STATISTICAL METHODS

Master in Industrial Management,

Operations and Sustainability (MIMOS)

2nd year/1st Semester

2025/2026

CONTACT

Professor: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

PROGRAM

1. Fundamental Concepts of Statistics
2. Descriptive Data Analysis
3. Introduction to Inferential Analysis
4. Parametric Hypothesis Testing
5. Non-Parametric Hypothesis Testing
6. Linear Regression Analysis

LECTURE 4 HOMEWORK: QUESTIONS AND SOLUTIONS

EXERCISE 3.25

3.25 The probability of A is 0.30, the probability of B is 0.40 and the probability of both is 0.30. What is the conditional probability of A given B? Are A and B independent in a probability sense?

Newbold et al (2013)

EXERCISE 3.25: SOLUTION

✅ Answer:

Given:

$$P(A) = 0.30, \quad P(B) = 0.40, \quad P(A \cap B) = 0.30$$

Step 1: Conditional probability $P(A \mid B)$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.30}{0.40} = 0.75$$

Step 2: Check for independence

- Two events A and B are **independent** if:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A) \cdot P(B) = 0.30 \cdot 0.40 = 0.12$$

- Given $P(A \cap B) = 0.30 \neq 0.12$, **A and B are not independent.**

EXERCISE 3.85

3.85 The Watts New Lightbulb Corporation ships large consignments of lightbulbs to big industrial users. When the production process is functioning correctly, which is 90% of the time, 10% of all bulbs produced are defective. However, the process is susceptible to an occasional malfunction, leading to a defective rate of 50%. If a defective bulb is found, what is the probability that the process is functioning correctly? If a nondefective bulb is found, what is the probability that the process is operating correctly?

Newbold et al (2013)

EXERCISE 3.83: SOLUTION

✅ Answer:

Given:

- Process correct: $P(C) = 0.90$
- Process malfunctioning: $P(M) = 0.10$
- Defective rate if correct: $P(D \mid C) = 0.10$
- Defective rate if malfunctioning: $P(D \mid M) = 0.50$

We want:

1. $P(C \mid D)$ → probability process is correct given defective bulb.
2. $P(C \mid D^c)$ → probability process is correct given nondefective bulb.

Step 1: Probability process correct given defective bulb

By Bayes' Theorem:

$$P(C \mid D) = \frac{P(C) \cdot P(D \mid C)}{P(C) \cdot P(D \mid C) + P(M) \cdot P(D \mid M)}$$

Substitute:

$$P(C \mid D) = \frac{0.90 \cdot 0.10}{0.90 \cdot 0.10 + 0.10 \cdot 0.50} = \frac{0.09}{0.09 + 0.05} = \frac{0.09}{0.14} \approx 0.643$$

✅ If a bulb is defective, there's about **64.3% probability** the process is still correct.

EXERCISE 3.83: SOLUTION

✅ Answer:

Step 2: Probability process correct given nondefective bulb

Now use Bayes' again:

$$P(C \mid D^c) = \frac{P(C) \cdot P(D^c \mid C)}{P(C) \cdot P(D^c \mid C) + P(M) \cdot P(D^c \mid M)}$$

- $P(D^c \mid C) = 1 - 0.10 = 0.90$
- $P(D^c \mid M) = 1 - 0.50 = 0.50$

Substitute:

$$P(C \mid D^c) = \frac{0.90 \cdot 0.90}{0.90 \cdot 0.90 + 0.10 \cdot 0.50} = \frac{0.81}{0.81 + 0.05} = \frac{0.81}{0.86} \approx 0.942$$

✅ If a bulb is nondefective, there's about **94.2% probability** the process is correct.

📌 **Final Answer:**

- $P(C \mid D) \approx 0.643$
- $P(C \mid D^c) \approx 0.942$

EXERCISE 4.18

4.18 An automobile dealer calculates the proportion of new cars sold that have been returned a various numbers of times for the correction of defects during the warranty period. The results are shown in the following table.

Number of returns	0	1	2	3	4
Proportion	0.28	0.36	0.23	0.09	0.04

a. Graph the probability distribution function.
b. Calculate and graph the cumulative probability distribution.
c. Find the mean of the number of returns of an automobile for corrections for defects during the warranty period.
d. Find the variance of the number of returns of an automobile for corrections for defects during the warranty period.

Newbold et al (2013)

EXERCISE 4.18 A): SOLUTION

✅ Answer:

We are given the **number of returns** (X) and their **proportions**:

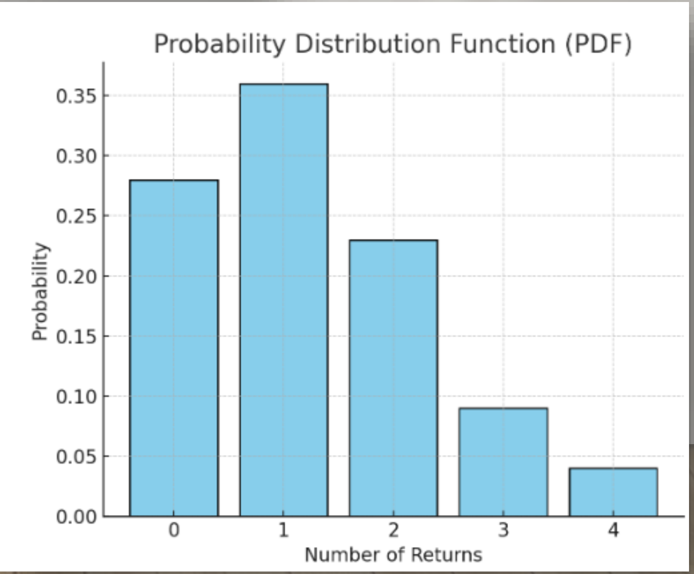
Number of returns x	**0**	**1**	**2**	**3**	**4**
Proportion $P(X = x)$	0.28	0.36	0.23	0.09	0.04

a. Graph the probability distribution function (PDF)

The **probability distribution function** is simply the probability of each value of X.

- On the x-axis: Number of returns (0,1,2,3,4)
- On the y-axis: Probability (0.28, 0.36, 0.23, 0.09, 0.04)

A simple **bar graph** works well, with **height of each bar = P(X=x)**.



Probability Mass Function (PMF) or Probability Distributionn Function (PDF) – shows the probability for each number of returns.

EXERCISE 4.18 B): SOLUTION

Answer:

We are given the **number of returns** (X) and their **proportions**:

Number of returns x	**0**	**1**	**2**	**3**	**4**
Proportion $P(X = x)$	0.28	0.36	0.23	0.09	0.04

b. Cumulative probability distribution

The **cumulative distribution function (CDF)** is:

$$F(x) = P(X \leq x)$$

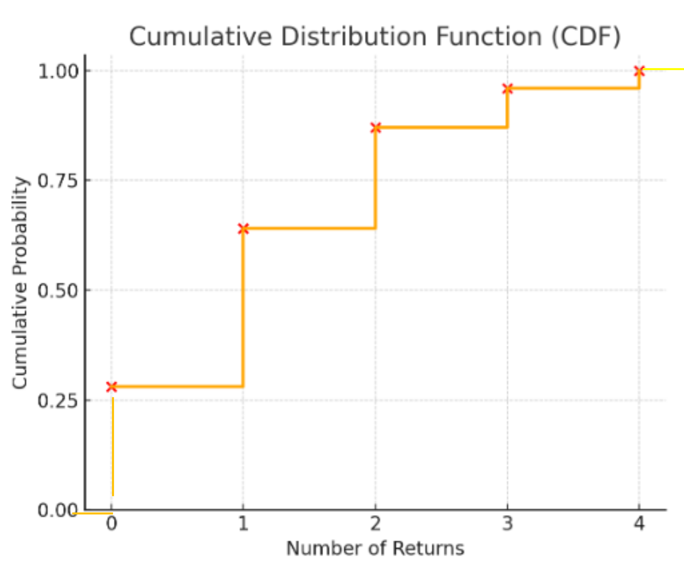
Compute step by step:

$$F(0) = P(X \leq 0) = 0.28$$
$$F(1) = P(X \leq 1) = 0.28 + 0.36 = 0.64$$
$$F(2) = P(X \leq 2) = 0.64 + 0.23 = 0.87$$
$$F(3) = P(X \leq 3) = 0.87 + 0.09 = 0.96$$
$$F(4) = P(X \leq 4) = 0.96 + 0.04 = 1.00$$

CDF Table:

x	**0**	**1**	**2**	**3**	**4**
F(x)	0.28	0.64	0.87	0.96	1.00

- Graph: Step function increasing from left to right, ending at 1.



Cumulative Distribution Function (CDF) – shows the cumulative probability as the number of returns increases.

EXERCISE 4.18 C): SOLUTION

✅ Answer:

We are given the **number of returns** (X) and their **proportions**:

Number of returns x	0	1	2	3	4
Proportion $P(X = x)$	0.28	0.36	0.23	0.09	0.04

c. Mean of the number of returns

The **mean** (expected value) is:

$$\mu = E[X] = \sum x \cdot P(X = x)$$

$$\mu = 0 \cdot 0.28 + 1 \cdot 0.36 + 2 \cdot 0.23 + 3 \cdot 0.09 + 4 \cdot 0.04$$

Step by step:

$$0 + 0.36 + 0.46 + 0.27 + 0.16 = 1.25$$

✅ **Mean = 1.25 returns**

EXERCISE 4.18 D): SOLUTION

✅ Answer:

We are given the **number of returns** (X) and their **proportions**:

Number of returns x	0	1	2	3	4
Proportion $P(X = x)$	0.28	0.36	0.23	0.09	0.04

d. Variance of the number of returns

The **variance** is:

$$\sigma^2 = E[(X-\mu)^2] = \sum (x-\mu)^2 \cdot P(X=x)$$

Compute each term:

$$\begin{aligned}
(0-1.25)^2 \cdot 0.28 &= 1.5625 \cdot 0.28 \approx 0.4375 \\
(1-1.25)^2 \cdot 0.36 &= 0.0625 \cdot 0.36 \approx 0.0225 \\
(2-1.25)^2 \cdot 0.23 &= 0.5625 \cdot 0.23 \approx 0.1294 \\
(3-1.25)^2 \cdot 0.09 &= 3.0625 \cdot 0.09 \approx 0.2756 \\
(4-1.25)^2 \cdot 0.04 &= 7.5625 \cdot 0.04 \approx 0.3025
\end{aligned}$$

Sum:

$$0.4375 + 0.0225 + 0.1294 + 0.2756 + 0.3025 \approx 1.1675$$

✅ **Variance ≈ 1.17**

- Standard deviation = $\sigma = \sqrt{1.17} \approx 1.08$

EXERCISE 4.33

4.33 For a binomial probability distribution with $P = 0.4$ and $n = 20$, find the probability that the number of successes is equal to 9 and the probability that the number of successes is fewer than 7.

Newbold et al (2013)

EXERCISE 4.33: SOLUTION

Answer:

Table 2 Probability Function of the Binomial Distribution

n	x	p = .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1419	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1442
	9	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.0365	.0148
	6	.0003	.0089	.0454	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.0746	.0370
	7	.0000	.0020	.0160	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602

Approximate value

- Binomial distribution: $X \sim \text{Binomial}(n = 20, p = 0.4)$
- Question 1: $P(X = 9)$
- Question 2: $P(X < 7) = P(X \leq 6)$

1 Probability that $X = 9$

Formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Substitute $n = 20, k = 9, p = 0.4$:

$$P(X = 9) = \binom{20}{9} (0.4)^9 (0.6)^{11}$$

- Step 1: Compute the combination:

$$\binom{20}{9} = \frac{20!}{9! \cdot 11!} = 167960$$

- Step 2: Compute powers:

$$(0.4)^9 \approx 0.000262144, \quad (0.6)^{11} \approx 0.003517$$

- Step 3: Multiply:

$$P(X = 9) \approx 167960 \cdot 0.000262144 \cdot 0.003517 \approx 0.155$$

✅ So, $P(X = 9) \approx 0.155$

Exact value

EXERCISE 4.33: SOLUTION

✅ Answer:

Table 3 Cumulative Binomial Probabilities (*Continued*)

n	x	p .05	.10	.15	.20	.25	.30	.35	.40	.45	.500
	12	1.00	1.00	1.00	1.00	1.00	.999	.997	.988	.966	.916
	13	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.989	.968
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.990
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	[illegible]
	16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	[illegible]
20	0	.358	.122	.039	.012	.003	.001	.00	.00	.00	[illegible]
	1	.736	.392	.176	.069	.024	.008	.002	.001	.00	[illegible]
	2	.925	.677	.405	.206	.091	.035	.012	.004	.001	[illegible]
	3	.984	.867	.648	[illegible]	[illegible]	[illegible]	[illegible]	.016	.005	[illegible]
	4	.997	.957	.83	[illegible]	[illegible]	[illegible]	[illegible]	.051	.019	[illegible]
	5	1.00	.989	.933	.804	.617	.416	.245	.126	.055	[illegible]
	6	1.00	.998	.978	.913	.786	.608	.417	.25	.13	[illegible]
	7	1.00	1.00	.994	.968	.898	.772	.601	.416	.252	[illegible]
	8	1.00	1.00	.999	.99	.959	.887	.762	.596	.414	[illegible]
	9	1.00	1.00	1.00	.997	.986	.952	.878	.755	.591	[illegible]
	10	1.00	1.00	1.00	.999	.996	.983	.947	.872	.751	[illegible]
	11	1.00	1.00	1.00	1.00	.999	.995	.98	.943	.869	[illegible]
	12	1.00	1.00	1.00	1.00	1.00	.999	.994	.979	.942	[illegible]
	13	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.979	[illegible]
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	[illegible]
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994
	16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
	17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000

Approximate value

- Binomial distribution: $X \sim \text{Binomial}(n = 20, p = 0.4)$
- Question 1: $P(X = 9)$
- Question 2: $P(X < 7) = P(X \leq 6)$

2 Probability that $X < 7$ (i.e., $X \leq 6$)

$$P(X < 7) = \sum_{k=0}^{6} \binom{20}{k} (0.4)^k (0.6)^{20-k}$$

- This is a sum of 7 terms ($k = 0, 1, \ldots, 6$)
- Usually we use **software or a calculator** to compute.

Approximate value using software:

$$P(X \leq 6) \approx 0.296$$

Exact value

EXERCISE 4.46

4.46 A campus finance officer finds that, for all parking tickets issued, fines are paid for 78% of the tickets. The fine is $2. In the most recent week, 620 parking tickets have been issued.

a. Find the mean and standard deviation of the number of these tickets for which the fines will be paid.
b. Find the mean and standard deviation of the amount of money that will be obtained from the payment of these fines.

Newbold et al (2013)

EXERCISE 4.46 A): SOLUTION

✅ Answer:

Given Data

- Probability that a ticket is paid: $p = 0.78$
- Probability not paid: $q = 1 - p = 0.22$
- Number of tickets: $n = 620$
- Fine per ticket: $2

Let X = number of tickets paid. Then $X \sim \text{Binomial}(n = 620, p = 0.78)$.

Part (a): Mean and standard deviation of tickets paid

Mean of a binomial distribution:

$$\mu_X = n \cdot p = 620 \cdot 0.78$$

$$\mu_X = 483.6$$

Standard deviation of a binomial distribution:

$$\sigma_X = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{620 \cdot 0.78 \cdot 0.22}$$

$$620 \cdot 0.78 = 483.6$$

$$483.6 \cdot 0.22 \approx 106.392$$

$$\sigma_X = \sqrt{106.392} \approx 10.32$$

✅ So, the **mean number of tickets paid** ≈ 483.6, and the **standard deviation** ≈ 10.32.

EXERCISE 4.46 B): SOLUTION

✅ Answer:

Given Data

- Probability that a ticket is paid: $p = 0.78$
- Probability not paid: $q = 1 - p = 0.22$
- Number of tickets: $n = 620$
- Fine per ticket: $2

Let X = number of tickets paid. Then $X \sim \text{Binomial}(n = 620, p = 0.78)$.

Part (b): Mean and standard deviation of the total money

Let $Y = 2 \cdot X$ = total money collected.

For a linear transformation $Y = aX$:

$$\mu_Y = a \cdot \mu_X$$

$$\sigma_Y = |a| \cdot \sigma_X$$

Here, $a = 2$ dollars per ticket:

$$\mu_Y = 2 \cdot 483.6 = 967.2$$

$$\sigma_Y = 2 \cdot 10.32 \approx 20.64$$

✅ So, the **mean amount of money** ≈ $967.20, and the **standard deviation** ≈ $20.64.

This topic has not been covered in class yet, but it will be addressed later.

LECTURE 5: POISSON DISTRIBUTION

POISSON DISTRIBUTION

- The Poisson distribution is used to determine the probability of a random variable which characterizes the number of occurrences or successes of a certain event in a given continuous interval (such as time, surface area, or length).

Newbold et al (2013)

Poisson distribution is useful for modeling the number of events in a fixed interval of time or space when events occur independently and with a constant average rate λ.

Let X be the number of events in a given interval, $x = 0, 1, 2, \dots$

The random variable X follows a Poisson distribution with parameter λ : $\boldsymbol{X \sim}$ **Poisson**$(\boldsymbol{\lambda})$

APPLICATIONS OF THE POISSON DISTRIBUTION

Content (bullet points):

- **Traffic Accidents:** Number of accidents on a highway in a given time period.
- **Call Centers:** Number of calls received per hour.
- **Insurance Claims:** Number of claims filed in a given period.
- **Machine Failures:** Number of breakdowns in a factory per day.
- **Hospital Emergencies:** Number of patients arriving at the ER per hour.
- **Rare Events:** Occurrence of mutations in a DNA sequence, or defects in manufactured items.

POISSON DISTRIBUTION ASSUMPTIONS

- Assume an interval is divided into a very large number of equal subintervals where the probability of the occurrence of an event in any subinterval is very small.

Poisson distribution assumptions

1. The probability of the occurrence of an event is constant for all subintervals.
2. There can be no more than one occurrence in each subinterval.
3. Occurrences are independent; that is, an occurrence in one interval does not influence the probability of an occurrence in another interval.

Newbold et al (2013)

PMF OF A POISSON DISTRIBUTION

For a discrete distribution like Poisson, PDF = PMF.

The expected number of events per unit is the parameter λ (lambda), which is a constant that specifies the average number of occurrences (successes) for a particular time and/or space

Probability Mass Function (PMF)

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

X ~ Poisson(λ)

where:

$P(x)$ = the probability of x successes over a given time or space, given λ

λ = the expected number of successes per time or space unit, $\lambda > 0$

e = base of the natural logarithm system $(2.71828\ldots)$

"$e \approx 2.718$, Euler's number"

Newbold et al (2013)

MEAN AND VARIANCE OF A POISSON RANDOM VARIABLE

Mean and variance of the Poisson distribution

- Mean

$$\boxed{\mu_x = E[X] = \lambda}$$

- Variance and Standard Deviation

$$\boxed{\sigma_x^2 = E\left[(X-\mu_x)^2\right] = \lambda}$$

$$\boxed{\sigma = \sqrt{\lambda}}$$

where $\lambda =$ expected number of successes per time or space unit

Newbold et al (2013)

USING POISSON TABLES

X	λ 0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	**0.0758**	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

X ~ Poisson(0.5)
λ = 0.5
x = 2

Example: Find $P(X=2)$ if $\lambda = .50$

$$P(X=2) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{e^{-0.50}(0.50)^2}{2!} = .0758$$

Note: Exact probabilities from the Poisson PMF can be computed manually or obtained from Poisson probability tables.

Newbold et al (2013)

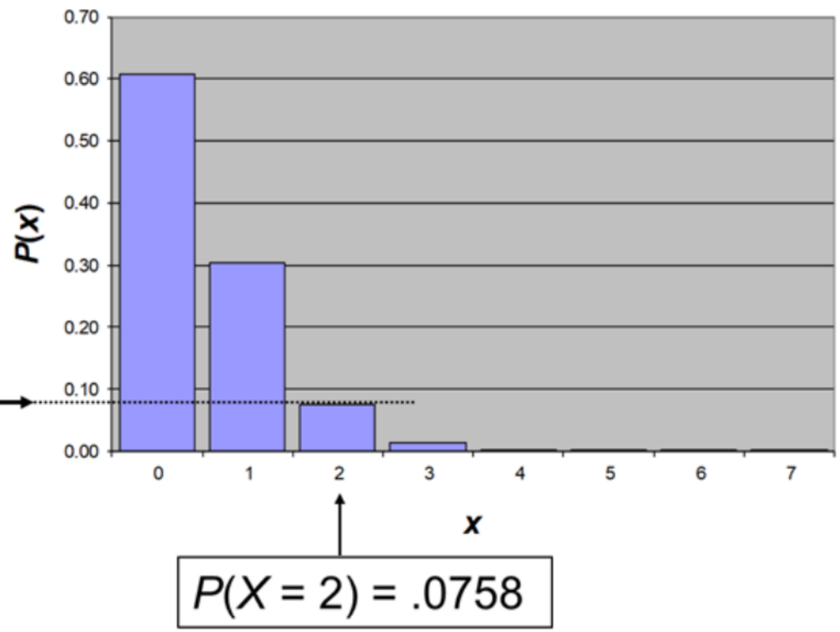
POISSON PMF GRAPH

Graphically:

X **~ Poisson(0.5)**
λ = 0.5
x = 2

$\lambda = .50$

x	$\lambda = 0.50$
0	0.6065
1	0.3033
2	0.0758
3	0.0126
4	0.0016
5	0.0002
6	0.0000
7	0.0000



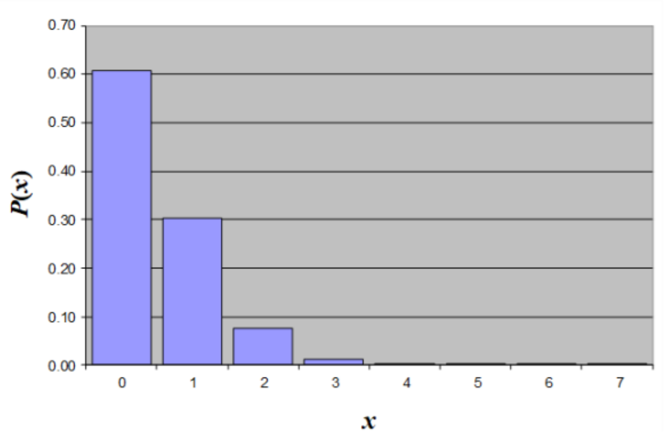
Newbold et al (2013)

POISSON DISTRIBUTION SHAPE

- The shape of the Poisson Distribution depends on the parameter λ :

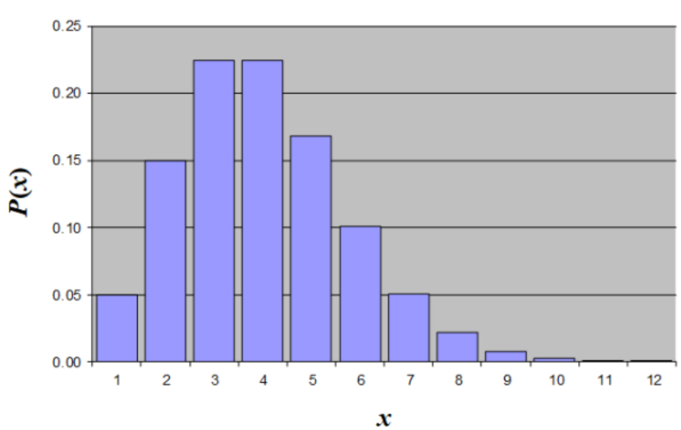
X ~ Poisson(0.50)

$$\lambda = 0.50$$



X ~ Poisson(3)

$$\lambda = 3.00$$



Newbold et al (2013)

EXERCISE 4.55

4.55 The number of accidents in a production facility has a Poisson distribution with a mean of 2.6 per month.

a. For a given month what is the probability there will be fewer than 2 accidents?
b. For a given month what is the probability there will be more than 3 accidents?

Newbold et al (2013)

EXERCISE 4.55 A): SOLUTION

Answer:

Table 6 Cumulative Poisson Probabilities

	Mean Arrival Rat[illegible]								
	2.1	2.2	2.3	2.4	2.5	2.6	2.[illegible]			
0	.1225	.1108	.1003	.0907	.0821	.0743	.0[illegible]			
1	.3796	.3546	.3309	.3084	.2873	.2674	.2[illegible]			
2	.6496	.6227	.5960	.5697	.5438	.5184	.4936	.4695	.4460	.4232
3	.8386	.8194	.79[illegible]	[illegible]	[illegible]	[illegible]	.7141	.6919	.6696	.6472
4	.9379	.9275	.91[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	.9796	.9751	.9700	.9643	.9580	.9510				
6	.9941	.9925	.9906	.9884	.9858	.9828				
7	.9985	.9980	.9974	.9967	.9958	.9947				
8	.9997	.9995	.9994	.9991	.9989	.9985				
9	.9999	.9999	.9999	.9998	.9997	.9996				
10	1.0000	1.0000	1.0000	1.0000	.9999	.9999				
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				

Approximate value

We are given:

$$X \sim Poisson(\lambda = 2.6)$$

The PMF of a Poisson variable is:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

a) Probability of fewer than 2 accidents

"Fewer than 2" means $X < 2 \Rightarrow X = 0$ or $X = 1$.

$$P(X < 2) = P(X = 0) + P(X = 1)$$

1. $P(X = 0) = \frac{e^{-2.6}2.6^0}{0!} = e^{-2.6} \approx 0.0743$
2. $P(X = 1) = \frac{e^{-2.6}2.6^1}{1!} = 2.6 \cdot e^{-2.6} \approx 0.1931$

Exact value

$$P(X < 2) \approx 0.0743 + 0.1931 = 0.2674$$

✅ So the probability of fewer than 2 accidents in a month is approximately **0.267**.

EXERCISE 4.55 B): SOLUTION

Answer:

Table 6 Cumulative Poisson Probabilities

	Mean Arrival Rate									
	2.1	2.2	2.3	2.4	2.5	2.6				
0	.1225	.1108	.1003	.0907	.0821	.0743				
1	.3796	.3546	.3309	.3084	.2873	.2674				
2	.6496	.6227	.5960	.5697	.5438	.5184				
3	.8386	.8194	.7993	.7787	.7576	.7360	.7141	.6919	.6696	.6472
4	.9379	.9275				.8774	.8629	.8477	.8318	.8153
5	.9796	.9751								
6	.9941	.9925	.9906	.9884	.9858					
7	.9985	.9980	.9974	.9967	.9958					
8	.9997	.9995	.9994	.9991	.9989					
9	.9999	.9999	.9999	.9998	.9997					
10	1.0000	1.0000	1.0000	1.0000	.9999					
11	1.0000	1.0000	1.0000	1.0000	1.0000					
12	1.0000	1.0000	1.0000	1.0000	1.0000					

Approximate value

We are given:

$$X \sim Poisson(\lambda = 2.6)$$

The PMF of a Poisson variable is:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

b) Probability of more than 3 accidents

"More than 3" means $X > 3 \Rightarrow 1 - P(X \leq 3)$.

$$P(X > 3) = 1 - P(X \leq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

We already have $P(X = 0) \approx 0.0743$, $P(X = 1) \approx 0.1931$.

3. $P(X = 2) = \frac{e^{-2.6}2.6^2}{2!} = \frac{6.76}{2}e^{-2.6} \approx 0.2511$
4. $P(X = 3) = \frac{e^{-2.6}2.6^3}{3!} = \frac{17.576}{6}e^{-2.6} \approx 0.2177$

Exact value

$$P(X \leq 3) \approx 0.0743 + 0.1931 + 0.2511 + 0.2177 = 0.7362$$

$$P(X > 3) = 1 - 0.7362 \approx 0.2638$$

So the probability of more than 3 accidents in a month is approximately **0.264**.

POISSON APPROXIMATION TO THE BINOMIAL DISTRIBUTION

Let *X* be the number of successes from *n* independent trials, each with probability of success *P*. The distribution of the number of successes, *X* , is binomial, with mean nP.
If the number of trials, *n*, is large and nP is of only moderate size (preferably $nP \le 7$), this distribution can be approximated by the Poisson distribution with $\lambda = nP$. The probability distribution of the approximating distribution is

$$\boxed{P(x) = \frac{e^{-nP}(nP)^x}{x!}} \quad \text{for} \quad x = 0, 1, 2, \ldots$$

Newbold et al (2013)

If **X ~ Binomial(n, p)** with large n and small p (preferably **n × p < 7**), then X can be approximated by a Poisson random variable with parameter **λ = n × p**. In other words, **X ~ Poisson(n × p).**

SUMMARY: CHARACTERISTICS OF A POISSON DISTRIBUTION

- **Discrete distribution** – takes non-negative integer values (0, 1, 2, ...).
- **Parameter** – $\lambda > 0$ (the average number of events in a fixed interval).
- **Mean and variance** – both equal to λ.
- **Probability mass function (PMF):**

$$\boxed{P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots}$$

- **Independent events** – the number of events in disjoint intervals are independent.
- **Skewness** – positively skewed (tail to the right), especially for small λ; becomes more symmetric as λ increases.
- **Sum property** – the sum of independent Poisson variables is also Poisson: $X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$.

***X* ~ Poisson(λ)**

- Possible values x = 0, 1, 2, ….
- E(X) = Var(X) = λ
- Independent events
- Positively Skewed

If **X1 ~ Poisson(λ1)** and **X2 ~ Poisson(λ2)** are two independent random variables, then **X1 + X2 ~ Poisson (λ1+ λ2).**

Example:

- "If $X \sim Poisson(\lambda)$ for 1 hour, then $Y \sim Poisson(2 \times \lambda)$ for 2 hours."

EXERCISE 4.60

4.60 An insurance company holds fraud insurance policies on 6,000 firms. In any given year the probability that any single policy will result in a claim is 0.001. Find the probability that at least 3 claims are made in a given year. Use the Poisson approximation to the binomial distribution.

Newbold et al (2013)

EXERCISE 4.60: SOLUTION

✅ Answer: "X is the number of complaints in 6,000 firms.

Given:

- Number of policies: $n = 6000$
- Probability of a claim for a single policy: $p = 0.001$
- We want: $P(X \geq 3)$, where X = number of claims in a year.

Step 1: Check conditions for Poisson approximation

Poisson approximation to the binomial works well when:

- n is large → $n = 6000$ ✅
- p is small → $p = 0.001$ ✅
- $\lambda = n \cdot p = 6000 \cdot 0.001 = 6$

So we can use:

$$X \sim Poisson(\lambda = 6)$$

EXERCISE 4.60: SOLUTION

✅ Answer:

Step 2: Express probability

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

Step 3: Compute Poisson probabilities

Poisson PMF:

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

1. $P(X = 0) = e^{-6}\frac{6^0}{0!} = e^{-6} \approx 0.0024788$
2. $P(X = 1) = e^{-6}\frac{6^1}{1!} = 6 \cdot e^{-6} \approx 0.014873$
3. $P(X = 2) = e^{-6}\frac{6^2}{2!} = 18 \cdot e^{-6} \approx 0.044618$

$$P(X \leq 2) = 0.0024788 + 0.014873 + 0.044618 \approx 0.0619698$$

Step 4: Compute $P(X \geq 3)$

$$P(X \geq 3) = 1 - 0.06197 \approx 0.938$$

✅ **Answer:** The probability that at least 3 claims are made in a given year is approximately **0.938**.

LECTURE 5: CONTINUOUS RANDOM VARIABLES

CONTINUOUS RANDOM VARIABLES

- A continuous random variable is a variable that can assume any value in an interval
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height, in inches
- These can potentially take on any value, depending only on the ability to measure accurately.

Newbold et al (2013)

CUMULATIVE DISTRIBUTION FUNCTION (CDF)

- The cumulative distribution function, $F(x)$, for a continuous random variable *X* expresses the probability that *X* does not exceed the value of *x*

Continuous CDF

$$F(x) = P(X \leq x)$$

- Let *a* and *b* be two possible values of *X*, with *a* < *b*. The probability that *X* lies between *a* and *b* is

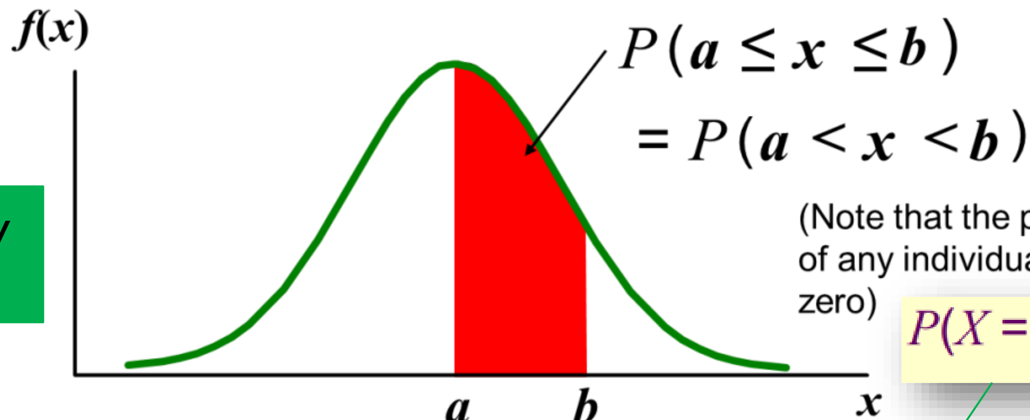
$$P(a < X < b) = F(b) - F(a)$$

For continuous random variables, $P(a < X < b)$ is the area under the Probability Density Function (PDF) curve between a and b.

Newbold et al (2013)

PROBABILITY AS AN AREA

Shaded area under the curve is the probability that *X* is between *a* and *b*



Probability Density Function (PDF)

Possible Values of the Variable

(Note that the probability of any individual value is zero) $P(X = x_0) = 0,$

Newbold et al (2013)

Since the probability that a continuous random variable takes any exact value is zero, we have: P(a < X < b) = P(a ≤ X ≤ b) = P(a < X ≤ b) = P(a ≤ X < b).

CDF: DISCRETE VS CONTINUOUS RANDOM VARIABLES

1. Discrete Random Variable

$$\boxed{F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} P(X = x)}$$

- **CDF**: sum of probabilities of all values ≤ x_0
- **PMF**: $P(X = x)$, probability of each specific value
- **Explanation:** "The CDF at x_0 gives the probability that X takes a value ≤ x_0."

2. Continuous Random Variable

$$\boxed{F(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)\, dx}$$

- **CDF**: integral of the PDF up to x_0
- **PDF**: $f(x)$, probability density function, does not give the probability of an exact value
- **Explanation:** "The CDF at x_0 gives the probability that X takes a value ≤ x_0, obtained by integrating the PDF up to x_0."

CDF: DISCRETE VS CONTINUOUS RANDOM VARIABLES

Summary Table

Aspect	Discrete	Continuous
CDF	Sum of PMF up to x_0	Integral of PDF up to x_0
PMF/PDF	PMF $P(X = x)$	PDF $f(x)$
Probability of exact value	Can be >0	Always 0
Visual	Step function	Smooth curve

PROPERTIES OF THE PROBABILITY DENSITY FUNCTION (PDF)

The probability density function, $f(x)$, of random variable *X* has the following properties:

1. $f(x) > 0$ for all values of x
2. The area under the probability density function $f(x)$ over all values of the random variable *X* within its range, is equal to 1.0
3. The probability that *X* lies between two values is the area under the density function graph between the two values

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

(a) $f(x) \geq 0$

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

Newbold et al (2013)

PROPERTIES OF THE PROBABILITY DENSITY FUNCTION (PDF)

The probability density function, $f(x)$, of random variable X has the following properties:

4. The cumulative density function $F(x_0)$, is the area under the probability density function $f(x)$ from the minimum x value up to x_0

Continuous CDF

$$F(x_0) = \int_{x_m}^{x_0} f(x)\,dx$$

Newbold et al (2013)

where x_m is the minimum value of the random variable x

Note that the **Probability Density Function** (PDF) is used only for continuous variables, and the **Probability Mass Function** (PMF) is used only for discrete variables. In some contexts, however, the acronym PDF is also used for discrete variables, with the meaning of **Probability Distribution Function**. **Cumulative Distribution Function** (CDF) applies to both discrete and continuous variables.

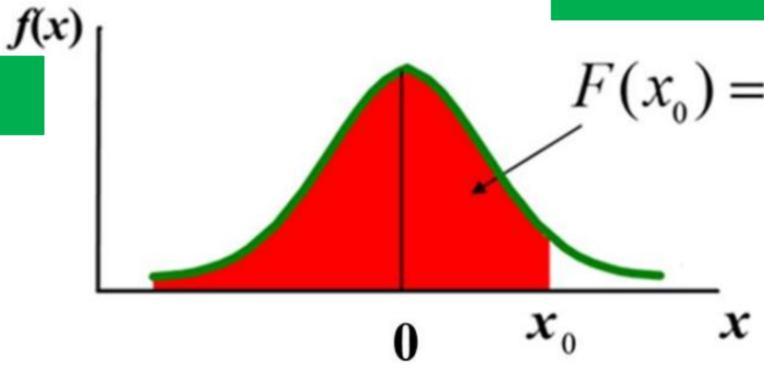
CDF: VISUAL REPRESENTATION

1. The total area under the curve $f(x)$ is 1
2. The area under the curve $f(x)$ to the left of x_0 is $F(x_0)$, where x_0 is any value that the random variable can take.

Continuous CDF

$$F(x_0) = P(X \le x_0)$$

Continuous PDF



Possible Values of the Variable

Newbold et al (2013)

EXPECTATIONS FOR CONTINUOUS RANDOM VARIABLES

- The mean of *X*, denoted μ_X, is defined as the expected value of *X*

$$\boxed{\mu_X = E[X]}$$

- The variance of *X*, denoted σ_X^2, is defined as the expectation of the squared deviation, $(X - \mu_X)^2$, of a random variable from its mean

$$\boxed{\sigma_X^2 = E\left[(X - \mu_X)^2\right]}$$

Newbold et al (2013)

MEAN, VARIANCE, AND STANDARD DEVIATION OF A CONTINUOUS VARIABLE

1. Mean (Expected Value)

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

2. Variance

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

König's Theorem

$$\mathrm{Var}(X) = \sigma^2 = E(X^2) - [E(X)]^2$$

3. Standard Deviation

$$\sigma = \sqrt{Var(X)} = \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx}$$

König's Theorem is also applied when calculating the variance of continuous random variables.

LINEAR FUNCTIONS OF RANDOM VARIABLES

- Let $W = a + bX$, where X has mean μ_X and variance σ_X^2, and a and b are constants
- Then the mean of W is

$$\boxed{\mu_W = E[a + bX] = a + b\mu_X}$$

- the variance is

$$\boxed{\sigma_W^2 = Var[a + bX] = b^2\sigma_X^2}$$

- the standard deviation of W is

$$\boxed{\sigma_W = |b|\sigma_X}$$

Newbold et al (2013)

LINEAR FUNCTIONS OF RANDOM VARIABLES

- An important special case of the result for the linear function of random variable is the standardized random variable

$$\boxed{Z = \frac{X - \mu_X}{\sigma_X}}$$

- which has a mean 0 and variance 1

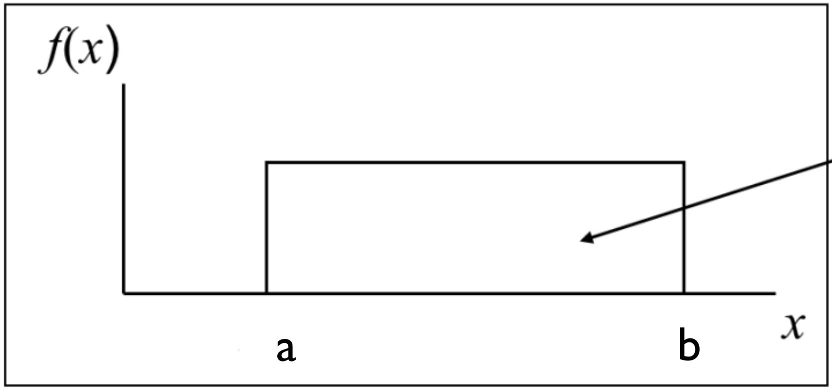
Newbold et al (2013)

LECTURE 5: UNIFORM DISTRIBUTION

UNIFORM DISTRIBUTION

- The uniform distribution is a probability distribution that has equal probabilities for all equal-width intervals within the range of the random variable

The random variable X follows a Uniform distribution with parameters a and b (a < b): ***X* ~ Uniform(a, b)**.



Total area under the uniform probability density function is 1.0

The uniform random variable takes values in the interval $[a, b]$.

Newbold et al (2013)

PDF OF A UNIFORM DISTRIBUTION

Probability Density Function (PDF)

The Continuous Uniform Distribution:

$X \sim$ Uniform(a, b)

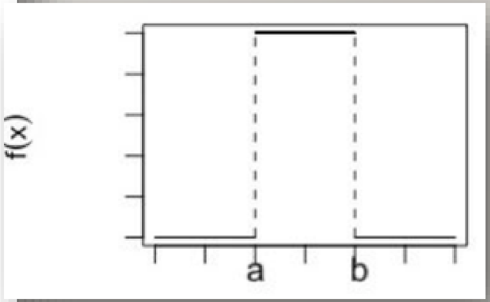
$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

where

$f(x)$ = value of the density function at any *x* value

a = minimum value of *x*

b = maximum value of *x*



Note: This graph shows the PDF of a Uniform Distribution.

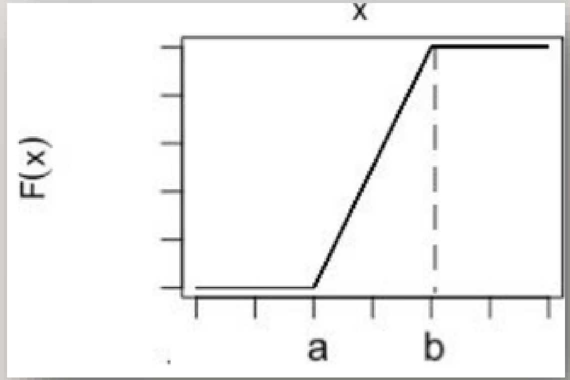
Newbold et al (2013)

CDF OF A UNIFORM DISTRIBUTION

X ~ Uniform(a, b)

Cumulative Distribution Function (CDF)

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b \end{cases}$$



Note: This graph shows the CDF of a Uniform Distribution.

MEAN AND VARIANCE OF THE UNIFORM DISTRIBUTION

- The mean of a uniform distribution is

$$\boxed{\mu = \frac{a+b}{2}}$$

- The variance is

$$\boxed{\sigma^2 = \frac{(b-a)^2}{12}}$$

Where a = minimum value of x

b = maximum value of x

Newbold et al (2013)

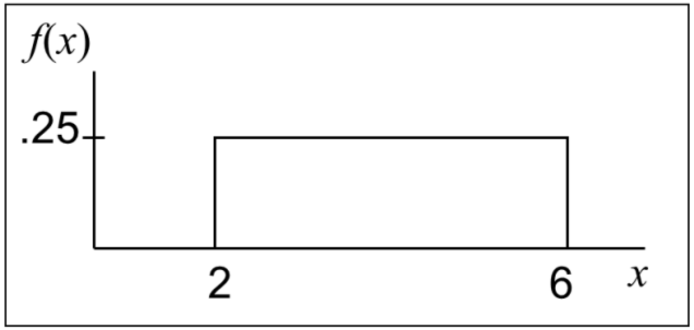
UNIFORM DISTRIBUTION EXAMPLE

Example: Uniform probability distribution over the range $2 \le x \le 6$:

X ~ Uniform(2, 6)

PDF

$$f(x) = \frac{1}{6-2} = .25 \text{ for } 2 \le x \le 6$$



$$\mu = \frac{a+b}{2} = \frac{2+6}{2} = 4$$

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(6-2)^2}{12} = 1.333$$

Newbold et al (2013)

PDF AND CDF OF THE UNIFORM(0,1) DISTRIBUTION – A SPECIAL CASE

$X \sim$ Uniform(0, 1)

PDF

$$f(x) = \begin{cases} 1 & (0 < x < 1) \\ 0 & (\text{outros } x) \end{cases}$$

CDF

$$F(x) = \begin{cases} 0 & (x < 0) \\ x & (0 \le x < 1) \\ 1 & (x \ge 1) \end{cases}$$

Mean and Variance

$$E(X) = \frac{1}{2}, \quad \mathrm{Var}(X) = \frac{1}{12}$$

EXERCISE 5.5

5.5 An analyst has available two forecasts, F_1 and F_2, of earnings per share of a corporation next year. He intends to form a compromise forecast as a weighted average of the two individual forecasts. In forming the compromise forecast, weight X will be given to the first forecast and weight $(1 - X)$, to the second, so that the compromise forecast is $XF_1 + (1 - X)F_2$. The analyst wants to choose a value between 0 and 1 for the weight X, but he is quite uncertain of what will be the best choice. Suppose that what eventually emerges as the best possible choice of the weight X can be viewed as a random variable uniformly distributed between 0 and 1, having the probability density function

$$f(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{for all other } x \end{cases}$$

a. Graph the probability density function.
b. Find and graph the cumulative distribution function.
c. Find the probability that the best choice of the weight X is less than 0.25.
d. Find the probability that the best choice of the weight X is more than 0.75.
e. Find the probability that the best choice of the weight X is between 0.2 and 0.8.

Newbold et al (2013)

EXERCISE 5.5 A): SOLUTION

Answer:

We are given:

- X = best possible weight for the compromise forecast.
- $X \sim \text{Uniform}(0,1)$
- PDF:

PDF

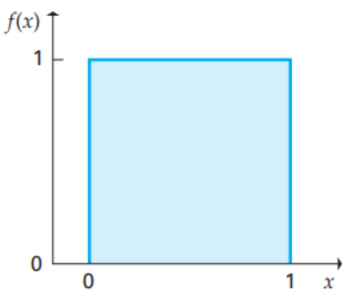
$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The **Uniform(0,1)** distribution is a special case of the **Uniform(a,b)** distribution with a = 0 and b = 1.

a) Graph the probability density function (PDF)

- For a **Uniform(0,1)** variable, the PDF is a **flat line at 1** between 0 and 1.
- Outside $[0,1]$, the PDF is 0.

(Slide tip: draw a rectangle from x = 0 to x = 1 with height = 1.)



(a)

Mean and Variance

$$E(X) = \frac{1}{2}, \quad \text{Var}(X) = \frac{1}{12}$$

The mean and variance of a Uniform(0,1) random variable are 1/2 and 1/12, respectively.

EXERCISE 5.5 B): SOLUTION

Answer:

We are given:

- X = best possible weight for the compromise forecast.
- $X \sim \text{Uniform}(0, 1)$
- PDF:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

b) Cumulative distribution function (CDF)

The CDF of a uniform variable:

CDF
$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

- Graph: straight line from (0,0) to (1,1).

EXERCISE 5.5 C): SOLUTION

Answer:

We are given:

- X = best possible weight for the compromise forecast.
- $X \sim \text{Uniform}(0, 1)$
- PDF:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

c) Probability $X < 0.25$

$$P(X < 0.25) = F(0.25) = 0.25$$

EXERCISE 5.5 D): SOLUTION

✅ Answer:

We are given:

- X = best possible weight for the compromise forecast.
- $X \sim \text{Uniform}(0, 1)$
- PDF:

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

d) Probability $X > 0.75$

$$P(X > 0.75) = 1 - F(0.75) = 1 - 0.75 = 0.25$$

EXERCISE 5.5 E): SOLUTION

Answer:

We are given:

- X = best possible weight for the compromise forecast.
- $X \sim \text{Uniform}(0, 1)$
- PDF:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

e) Probability $0.2 \leq X \leq 0.8$

$$P(0.2 \leq X \leq 0.8) = F(0.8) - F(0.2) = 0.8 - 0.2 = 0.6$$

HOMEWORK OF LECTURE 5: QUESTIONS

EXERCISE 4.57

4.57 Records indicate that, on average, 3.2 breakdowns per day occur on an urban highway during the morning rush hour. Assume that the distribution is Poisson.

a. Find the probability that on any given day there will be fewer than 2 breakdowns on this highway during the morning rush hour.
b. Find the probability that on any given day there will be more than 4 breakdowns on this highway during the morning rush hour.

Newbold et al (2013)

EXERCISE 5.6

5.6 The jurisdiction of a rescue team includes emergencies occurring on a stretch of river that is 4 miles long. Experience has shown that the distance along this stretch, measured in miles from its northernmost point, at which an emergency occurs can be represented by a uniformly distributed random variable over the range 0 to 4 miles. Then, if X denotes the distance (in miles) of an emergency from the northernmost point of this stretch of river, its probability density function is as follows:

$$f(x) = \begin{cases} 0.25 & \text{for } 0 < x < 4 \\ 0 & \text{for all other } x \end{cases}$$

a. Graph the probability density function.
b. Find and graph the cumulative distribution function.
c. Find the probability that a given emergency arises within 1 mile of the northernmost point of this stretch of river.
d. The rescue team's base is at the midpoint of this stretch of river. Find the probability that a given emergency arises more than 1.5 miles from this base.

Newbold et al (2013)

THANKS!

Questions?