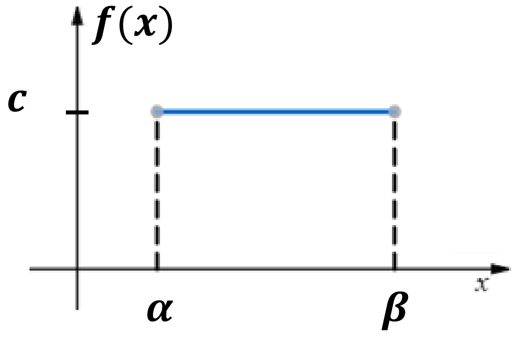
Lecture 6 - Outline

- Special Continuous Distributions
  - Uniform Distribution
  - Exponential Distribution
  - Normal (Gaussian) Distribution

I. Uniform Distribution

- Suppose that a continuous random variable $\boldsymbol{X}$ can assume values only in a bounded interval, say the open interval $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and suppose that the probability density function (PDF) is **constant**, say

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c} \qquad \text{for} \qquad \boldsymbol{\alpha} \leq \boldsymbol{x} \leq \boldsymbol{\beta}$$



- This implies that $\boldsymbol{c} = \mathbf{1}/(\boldsymbol{\beta} - \boldsymbol{\alpha})$, since

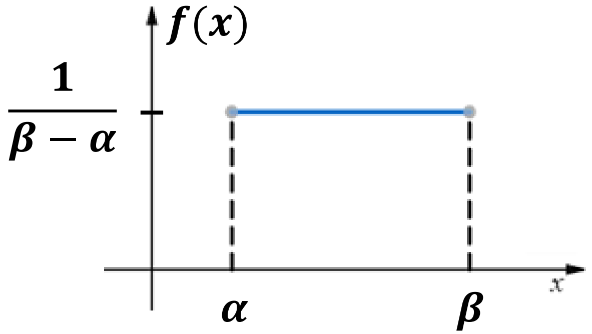
$$\int_{-\infty}^{\infty} \boldsymbol{f}(\boldsymbol{x})\ \boldsymbol{dx} = \int_{\boldsymbol{\alpha}}^{\boldsymbol{\beta}} \boldsymbol{c}\ \boldsymbol{dx} = \boldsymbol{c}\,(\boldsymbol{\beta} - \boldsymbol{\alpha}) = 1$$

- If we define $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0}$ outside the interval, then we obtain $\boldsymbol{f}(\boldsymbol{x}) \geq \mathbf{0}$ for all $\boldsymbol{x}$.

I. Uniform Distribution – Def. - PDF

- Definition 1: A continuous random variable X is said to have *uniform distribution* with parameters α and β $(\alpha < \beta)$ if and only if its PDF is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{for} \quad \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

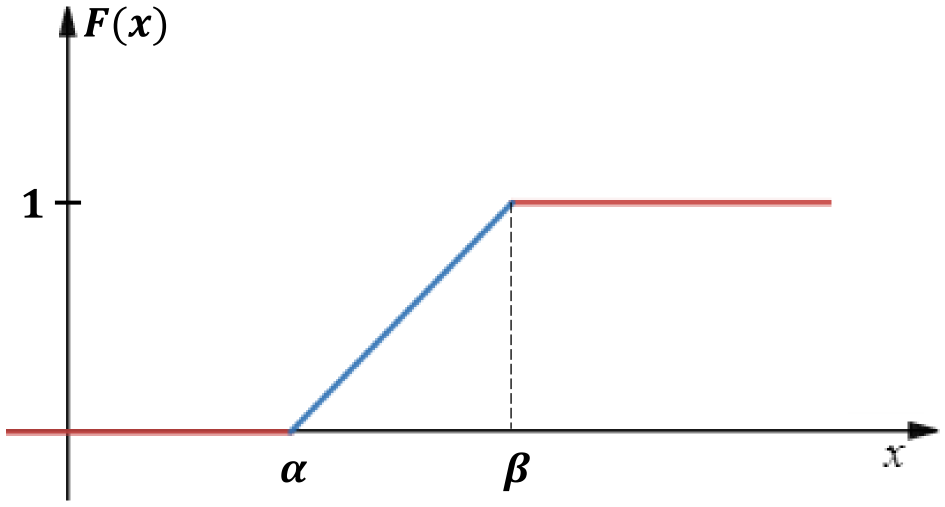


- This distribution provides a probability model for selecting a point "at random" from an interval (α, β).

I. Uniform Distribution – CDF

❖ The cumulative distribution function (CDF) of a uniformly distributed random variable X has the form:

$$F(x) = \begin{cases} 0 & \text{for} \quad x \leq \alpha \\ \dfrac{x-\alpha}{\beta-\alpha} & \text{for} \quad \alpha < x < \beta \\ 1 & \text{for} \quad x \geq \beta \end{cases}$$



I. Uniform Dist. – Mean and Var.

- The mean and the variance of *uniform distribution* are given by:

$$E(X) = \mu = \frac{\beta + \alpha}{2} \quad \text{and} \quad Var(X) = \sigma^2 = \frac{(\beta - \alpha)^2}{12}$$

I. Uniform Dist. – Mean

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\; dx$$

$$= \int_{\alpha}^{\beta} x \left(\frac{1}{\beta - \alpha}\right) dx = \left(\frac{1}{\beta - \alpha}\right) \frac{x^2}{2}\Bigg|_{\alpha}^{\beta} = \left(\frac{1}{\beta - \alpha}\right) \frac{\beta^2 - \alpha^2}{2} = \frac{\beta + \alpha}{2}$$

I. Uniform Dist. – Variance

- $E(X) = \frac{\beta+\alpha}{2}$

$$E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\; dx$$

$b^3 - a^3$
$= (b-a)(b^2 + ab + a^2)$

$$= \int_{\alpha}^{\beta} x^2 \left(\frac{1}{\beta-\alpha}\right) dx = \left(\frac{1}{\beta-\alpha}\right)\frac{x^3}{3}\Bigg|_{\alpha}^{\beta} = \left(\frac{1}{\beta-\alpha}\right)\frac{\beta^3-\alpha^3}{3} = \frac{\beta^2+\beta\alpha+\alpha^2}{3}$$

- $\therefore Var(X) = E(X^2) - [E(X)]^2$

$$Var(X) = \frac{\beta^2+\beta\alpha+\alpha^2}{3} - \left(\frac{\beta+\alpha}{2}\right)^2 = \frac{\beta^2-2\beta\alpha+\alpha^2}{12} = \frac{(\beta-\alpha)^2}{12}$$

Example 1

- Suppose a train arrives at a subway station regularly every $\mathbf{10}$ min. If a passenger arrives at the station without knowing the timetable, then the waiting time to catch the train is uniformly distributed with density

$$\boldsymbol{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{10}} & \text{for} \quad \mathbf{0} \le \boldsymbol{x} \le \mathbf{10} \\ \mathbf{0} & \text{otherwise} \end{cases}$$

i. Calculate the “average” waiting time.

ii. Calculate the probability of waiting for the train for less than $\mathbf{3}$ min.

Example 1 – Sol.

- Since X is uniformly distributed with PDF:

$$\boldsymbol{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{10}} & \text{for} \quad \mathbf{0 \le x \le 10} \\ \mathbf{0} & \text{otherwise} \end{cases}$$

- Therefore, $\alpha = 0, \beta = 10$.
- The "average" waiting time is

$$\boldsymbol{E(X) = \frac{\beta+\alpha}{2} = (10+0)/2 = 5} \text{ min.}$$

Example 1 – Sol.

- Since X is uniformly distributed with PDF:

$$\boldsymbol{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{10}} & \text{for} \quad \mathbf{0 \le x \le 10} \\ \mathbf{0} & \text{otherwise} \end{cases}$$

- Therefore, $\alpha = 0, \beta = 10.$
- The probability of waiting for the train for less than 3 min is obviously $\mathbf{0.3}$ (or $\mathbf{30}\,\%$) which can be calculated as follows:

$$\boldsymbol{P(X \le 3) = F(3) = \frac{3-0}{10-0} = 0.3}$$

$$\boldsymbol{F(x) = \frac{x-\alpha}{\beta-\alpha}}$$

Example 2

- Suppose the time it takes a data collection operator to fill out an electronic form for a database is uniformly between $\mathbf{1.5}$ and $\mathbf{2.2}$ minutes.

(a) What is the mean and the variance of time, it takes an operator to fill out the form?

(b) What is the probability that it takes less than two minutes to fill out the form?

Example 2 – Sol.

- Since X is uniformly between $\mathbf{1.5}$ and $\mathbf{2.2}$ minutes; $\alpha = 1.5, \beta = 2.2$
- Therefore,

$$\boldsymbol{E(X) = \mu = \frac{\beta + \alpha}{2} = \frac{2.2 + 1.5}{2} = 1.85}$$

$$\boldsymbol{Var(X) = \sigma^2 = \frac{(\beta - \alpha)^2}{12} = \frac{(2.2 - 1.5)^2}{12} = 0.041}$$

- The probability that it takes less than two minutes to fill out the form:

$$\boldsymbol{P(X \leq 2) = F(2) = \frac{2 - 1.5}{2.2 - 1.5} = 0.714}$$

$$\boldsymbol{F(x) = \frac{x - \alpha}{\beta - \alpha}}$$

Example 3

- Suppose $\boldsymbol{X}$ has a continuous uniform distribution over the interval $[-\mathbf{1},\mathbf{1}]$. Find the value for $\boldsymbol{k}$ such that

$$\boldsymbol{P(-k \leq X \leq 2k) = 0.6}$$

Example 3 – Sol.

- Suppose $\boldsymbol{X}$ has a continuous uniform distribution over the interval $[-\mathbf{1},\mathbf{1}]$. Find the value for $\boldsymbol{k}$ such that

$$\boldsymbol{P(-k \leq X \leq 2k) = 0.6}$$

- Since X is uniformly distributed, $\alpha = -1, \beta = 1,$
- Therefore, its PDF is:

$$\boldsymbol{f(X) = \frac{1}{\beta - \alpha} = \frac{1}{2} \qquad -1 \leq x \leq 1}$$

$$\boldsymbol{P(-k \leq X \leq 2k) = \int_{-k}^{2k} 0.5 \; dx = 0.5(2k + k) = 1.5\,k = 0.6}$$

$$\boldsymbol{\therefore k = 0.6/1.5 = 0.4}$$

Example 3 – Another Sol.

- Suppose $\boldsymbol{X}$ has a continuous uniform distribution over the interval $[-\mathbf{1},\mathbf{1}]$. Find the value for $\boldsymbol{k}$ such that

$$\boldsymbol{P(-k \le X \le 2k) = 0.6}$$

- Since X is uniformly distributed, $\alpha = -1, \beta = 1$,
- Therefore, its PDF is:

$$\boldsymbol{f(X) = \frac{1}{\beta - \alpha} = \frac{1}{2} \qquad -1 \le x \le 1}$$

$$\boldsymbol{P(-k \le X \le 2k) = F(2k) - F(-k) = \left(\frac{2k+1}{2}\right) - \left(\frac{-k+1}{2}\right) = 1.5\,k}$$

$$\boldsymbol{\therefore k = 0.6/1.5 = 0.4}$$

Poisson Distribution – Review

- Poisson distribution is used to approximate Binomial probabilities when the number of trials $\boldsymbol{n}$ is large ($\boldsymbol{n} \to \infty$), the probability of success $\boldsymbol{p}$ is small ($\boldsymbol{p} \to \mathbf{0}$) and $\boldsymbol{np}$ remains constant.
- Such an approximation is acceptable, say, for $\boldsymbol{n} \geq \mathbf{30}$ and $\boldsymbol{p} \leq \mathbf{0.05}$, and it becomes more accurate for larger $\boldsymbol{n}$.

Poisson Distribution – Def. - Review

- A random variable $\boldsymbol{X}$ is said to have ***Poisson distribution*** with parameter $\boldsymbol{\lambda > 0}$
  if and only if its probability mass function (PMF) is given by

$$\boldsymbol{f(x) = \frac{e^{-\lambda}\ \lambda^{x}}{x!}} \quad \text{for} \quad \boldsymbol{x = 0, 1, 2, \cdots}$$

Poisson Distribution – Mean and Variance - Review

❖ Hence, the mean and the variance of Poisson distribution are given by

$$\boldsymbol{E(X) = \mu = \lambda}$$

$$\boldsymbol{Var(X) = \sigma^2 = \lambda}$$

As $\boldsymbol{q \to 1}$

Illustrative Example

- Imagine that you are a traffic engineer and you need to model the random variable that represents the number of cars that pass a certain point in an hour.
- Assume the following:
  - The number of cars passing in any hour does not affect another hour (independent events/occurrences)
  - The probability of success (car passing) is constant among different hours, it is p.
- You can try to model it by a Binomial distribution, because every hour, either success (car passes) or failure (car does not pass) occurs.
- So, $\mathrm{E(X)} = \lambda = np = 60.\frac{\lambda}{60}$

Illustrative Example

- You can try to model it by a Binomial distribution, because every hour, either success (car passes) or failure (car does not pass) occurs.
- So, $\mathrm{E(X)} = \lambda = np = 60.\frac{\lambda}{60}$
- $P(X = k) = \binom{60}{k}\left(\frac{\lambda}{60}\right)^k\left(1 - \frac{\lambda}{60}\right)^{60-k}$
- But, what if two or more cars pass within the same minute?! We have to increase granularity… using seconds…
- $P(X = k) = \binom{3600}{k}\left(\frac{\lambda}{3600}\right)^k\left(1 - \frac{\lambda}{3600}\right)^{3600-k}$
- But, what if two or more cars pass within the same second?! We have to increase granularity…
- As $n \to \infty$ and $p \to 0$, the Binomial distribution will become **Poisson Distribution** (as proved last lecture).

II. Exponential Distribution

- Exponential distribution is often used to model time: waiting time, interarrival time, hardware lifetime, failure time, time between telephone calls, etc. As we shall see below, in a sequence of rare events, when the number of events is **Poisson**, the time between events is **Exponential**.

Relation between Poisson Dist. and Exponential Dist.

- Conditions for a **Poisson Process**:
  - Events must occur at a **constant rate** (constant λ).
  - Events must be **independent**.
  - Events **cannot** occur **simultaneously**.

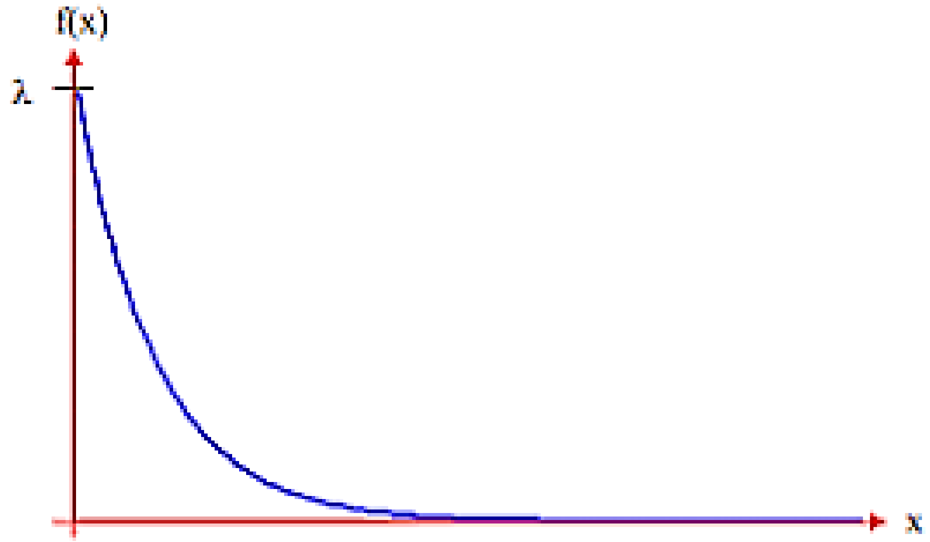
Relation between Poisson Dist. and Exponential Dist.

Exponential Distribution	Poisson Distribution
Continuous distribution	Discrete distribution
It represents time between a number of events in a Poisson Process (inter-arrival time)	It represents number of events (occurrences) in a time interval (rare events, $p \to 0$).
Parameter: λ: rate/frequency of occurrences per unit time – unit: 1/time (sec^{-1}, min^{-1}, ... etc.)	Parameter: λ: average (expected) number of events (occurrences)
Example: Time between two cars' arrivals.	Example: Number of cars passing a certain square in an hour.

II. Exponential Distribution – Def. - PDF

❖ A random variable X is said to have *Exponential distribution* with parameter $\lambda > 0$ if and only if its PDF is given by

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



II. Exponential Distribution – Def. - CDF

- Its PDF is:

$$f(x)=\begin{cases}\lambda\, e^{-\lambda x} & \text{for} \quad x>0\\[2ex] 0 & \text{otherwise}\end{cases}$$

- Its CDF can be obtained by:

$$F(x)=\int_{-\infty}^{x} f(t)\;dt=\int_{0}^{x}\lambda\, e^{-\lambda t}\;dt=\lambda\left(\left.\frac{e^{-\lambda t}}{-\lambda}\right|_{0}^{x}\right)=1-e^{-\lambda x}$$

$$\therefore\; F(x)=\begin{cases}1-e^{-\lambda x} & \text{for} \quad x>0\\[2ex] 0 & \text{for} \quad x\le 0\end{cases}$$

II. Exponential Distribution – Mean

❖ Its PDF is:

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

❖ Its Mean is:

$$E(X) = \mu = \int_{-\infty}^{\infty} x\, f(x)\; dx = \int_{0}^{\infty} x\left(\lambda\, e^{-\lambda x}\right)\; dx = \frac{1}{\lambda}$$

Using integration by parts

II. Exponential Distribution – Variance

- Its PDF is:

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

- Its Mean is: $E(X) = \frac{1}{\lambda}$

$$E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\; dx = \int_{0}^{\infty} x^2\, (\lambda\, e^{-\lambda x})\; dx = \frac{2}{\lambda^2}$$

Using integration by parts twice

- Its Variance:

$$\therefore Var(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

II. Exponential Dist. – Mean and Var.

$$E(X) = \mu = \frac{1}{\lambda}$$

$\mu = \sigma$

$$Var(X) = \sigma^2 = \frac{1}{\lambda^2}$$

II. Exponential Dist. - λ Parameter

- The quantity $\boldsymbol{\lambda}$ is a parameter of *Exponential distribution*.
- It is clear from the fact that $\boldsymbol{E(X) = 1/\lambda}$, that λ is the **reciprocal** of the mean of the exponentially distributed random Variable X.
- If $\boldsymbol{X}$ is time, measured in minutes, then $\boldsymbol{\lambda}$ is a **frequency (rate)**, measured in $\text{min}^{-\mathbf{1}}$.
- For example, if arrivals occur every half a minute, on the average, then $\boldsymbol{E(X) = 0.5}$, therefore $\boldsymbol{\lambda = 2}$, meaning that they occur with a frequency (arrival rate) of $\mathbf{2}$ arrivals per minute.
- This $\boldsymbol{\lambda}$ has the **<u>same</u>** meaning as the parameter of **Poisson distribution**.
- $\boldsymbol{\lambda} =$ Frequency parameter, the **<u>number of events per time unit</u>**.

Example 4

❖ Jobs are sent to a printer at an average rate of 3 jobs per hour.

(a) What is the expected time between jobs?

(b) What is the probability that the next job is sent within 5 minutes?

Example 4 – Sol.

❖ Job arrivals represent rare events, thus the time $\boldsymbol{X}$ between them is Exponential with the given parameter $\boldsymbol{\lambda} = \mathbf{3}\ \text{hr}^{-\mathbf{1}}$ (jobs per hour).

a) $E(X) = 1/\lambda = 1/3$ hr or 20 min between jobs;

b) Convert to the same measurement unit: $\mathbf{5}$ min $= (\mathbf{1/12})$ hr. Then,

$$\boldsymbol{P(X < 1/12) = F(1/12) = 1 - e^{-\lambda(1/12)} = 1 - e^{-1/4} = 0.2212}$$

Example 5

- The time to failure of fans in a PC can be modeled by an exponential distribution with mean $\mathbf{4000}$ hours. What proportion of the fans will last at least $\mathbf{10,000}$ hours?

Example 5 – Sol.

- The time to failure of fans in a PC can be modeled by an exponential distribution with mean $\mathbf{4000}$ hours. What proportion of the fans will last at least $\mathbf{10,000}$ hours?

$$\boldsymbol{\mu = 4000 \qquad \rightarrow \qquad \lambda = 1/4000}$$

$$\boldsymbol{P(X \geq 10,000) = 1 - F(10,000)}$$

$$\boldsymbol{= 1 - [1 - e^{-(10000/4000)}] = e^{-(2.5)} = 0.082}$$

- Another Approach

$$\boldsymbol{P(X \geq 10,000) = \int_{10000}^{\infty} \frac{1}{4000} e^{-(x/4000)}\, dx = e^{-(2.5)} = 0.082}$$

II. Exp. Dist. - Memoryless Property

❖Theorem

❖ If the random variable X has *Exponential distribution*, then

$$P(X > T + t \mid X > T) = P(X > t)$$

for all $T > 0$ and $t > 0$.

❖ Proof:

$$P(X > T + t \mid X > T) = \frac{P(X > T + t \cap X > T)}{P(X > T)} = \frac{P(X > T + t)}{P(X > T)}$$

$$= \frac{1 - F(T + t)}{1 - F(T)} = \frac{e^{-\lambda\,(T+t)}}{e^{-\lambda\,(T)}} = e^{-\lambda\,(t)} = P(X > t)$$

II. Exp. Dist. - Memoryless Property - Meaning

❖ Note that: If X is the lifetime of a component has *Exponential distribution*, then the probability that the component will last more than $T+t$ time's units given that it has already lasted more than T units is the same as that of a new component lasting more than t units.

❖ In other words, an old component which is still working is **just as reliable as** a new component.

Example 6

- Suppose a certain solid-state component has a lifetime time X with mean $\mathbf{100}$ hours. Find the probability that the component will last at least $\mathbf{80}$ hours given that it is already worked more than $\mathbf{30}$ hours.

Example 6 – Sol.

- Suppose a certain solid-state component has a lifetime time X with mean 100 hours. Find the probability that the component will last at least 80 hours given that it is already worked more than 30 hours.

$$\mu = 100 \quad \rightarrow \quad \lambda = 0.01$$

$$P(X > 80 \mid X > 30) = P(X > 50)$$

$$= 1 - F(50) = e^{-0.5} = 0.6065$$

Thank You ☺