ISEG
Lisbon School of Economics & Management
Universidade de Lisboa

STATISTICS I

Economics / Finance/ Management

2nd Year/2nd Semester

2024/2025

LESSON 3

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

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

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

Roadmap:

- Probability
- Random variable and two dimensional random variables:
  - Distribution
  - Joint distribution
  - Marginal distribution
  - Conditional distribution functions
- Expectations and parameters for a random variable and two dimensional random variables
- Discrete Distributions
- Continuous Distributions

Bibliography: Miller & Miller, John E. (2014) Freund’s Mathematical Statistics with applications, 8th Edition, Pearson Education, [MM]

Discrete vs Continuous Data

Discrete Data	Continuous Data
It can only have specific values. “Counted” **Example**: Random Variable X – Faces of a single dice. $x \in D = \{1, 2, 3, 4, 5, 6\}$	It can take on any value in an interval. “Measured” **Example**: Random Variable X – Temperature. $x \in D = \mathbb{R}$

Discrete vs Continuous Data

Discrete Data

Probability Function or PF

f(x) = P(X = x)

Example:

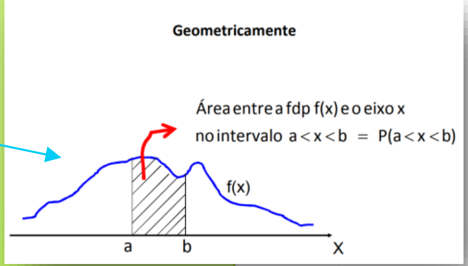
x	0	1	2
$f(x) = P(X=x)$	0.3	0.2	0.5

- $0 \leq f_X(x_j) \leq 1,\ j = 1, 2, 3, \ldots$
- $\sum_{j=1}^{\infty} f_X(x_j) = 1.$

Continuous Data

Probability Density Function or PDF

f(x)



Example:

f(x) = 1/2 , x ∈ D = (0,2)

- $f_X(x) \geq 0$ for $-\infty < x < +\infty$
- $\int_{-\infty}^{+\infty} f_X(x)dx = 1.$

Discrete vs Continuous Data

Discrete Data	Continuous Data
Cumulative Distribution Function or CDF	Cumulative Distribution Function or CDF

Discrete Data

$$F_X(x) = P(X \leq x) = \sum_{x_j \leq x} f_X(x_j).$$

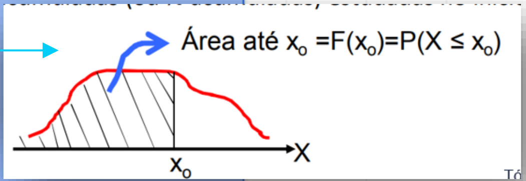
Example:

x	0	1	2
$f(x) = P(X=x)$	0.3	0.2	0.5

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.3, & 0 \leq x < 1 \\ 0.5, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

Continuous Data

$$F_X(x) = \int_{-\infty}^{x} f_X(s)\,ds.$$



$$f_X(x) = \frac{dF_X(x)}{dx}$$

Example:

f(x) = 1/2 , x ∈ D = (0,2)

$$F(x) = \begin{cases} 0, & x < 0 \\ x/2, & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

Discrete vs Continuous Data

Properties of CDFs:

1) $0 \leq F_X(x) \leq 1;$
2) $F_X(x)$ is non-decreasing: $\forall \Delta_x > 0 : F_X(x) \leq F_X(x + \Delta_x)$.
3) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$.
4) $P(a < X \leq b) = F_X(b) - F_X(a)$, for $b > a$
5) $\lim_{x \to a^+} F_X(x) = F_X(a)$; therefore X is **right continuous**
6) $P(X = a) = F_X(a) - \lim_{x \to a^-} F_X(x)$ for any real finite number.

Cumulative Distribution Function: Notes

1. $P(X \in B) = \sum_{x_i \in B} f_X(x_i);$
2. $F_X(x) = \sum_{x_i \leq x} f_X(x_i);$
3. $f_X(x) = F_X(x) - F_X(x^-)$, onde $F_X(x^-) \equiv P(X < x);$
4. $P(a \leq X \leq b) = F_X(b) - F_X(a) + f_X(a);$
5. $P(a < X < b) = F_X(b) - F_X(a) - f_X(b):$
6. $P(a \leq X < b) = F_X(b) - F_X(a) - f_X(b) + f_X(a);$
7. $P(a < X \leq b) = F_X(b) - F_X(a).$

Notas de Probabilidades e Estatística dos Professores Giovani Silva e Carlos Daniel

Discrete Random Variable: Exercises

Probability Funcion (PF) and Cumulative Distribution Function (CDF)

1

Remarks: PF and CDF

Exercises: 1, 5, 6, 7, 8, 9, 12

$$\begin{cases} 1) \quad f_X(x) \geq 0 \quad \forall x \in D_X \\ 2) \quad \sum_{x \in D_X} f_X(x) = 1 \end{cases}$$

Answer: F can be a cumulative distribution function because it seems that F is non-decreasing, $0 \leq F \leq 1$ ~~and F is right continuous~~.

1. For each of the following, determine whether the given values can serve as the values of a probability function of a random variable with the range $x = 1, 2, 3$, and 4:

   a) $f(1) = 0.25$, $f(2) = 0.75$, $f(3) = 0.25$, and $f(4) = -0.25$;
   b) $f(1) = 0.15$, $f(2) = 0.27$, $f(3) = 0.29$, and $f(4) = 0.29$;
   c) $f(1) = 1/19$, $f(2) = 10/19$, $f(3) = 2/19$, and $f(4) = 5/19$.

Exercise 1 a), b) and c)

$D_X = \{1, 2, 3, 4\}$

a) No, because $f_X(4) < 0$

b) Yes, because $\sum_{x \in D_X} f_X(x) = 1$ and $f_X(x) \geq 0 \quad \forall \, x \in D_X$

c) No, because $\sum_{x \in D_X} f_X(x) = \frac{18}{19} \neq 1$

2. Verify that $f(x) = 2x/[k(k+1)]$ for $x = 1, 2, 3, ..., k$ can serve as the probability function of a random variable with the given range.

Exercise 2

$$f_X(x) = \frac{2x}{K(K+1)} \qquad (x = 1, 2, 3, \cdots, K)$$

$$f_X(x) > 0 \quad \forall\, K \in \mathbb{N},\ \forall\, x \in D_X$$

$$\sum_{x \in D_X} f_X(x) = \sum_{x=1}^{K} \frac{2x}{K(K+1)} = \frac{2}{K(K+1)} \sum_{x=1}^{K} x =$$

(K is fixed) (arithmetic series)

$$= \frac{2}{K(K+1)} \frac{1}{2} K(K+1) = 1$$

Conclusion: Since $f_X(x) > 0 \ \forall x \in D_X$ and $\sum_{x \in D_X} f_X(x) = 1$ it is confirmed that $f_X(x)$ is a valid probability function.

Note (arithmetic series):

$$\sum_{k=1}^{n} k = \frac{1}{2} n (n+1)$$

3. For what values of k can $f(x) = (1-k)k^x$ serve as the values of the probability function of a random variable with the countably infinite range $x = 0, 1, 2, ...$?

Exercise 3

$$f_X(x) = (1-K)K^x \qquad (x = 0, 1, 2, \dots)$$

We need:

$$\begin{cases} 1) & f_X(x) \geq 0 \quad \forall x \in D_X \\ 2) & \sum_{x \in D_X} f_X(x) = 1 \end{cases}$$

Infinite geometric series (ratio K)

$$\sum_{x \in D_X} f_X(x) = \sum_{x=0}^{+\infty} (1-K)K^x = (1-K)\underbrace{\sum_{x=0}^{+\infty} K^x}_{} =$$

If $|K| < 1 \longrightarrow = (1-K)\dfrac{1}{1-K} = 1$

Exercise 3

$$f_X(x) = (1-K)K^x \geq 0 \quad \forall K \in (0,1)$$

<u>Conclusion</u>: $f_X(x) = (1-K)K^x$ serves as a probability function if $0 < K < 1$

Note (geometric series):

For $-1 < r < 1$, the sum converges as $n \to \infty$, in which case

$$S \equiv S_\infty = \sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

4. Show that $f(x) = 1/x$ cannot serve as the values of the probability function of a random variable with the countably infinite range $x = 1, 2, 3, \ldots$.

Exercise 4

$$f_X(x) = \frac{1}{x} \qquad (x = 1, 2, 3, \dots)$$

$$\sum_{x \in D_X} f_X(x) = \underbrace{\sum_{x=1}^{+\infty} \frac{1}{x}}_{\text{Harmonic series (know to diverge)}} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots > 1$$

Since $\sum_{x \in D_X} f_X(x)$ diverges, $f_X(x)$ is not a valid probability function.

5. For each of the following, determine whether the given values can serve as the values of a cumulative distribution function of a random variable with the range $x = 1, 2, 3,$ and 4:

   (a) $F(1) = 0.3$, $F(2) = 0.5$, $F(3) = 0.8$, and $F(4) = 1.2$;
   (b) $F(1) = 0.5$, $F(2) = 0.4$, $F(3) = 0.7$, and $F(4) = 1.0$;
   (c) $F(1) = 0.25$, $F(2) = 0.61$, $F(3) = 0.83$, and $F(4) = 1.0$.

Exercise 5 a), b) and c)

(a) $F(1) = 0.3$, $F(2) = 0.5$, $F(3) = 0.8$, and $F(4) = 1.2$;

Answer: F is not a cumulative distribution function because $F(4) > 1$, which is not allowed because a CDF has to verify $0 \leq F \leq 1$.

(b) $F(1) = 0.5$, $F(2) = 0.4$, $F(3) = 0.7$, and $F(4) = 1.0$;

Answer: F is not a cumulative distribution function because $F(2) < F(1)$, which means that F decreases, which is not allowed.

(c) $F(1) = 0.25$, $F(2) = 0.61$, $F(3) = 0.83$, and $F(4) = 1.0$.

Answer: F can be a cumulative distribution function because it seems that F is non-decreasing, $0 \leq F \leq 1$ ~~and F is right continuous~~.

We cant know for sure with the information that we have

6. If X has the cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1/3 & \text{for } 1 \leq x < 4 \\ 1/2 & \text{for } 4 \leq x < 6 \\ 5/6 & \text{for } 6 \leq x < 10 \\ 1 & \text{for } x \geq 10 \end{cases}$$

find

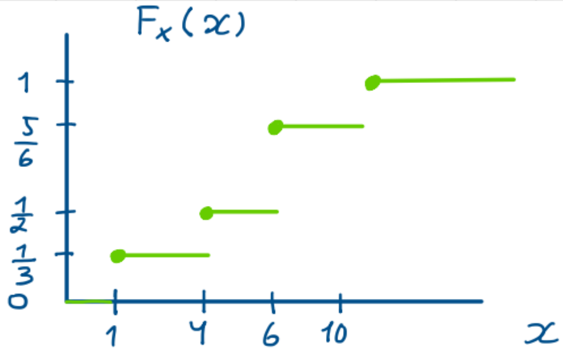
(a) $P(2 < X \leq 6)$;

(b) $P(X = 4)$;

(c) the probability function of X.

Exercise 6 a) and b)

$$P(X \le x) = F_X(x) = \begin{cases} 0 & \text{for } x < 1 \\ 1/3 & \text{for } 1 \le x < 4 \\ 1/2 & \text{for } 4 \le x < 6 \\ 5/6 & \text{for } 6 \le x < 10 \\ 1 & \text{for } x \ge 10 \end{cases}$$



a)

$$P(2 < X \le 6) = F_X(6) - F_X(2) = \frac{5}{6} - \frac{1}{3} = \frac{5}{6} - \frac{2}{6} = \frac{3}{6} = \frac{1}{2}$$

b)

$$\underbrace{P(X = 4)}_{f_X(4)} = F_X(4) - F_X(4^-) = \frac{1}{2} - \frac{1}{3} = \frac{3}{6} - \frac{2}{6} = \frac{1}{6}$$

Exercise 6 c)

c)

$$D_X = \{1, 4, 6, 10\}$$

$$f_X(x) = P(X = x) = F_X(x) - F_X(x^-)$$

$$f_X(1) = \frac{1}{3} - 0 = \frac{1}{3}$$

$$f_X(6) = \frac{5}{6} - \frac{1}{2} = \frac{5}{6} - \frac{3}{6} = \frac{2}{6} = \frac{1}{3}$$

$$f_X(4) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

$$f_X(10) = 1 - \frac{5}{6} = \frac{1}{6}$$

Conclusion:
$$f_X(x) = \begin{cases} \frac{1}{3} & (x = 1, 6) \\ \frac{1}{6} & (x = 4, 6) \\ 0 & \text{otherwise} \end{cases}$$

7. Find the cumulative distribution function of the random variable that has the probability function $f(x) = x/15$ for $x = 1, 2, 3, 4, 5$.

Exercise 7

$$f_x(x) = \frac{x}{15} \qquad (x = 1, 2, 3, 4, 5)$$

$$F_x(1) = f_x(1) = \frac{1}{15}$$

$$F_x(2) = f_x(1) + f_x(2) = \frac{1}{15} + \frac{2}{15} = \frac{3}{15}$$

$$F_x(3) = f_x(1) + f_x(2) + f_x(3) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} = \frac{6}{15}$$

$$F_x(4) = f_x(1) + f_x(2) + f_x(3) + f_x(4) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} + \frac{4}{15} = \frac{10}{15}$$

$$F_x(5) = f_x(1) + f_x(2) + f_x(3) + f_x(4) + f_x(5) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} + \frac{4}{15} + \frac{5}{15} = 1$$

Exercise 7

Conclusion:

$$F_X(x) = \begin{cases} 0 & (x < 1) \\ \frac{1}{15} & (1 \leq x < 2) \\ \frac{3}{15} & (2 \leq x < 3) \\ \frac{6}{15} & (3 \leq x < 4) \\ \frac{10}{15} & (4 \leq x < 5) \\ 1 & (x \geq 5) \end{cases}$$

Note: In the solutions, the intervals are wrong

8. To make a study about the quality of public transports in a certain city, the Mayor wants to know how many people arrive at a bus stop to catch a bus between two consecutive bus arrivals. Let X be a random variable that provides this information, with the following probability function:

x	0	1	2	3	4	5	6 or more
$P(X = x)$	0.1	0.15	0.20	0.25	a	b	0.05

Find a and b such that

a) $P(X \geq 5) = 0.15$;

b) $P(X \in \{1, 4\}) = 0.35$;

c) $F_X(4) = 0.8$.

Exercise 8 a) and b)

a)

$$P(X \geq 5) = b + 0.05 = 0.15 \Leftrightarrow b = 0.1$$

$$\sum_{x \in D_X} f_X(x) = 1 \Leftrightarrow 0.1 + 0.15 + 0.2 + 0.25 + a + 0.1 + 0.05 = 1 \Leftrightarrow$$

$$\Leftrightarrow a + 0.85 = 1 \Leftrightarrow a = 0.15$$

b)

$$P(X \in \{1, 4\}) = f_X(1) + f_X(4) = 0.15 + a = 0.35 \Leftrightarrow$$

$$\Leftrightarrow a = 0.2$$

$$\sum_{x \in D_X} f_X(x) = 1 \Leftrightarrow 0.1 + 0.15 + 0.2 + 0.25 + 0.2 + b + 0.05 = 1 \Leftrightarrow$$

$$\Leftrightarrow 0.95 + b = 1 \Leftrightarrow b = 0.05$$

Exercise 8 c)

e)

$$F_X(4) = 0.1 + 0.15 + 0.2 + 0.25 + a = 0.8 \Leftrightarrow$$

$$\Leftrightarrow a + 0.7 = 0.8 \Leftrightarrow a = 0.1$$

$$1 - F_X(4) = P(X > 4) = b + 0.05 = 0.2 \Leftrightarrow b = 0.15$$

2

Continuous Random Variable: Exercises

Probability Density Funcion (PDF) and Cumulative Distribution Function (CDF)

Integral: Formules

(a) Temos que $f(x) \geq 0$ e $\int_{-\infty}^{+\infty} f(x)\,dx = \int_1^4 \frac{1}{3}\,dx = \frac{1}{3}\,x\Big|_1^4 = \frac{1}{3}\,(4\text{-}1) = 1$

portanto f(x) é uma fdp.

(b) Temos: $P(2 < x < 3) = \int_2^3 \frac{1}{3}\,dx = \frac{1}{3}\,x\Big|_2^3 = \frac{1}{3}\,(3\text{-}2) = \frac{1}{3}$

Tabela 1.1: Tabela de Primitivas Elementares

f	$Pf = F$
$c, c \in IR$	$c\,x$
x^{α} $(\alpha \neq -1)$	$\frac{x^{\alpha+1}}{\alpha+1}$
$\frac{1}{x}$	$\log\lvert x\rvert$
e^x	e^x
$\cos x$	$\sin x$
$\sin x$	$-\cos x$
$\sec^2 x$	$\mathrm{tg}x$
$\mathrm{cosec}^2 x$	$-\mathrm{cotg}x$
$\frac{1}{1+x^2}$	$\mathrm{arctg}\,x$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$
$\cosh x$	$\sinh x$
$\sinh x$	$\cosh x$

Tópico_08_.pdf (usp.br)

Remarks: PDF and CDF

Since $f_X(x) \geqslant 0 \ \forall x \in D_X$ and $\int_{x \in D_X} f_X(x) = 1$

Answer: F can be a cumulative distribution function because it seems that F is non-decreasing, $0 \leq F \leq 1$ ~~and F is right continuous~~.

9. The probability density of the continuous random variable X is given by

$$f_X(x) = \begin{cases} 1/5 & 2 < x < 7 \\ 0 & elsewhere \end{cases}$$

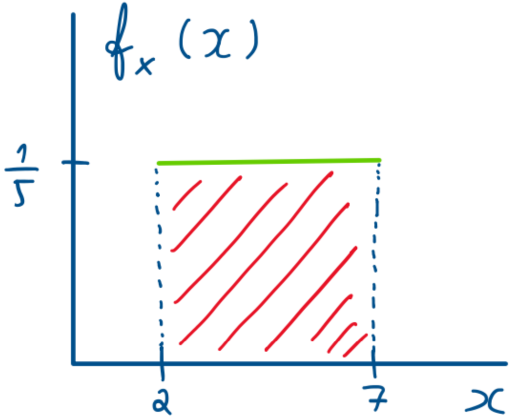
(a) Draw its graph and verify that the total area under the curve (above the x-axis) is equal to 1.

(b) Find $P(3 < X < 5)$.

Exercise 9 a)

$$f_X(x) = \begin{cases} \frac{1}{5} & (2 < x < 7) \\ 0 & (\text{elsewhere}) \end{cases}$$

a)

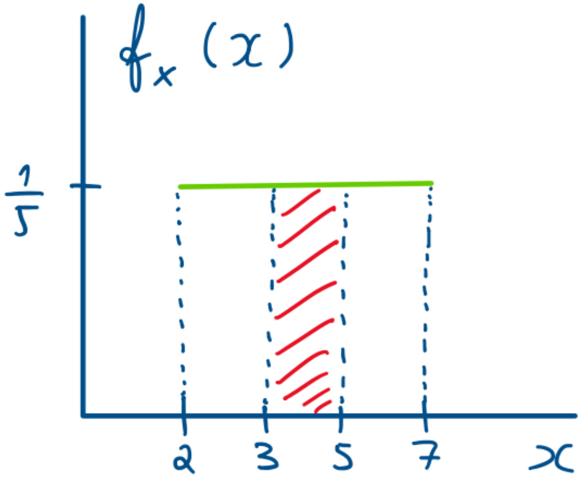


— $f_X(x)$

$$\text{▨} = \frac{1}{5} \cdot (7-2) = 1$$

Exercise 9 b)

b)



$$P(3 < X < 5) = \frac{1}{5} \times (5-3) = \frac{2}{5}$$

10. Let $f_X(x) = e^{-x}$ for $0 < x < +\infty$.

(a) Show that $f_X(x)$ represents a probability density function.

(b) Sketch a graph of this function and indicate the area associated with the probability that $X > 1$.

(c) Calculate the probability that $X > 1$.

Exercise 10 a)

$f_X(x) = e^{-x} \quad (x > 0)$

a)

$D_X = (0, +\infty)$

$f_X(x) > 0 \quad \forall \quad x \in D_X$

$$\int_{x \in D_X} f_X(x)\,dx = \int_0^{+\infty} e^{-x}\,dx = \lim_{b \to +\infty} [-e^{-x}]_0^b =$$

$$= \lim_{b \to +\infty} (-e^{-b} - (-e^0)) =$$

$$= \lim_{b \to +\infty} (1 - e^{-b}) = 1 - \underbrace{\lim_{b \to +\infty} e^{-b}}_{0} = 1$$

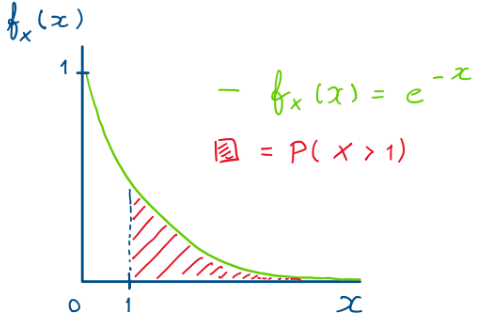
Note: $e^{-b} = (e^b)^{-1} = \frac{1}{e^b}$

Conclusion:

Since $f_X(x) \geq 0 \ \forall \ x \in D_X$ and $\int_{x \in D_X} f_X(x) = 1$

we conclude that $f_X(x)$ is a valid

Exercise 10 b) and c)



c)

$$P(X>1) = \int_1^{+\infty} e^{-x}\,dx = \lim_{b \to +\infty} \left[-e^{-x}\right]_1^b =$$

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

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

$$= e^{-1} - \underbrace{\lim_{b \to +\infty} e^{-b}}_{0} = e^{-1}$$

11. Let $f_X(x) = 3x^2$ for $0 < x < 1$.

(a) Show that $f_X(x)$ represents a density function.

(b) Sketch a graph of this function, and indicate the area associated with the probability that $0.1 < X < 0.5$.

(c) Calculate the probability that $0.1 < X < 0.5$.

Exercise 11 a)

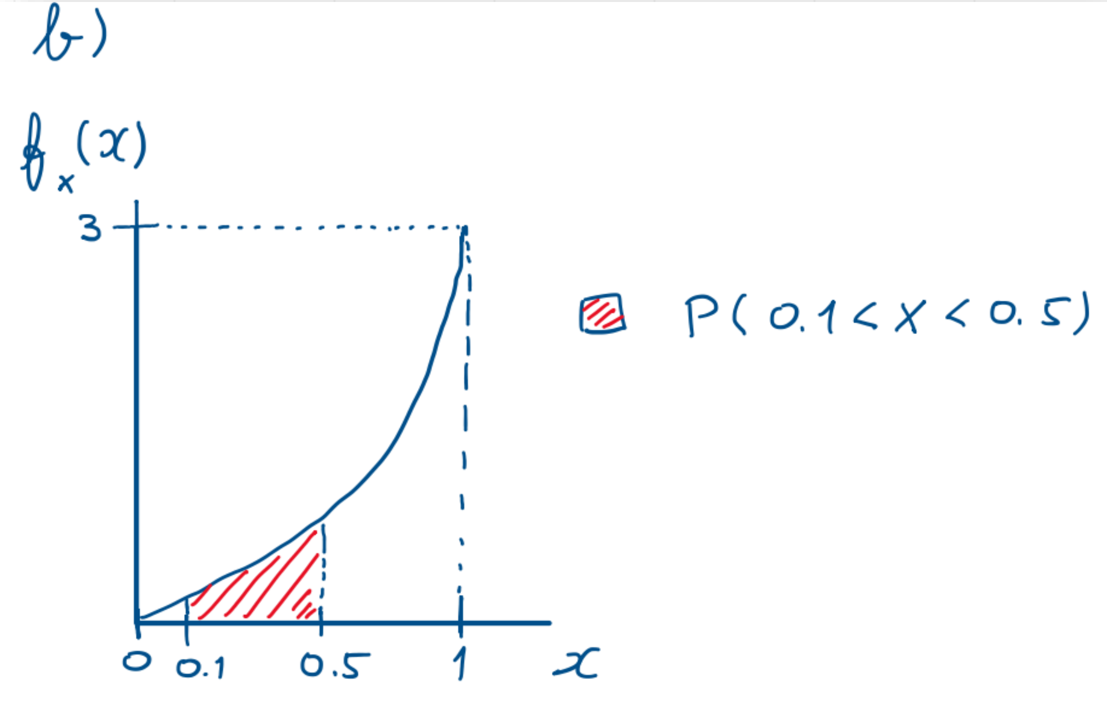
$$f_X(x) = 3x^2 \qquad (0 < x < 1)$$

(a)

$$f_X(x) = 3x^2 > 0 \quad \forall\ x \in D_X = (0, 1)$$

$$\int_{x \in D} f_X(x)\, dx = \int_0^1 3x^2\, dx = [x^3]_0^1 = 1^3 - 0^3 = 1$$

Exercise 11 b)



Exercise 11 c)

c)

$$P(0.1 < X < 0.5) = \int_{0.1}^{0.5} f_X(x)\,dx = \int_{0.1}^{0.5} 3x^2\,dx$$

$$= \left[x^3\right]_{0.1}^{0.5} = 0.5^3 - 0.1^3 = 0.124$$

Thanks!

Questions?