CS 107, Probability, Spring 2020
Lecture 22

Michael Poghosyan
mpoghosyan@aua.am

AUA

17 March 2020

Content

- Continuous Random Variables

Last Lecture ReCap

- R.v X is Discrete, if the Range of X, i.e., the set of all possible values of X is finite or countably infinite.
- Discrete r.v. X can be described through its PMF:

Values of X	x_1	x_2	x_3	...
$\mathbb{P}(X = x)$	p_1	p_2	p_3	...

- Using the PMF, we can calculate Probabilities $\mathbb{P}(X \in A) = \sum_{x_k \in A} p_k$;
- the CDF of a Discrete r.v. is a piecewise constant (step) function.

Continuous R.V.

Intro: Continuous Random Variables

There are many r.v.s that are not Discrete. For example, the lifetime of a light bulb, the height of a (randomly chosen) person, the exact age (not rounded) of a person etc, the waiting time for the quiz grades- I mean, metro train at the bus station $\ddot\smile$ are all examples of non-Discrete r.v.s.

Prelude, Continuous Random Variables

Now we introduce Continuous r.v.s: unlike Discrete r.v.s, we cannot describe them using the probabilities $\mathbb{P}(X = x)$, since these probabilities are 0 for any continuous r.v. X, and any real number x. So we will not consider the Probability of $X = x$, rather the combined Probabilities of the form $\mathbb{P}(a \leq X \leq b)$.

The idea of defining Probabilities for Continuous r.v.s is similar to defining the Mass of a physical body: we cannot define it through the mass of points, because each point's mass will be 0. We define the Mass of some part of our body by integrating the Density. In the analogy, we will define the Probability $\mathbb{P}(a \leq X \leq b)$ by integrating the Probability Density Function over $[a, b]$.

Zeno's paradox: Consider an arrow in the flight. At any time instant, the arrow is not moving, is motionless. Think as recording the flight and then considering just one frame at a time. Then how can it move at all?

Continuous Random Variables

Assume X is a r.v. and $F(x)$ is its CDF.

Continuous r.v.

We say that the r.v. X is **absolutely continuous** or just **continuous**, if there exists an integrable non-negative function $f : \mathbb{R} \to \mathbb{R}$ such that

$$F(x) = \int_{-\infty}^{x} f(t)dt, \qquad \forall x \in \mathbb{R}.$$

Here f is called the **Probability Density Function (PDF)** of X. Sometimes we will write $f_X(x)$ to stress that f is the PDF of X.

Remark: By a Theorem of $RA \cup Calc2$, if X is a continuous r.v., then $F(x)$ will be continuous at every x.

Relation between PDF and CDF for Continous r.v.

Assume $F(x)$ is the CDF and $f(x)$ is the PDF of a r.v. X. Then, by definition,

$$F(x) = \int_{-\infty}^{x} f(t)dt, \qquad \forall x \in \mathbb{R}.$$

Hence, using $RA \cup Calc2$, we get

$$f(x) = F'(x), \qquad \text{if } f \text{ is countinuous at } x.$$

Characterization of PDF

What are the characteristic properties of a PDF? The following Theorem give necessary and sufficient conditions in a function for being a PDF for some r.v.

Characterization of PDF

If X is a continuous r.v. with a PDF $f(x)$, then

a. $f(x) \geq 0$ for all $x \in \mathbb{R}$;

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

Inversely, if f is an integrable function satisfying the above two conditions, then there is a Probability Space and a r.v. on that Space such that f is the PDF of that r.v.

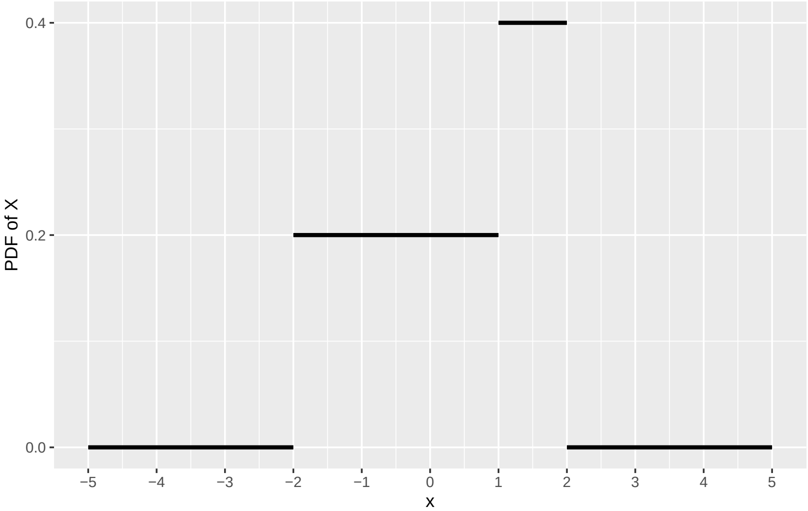
Notes on PDFs

Note: If X is a continuous r.v., that doesn't mean that its PDF, $f(x)$, is continuous. Rather, it is a non-negative integrable function, with integral 1 over $\mathbb{R}$. Moreover, if x_0 is a discontinuity point of f, it is not necessary that f is right-continuous at x_0. It can be left-continuous, or f can be discontinuous from both sides. In fact, the value of f at finite or countably many number of points is not important.

Note: If X is a Continuous r.v., then its CDF is uniquely determined, but its PDF is not uniquely determined: we can change the value of its PDF at some point, and this new function will still be a PDF of X.

Notes on PDFs

Example 22.1:



Notes on PDFs

Note: If X is continuous r.v., then its CDF, $F(x)$, is a continuous function in $\mathbb{R}$. But the inverse is not true, in general: if the CDF $F(x)$ is continuous everywhere, that doesn't mean that X is continuous. But this is a $18+$ topic, and we will skip it $\ddot\smile$

Note: In all of our examples we will consider, PDFs either will be equal to 0 outside some interval, or will tend to 0 when $x \to +\infty$ and $x \to -\infty$. But, in general case, this is not true: if $f(x)$ is the PDF of some r.v., then not necessarily $f(x) \to 0$ as $x \to +\infty$ or as $x \to -\infty$.

Note: Recall that for the CDF $F(x)$ of a r.v. we have that $0 \leq F(x) \leq 1$, for any x. Does this inequality hold true also for PDFs? I.e., if $f(x)$ is the PDF of some continuous r.v. X, is it true that $0 \leq f(x) \leq 1$ for any x? The answer is: No, in general.

Geometric Interpretation of PDF

Recall that

$$\mathbb{P}(X \le x) = F(x) = \int_{-\infty}^{x} f(t)\,dt, \qquad \forall x \in \mathbb{R}.$$

And from $RA \cup Calc2$ you know that the integral on the RHS is the *Area under the curve* $y = f(x)$ *over the interval* $(-\infty, x]$. Hence,

$$\mathbb{P}(X \le x) = F(x) = Area(\text{region under } y = f(x) \text{ on } (-\infty, x]).$$

Geometric Interpretation of PDF

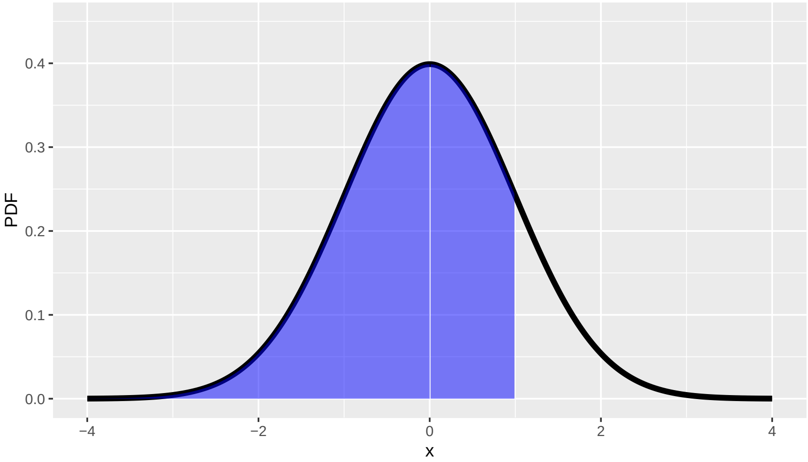
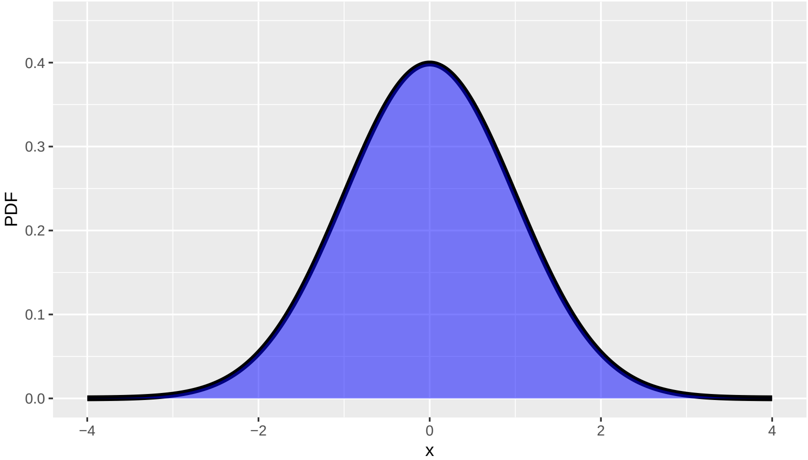


Figure: The Shaded Region area is $\mathbb{P}(X \leq 1) = F(1)$

Geometric Interpretation of PDF

Fact: The area under the PDF graph is 1.



Using PDF to calculate Probabilities

Recall again that, if f and F are the PDF and CDF of r.v. X, then for any $x \in \mathbb{R}$,

$$\mathbb{P}(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt.$$

And we know that

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx =$$

$$= \int_{a}^{b} f(x)dx.$$

So we have obtained the following relation:

$$\mathbb{P}(a < X \leq b) = \int_{a}^{b} f(x)dx.$$

Using PDF to calculate Probabilities

The following properties are true:

Calculation of Probabilities by PDF

If X is a r.v. with the PDF $f(x)$ and CDF $F(x)$, then

- $\mathbb{P}(X = x) = 0$;
- $\mathbb{P}(a < X \le b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a \le X \le b) =$

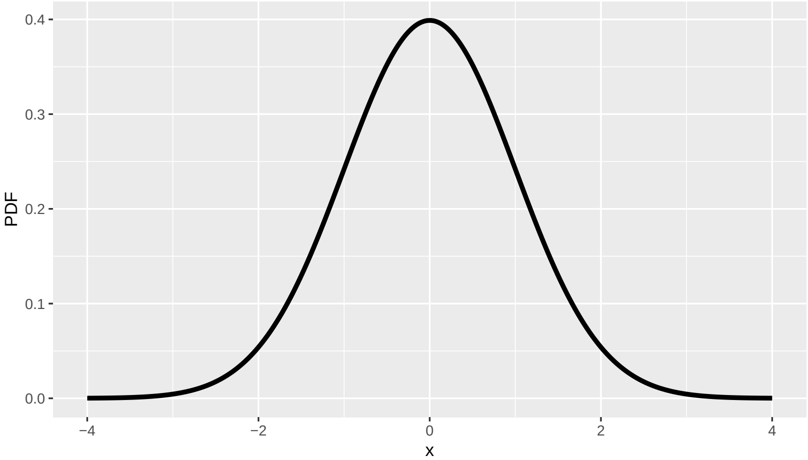
$$= \mathbb{P}(a < X < b) = \int_a^b f(x)dx$$

And, in general,

$$\mathbb{P}(X \in A) = \int_A f(x)dx.$$

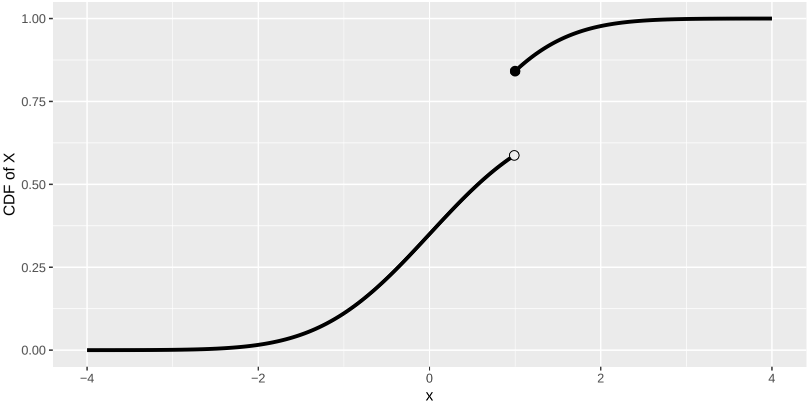
Probability by a PDF

Example 22.2: Below is the graph of the PDF $f(x)$ for some r.v. X. Which one is larger: $\mathbb{P}(X = 0)$ or $\mathbb{P}(X = 2)$?



CDF of a continuous r.v.

Example 22.3: Below is the graph of the CDF $F(x)$ for some r.v. X. Is X Continuous?

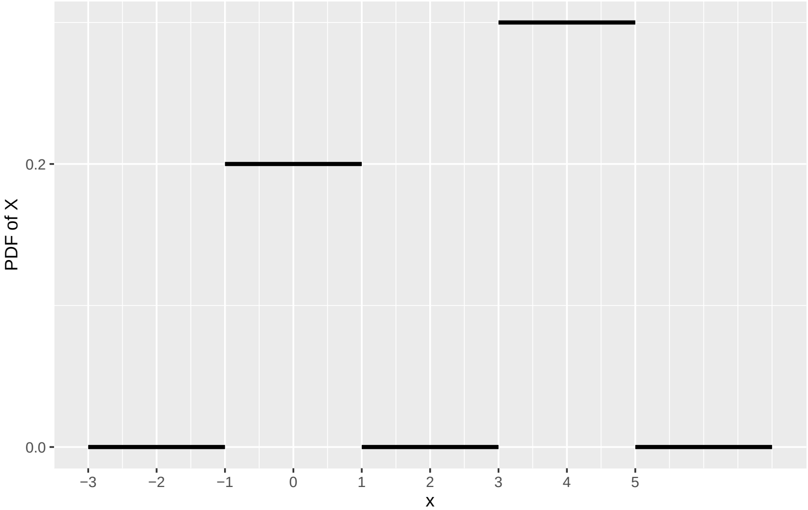


Probability by a PDF

Example 22.4: On the next slide you will find the graph of some function $f(x)$.

a. Is it a PDF of some r.v. X?
b. If it is, show geometrically and calculate $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 1)$, $\mathbb{P}(X < -1)$, $\mathbb{P}(X \geq 0)$, $\mathbb{P}(0 \leq X \leq 2)$ and $\mathbb{P}(X > 4)$;
c. Find $F(0)$, where F is the CDF of X;
d. Find, analytically, $F(x)$, for any x;
e. Plot the graph the CDF $F(x)$.

Example 22.4



Probability by a PDF

Example 22.5: Which of the following functions will be a PDF of some r.v.?

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad g(x) = \begin{cases} 1, & x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

$$h(x) = sigmoid(x), \qquad m(x) = \begin{cases} 0, & x < 0 \text{ or } x > 2 \\ \frac{x}{2}, & x \in [0,2] \end{cases}$$

$$k(x) = \begin{cases} x, & x \in [-1, \sqrt{3}] \\ 0, & otherwise. \end{cases}$$

Probability by a PDF

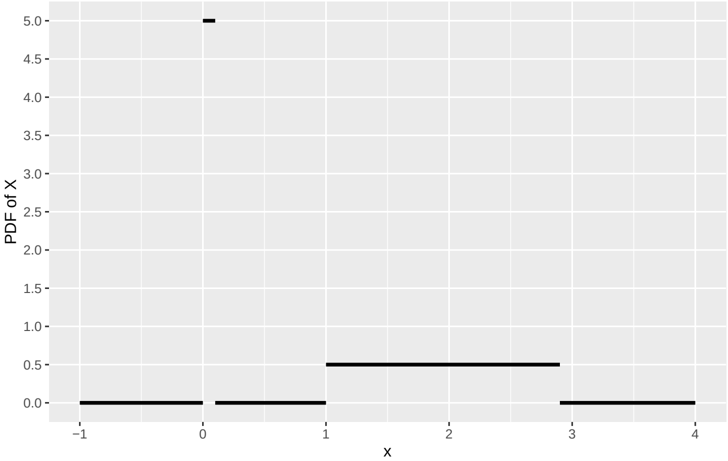
Example 22.6: Assume

$$f(x) = \begin{cases} K \cdot (x^2+1), & x \in [0,1] \\ 0, & \textit{otherwise}. \end{cases}$$

a. For which value of K, f will be a PDF of some r.v. X?
b. For that K and X, calculate $\mathbb{P}(X < 0.5)$ and $\mathbb{P}(X \in [0.3, 0.5])$;
c. Calculate $\mathbb{P}(X < 0.7 | X > 0.5)$;
d. Find $F(0.5)$ and $F(1)$, where F is the CDF of X;
e. Find analytically $F(x)$ for any $x \in \mathbb{R}$;
f. Plot the graphs of f and F (using **R**!).

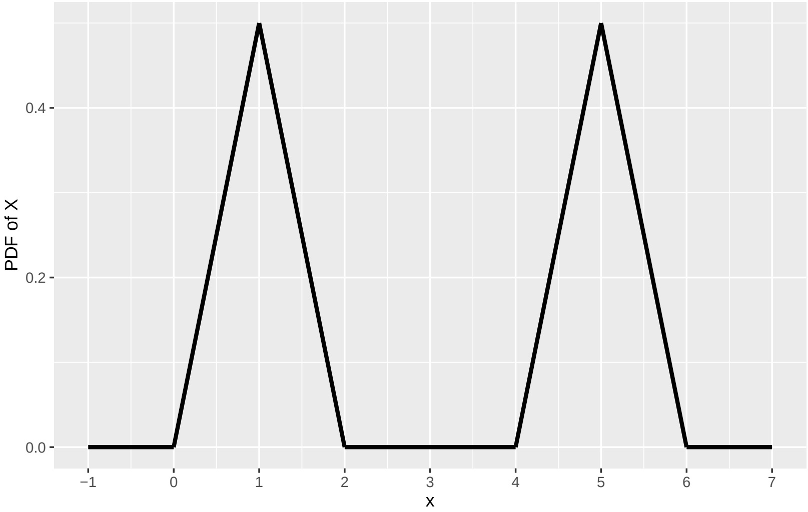
Reading info about a r.v. from its PDF

Example 22.7: We consider the r.v. X given by its PDF $f(x)$ below (the discontinuity points of f are 0, 0.01, 1 and 2.9). Are the chances that X is close to 0 high?



Reading info about a r.v. from its PDF

Example 22.8: Describe the r.v. X given by its PDF $f(x)$:



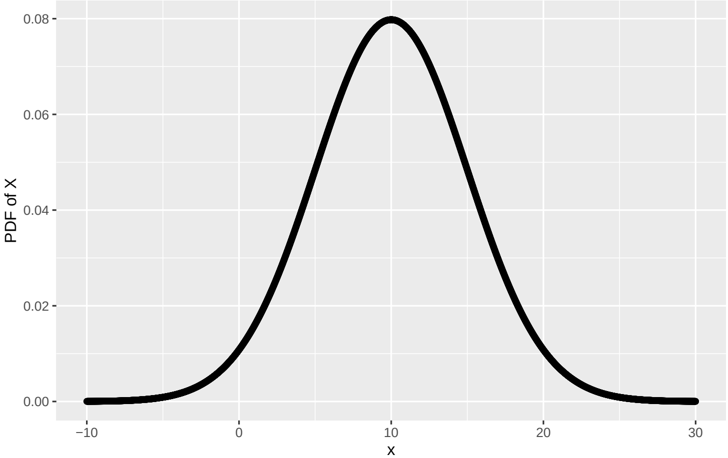
PDF and CDF relation, geometrically

Example 22.9:

- Given the graph of the PDF of the r.v. X, construct the graph of its CDF.
- Given the graph of the CDF of the r.v. X, construct the graph of its PDF.

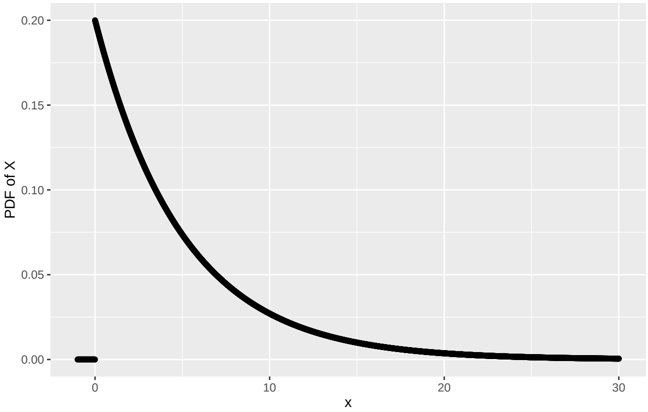
Reading info about a r.v. from its PDF

Example 22.10: Let X be the r.v. showing the closing price of some stock today, in $\$K$. Can the function given by the graph below be the PDF of X? (The area under the graph is 1).



Reading info about a r.v. from its PDF

Example 22.11: Assume X is the r.v. representing the weight in Kg of a randomly chosen person at AUA. Can the function given by the graph below be the PDF of X? (The area under the graph is 1).



Summary on Discrete and Continuous r.v.s

Summary on Discrete and Cont. r.v.s

We have learned that:

- we can describe a r.v.s by its CDF - this will work for the general case, for **any** r.v., both Discrete and Continuous, and even for Singular or Mixture r.v.s;
- if r.v. is Discrete, then we can describe it by its PMF
- if r.v. is Continuous, then we can describe it by its PMF

That is, we can calculate Probabilities about the possible values of a r.v. through the CDF or PMF/PDF.

Probabilities through the CDF

If X is any r.v., then the CDF of X is defined by

$$F(x) = \mathbb{P}(X \leq x), \qquad x \in \mathbb{R}.$$

And using the CDF, we can calculate probabilities like

$$\mathbb{P}(X = a) = F(a) - F(a-),$$

$$\mathbb{P}(a < X \leq b) = F(b) - F(a).$$

Probabilities through the PM(D)F

Let's consider now two cases - when X is Discrete or Continuous:

Discrete, by PMF	Continuous, by PDF
$\mathbb{P}(X = x_k) = F(x_k) - F(x_k-)$	PDF: $f(x) = F'(x)$
$\mathbb{P}(X = x_k) = p_k$	$\mathbb{P}(X = x) = 0, \ \forall x \in \mathbb{R}$
$\displaystyle\sum_k p_k = 1$	$\displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1$
$\displaystyle F(x) = \sum_{x_k \leq x} p_k$	$\displaystyle F(x) = \int_{-\infty}^{x} f(t)dt$
$\displaystyle\mathbb{P}(a \leq X \leq b) = \sum_{a \leq x_k \leq b} p_k$	$\displaystyle\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$

A warning about a common mistake

Assume X is a r.v., and a is a number. Then X **is not an Event**. Instead, the sets $\{X = a\}$, $\{X < a\}$, $\{X \le a\}$,... are Events. So:

- Writing $\mathbb{P}(X)$ is **incorrect**;
- Writing $\mathbb{P}(a)$ is **incorrect**;
- Writing $\mathbb{P}(X = a)$ or $\mathbb{P}(X < a)$ or $\mathbb{P}(X \ge a)$, ... are **correct**.

Example: Let X be the number of emails I will get today. Then X is a r.v. Is it Discrete or Continuous? Then $\mathbb{P}(X)$ denotes "the Probability of number of emails I will get today". Nonsense, of course $\ddot\smile$ But we can talk about $\mathbb{P}(X = 0)$ or $\mathbb{P}(X > 10)$, and they have clear meanings.

Important Discrete R.V.