LESSON 13

Continuous Random Variables

CSCI 3022

Course Logistics: Your Fifth Week At A Glance

Mon	Tues	Wed	Thurs	Fri
Complete Recorded Lesson Video Assignment(Canvas)		Attend & Participate in Class	HW 5 Due 11:59pm via Gradescope	In Class Quiz 4 Scope: Lessons 7-8 HW 4 Attend & Participate in Class
	Quiz 3 feedback/ grades posted		HW 4 feedback/ grades posted	HW 6 released

HW Note:

To receive credit on problems involving integration, you must show all steps evaluating the integral and simplifying.

No Credit:

$$\int_1^3 x^2\,dx = \frac{7}{3}$$

Full Credit:

$$\int_1^3 x^2\,dx$$

$$= \frac{x^3}{3}\Bigg|_1^2$$

$$= \frac{2^3}{3} - \frac{1^3}{3}$$

$$= \boxed{\frac{7}{3}}$$

Learning Objectives:

- Explain the difference between Probability Density Functions (PDFs) and Cumulative Distribution Functions (CDFs) and use both to calculate probabilities for continuous random variables
- Calculate Expected Value and Variance of Continuous RV

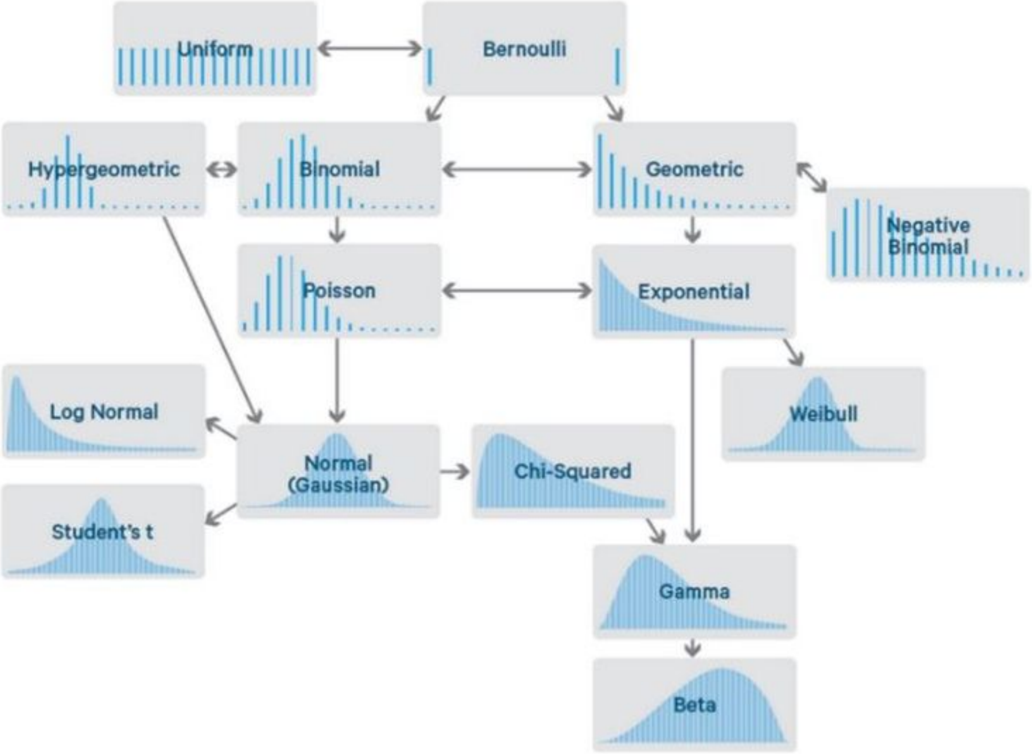
RoadMap

Introduction to Continuous RVs

Probability Density Functions (PDF)

- Expected Value and Variance of Continuous RV
- Cumulative Distribution Functions (CDF)

Scope



In this class we're going to focus on a small subset of common distributions:

Discrete:
- Bernoulli
- Binomial
- Poission

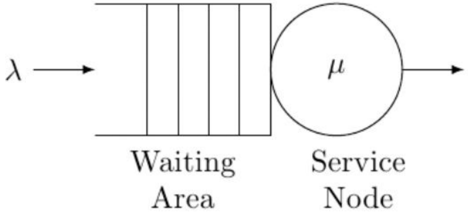
Continuous:
- Uniform
- Exponential
- Normal

*See [supporting materials] for info on Geometric and Negative Binomial

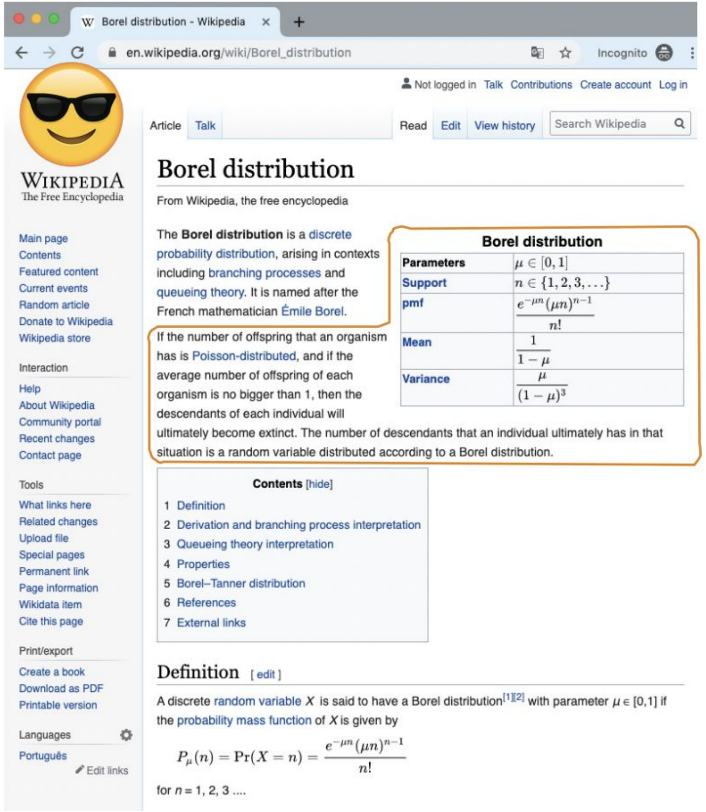
CSCI 3022 Learning Goal: Use New Random Variables

Let's say you are learning about servers and networks.

You read about the M/D/1 queue:



"The service time busy period is distributed as a Borel with parameter $\mu = 0.2$."


Borel distribution - Wikipedia
en.wikipedia.org/wiki/Borel_distribution
Incognito
Not logged in Talk Contributions Create account Log in
Article Talk Read Edit View history Search Wikipedia
WIKIPEDIA
The Free Encyclopedia
Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page
Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page
Print/export
Create a book
Download as PDF
Printable version
Languages
Português
Edit links

Borel distribution

From Wikipedia, the free encyclopedia

The **Borel distribution** is a discrete probability distribution, arising in contexts including branching processes and queueing theory. It is named after the French mathematician Émile Borel.

If the number of offspring that an organism has is Poisson-distributed, and if the average number of offspring of each organism is no bigger than 1, then the descendants of each individual will ultimately become extinct. The number of descendants that an individual ultimately has in that situation is a random variable distributed according to a Borel distribution.

Borel distribution	
Parameters	$\mu \in [0,1]$
Support	$n \in \{1,2,3,\ldots\}$
pmf	$\frac{e^{-\mu n}(\mu n)^{n-1}}{n!}$
Mean	$\frac{1}{1-\mu}$
Variance	$\frac{\mu}{(1-\mu)^3}$

Contents [hide]

Definition [edit]

A discrete random variable X is said to have a Borel distribution[1][2] with parameter $\mu \in [0,1]$ if the probability mass function of X is given by

$$P_\mu(n) = \Pr(X = n) = \frac{e^{-\mu n}(\mu n)^{n-1}}{n!}$$

for $n = 1, 2, 3$ ....

Goal: You can recognize terminology and understand experiment setup.

Probability Density Functions (PDF)

Introduction to Continuous RVs

- **Probability Density Functions (PDF)**
- Expected Value and Variance of Continuous RV
- Cumulative Distribution Functions (CDF)

People heights

You are volunteering at the local elementary school fundraiser.

- To buy a t-shirt for your friend Vanessa, you need to know her height.

1. What is the probability that your friend is 54.0923857234 inches tall?

People heights

You are volunteering at the local elementary school fundraiser.

- To buy a t-shirt for your friend Vanessa, you need to know her height.

1. What is the probability that your friend is 54.0923857234 inches tall?

Essentially 0

People heights

You are volunteering at the local elementary school fundraiser.

- To buy a t-shirt for your friend Vanessa, you need to know her height.

1. What is the probability that your friend is 54.0923857234 inches tall?

   Essentially 0

2. What is the probability that Vanessa is between 52–56 inches tall?

People heights

You are volunteering at the local elementary school fundraiser.

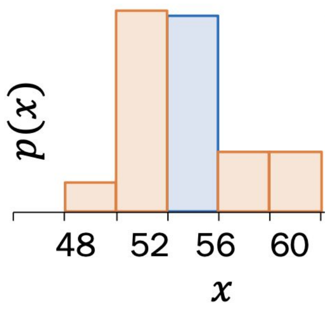
- To buy a t-shirt for your friend Vanessa, you need to know her height.

1. What is the probability that your friend is 54.0923857234 inches tall?

Essentially 0

2. What is the probability that Vanessa is between 52–56 inches tall?

$P(52 < X \leq 56)$



People heights

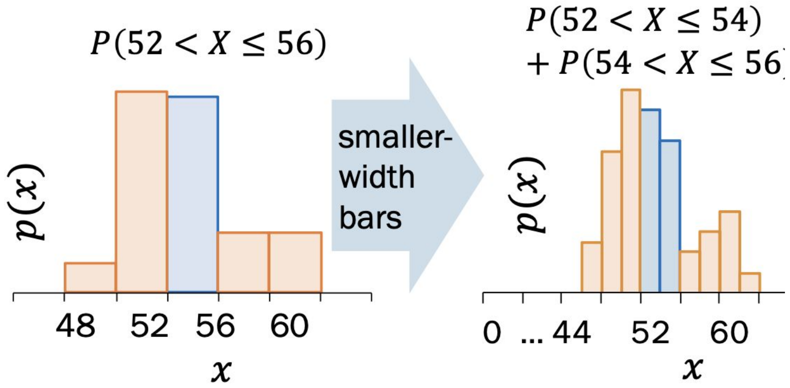
You are volunteering at the local elementary school fundraiser.

- To buy a t-shirt for your friend Vanessa, you need to know her height.

1. What is the probability that your friend is 54.0923857234 inches tall?

Essentially 0

2. What is the probability that Vanessa is between 52–56 inches tall?



People heights

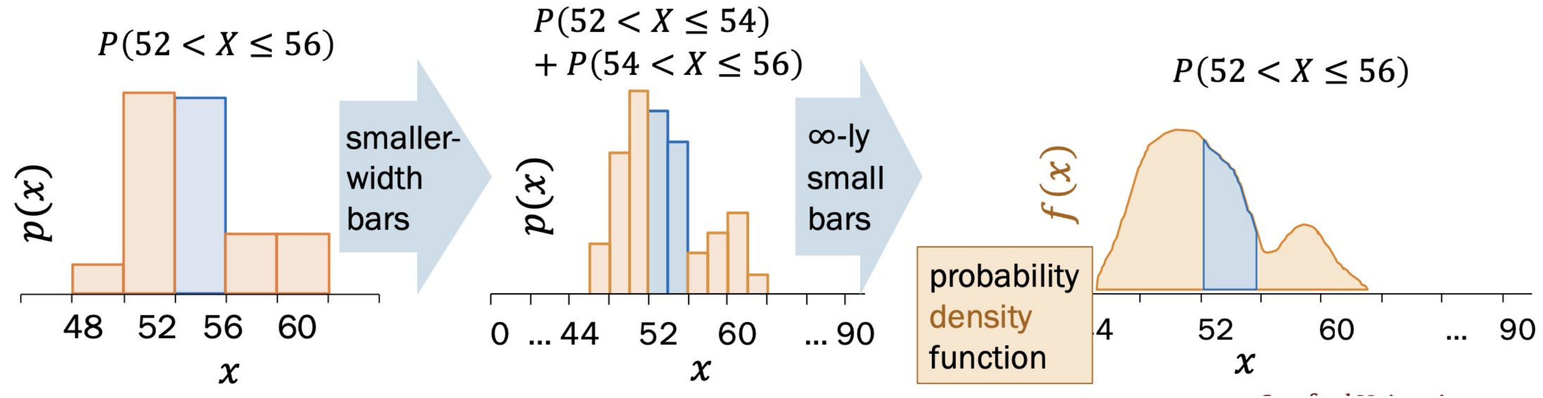
You are volunteering at the local elementary school fundraiser.

- To buy a t-shirt for your friend Vanessa, you need to know her height.

1. What is the probability that your friend is 54.0923857234 inches tall?

   Essentially 0

2. What is the probability that Vanessa is between 52–56 inches tall?



Continuous Random Variable Definition:

A random variable X is continuous if there is a probability density function $f(x) \geq 0$ such that for $-\infty < x < \infty$:

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

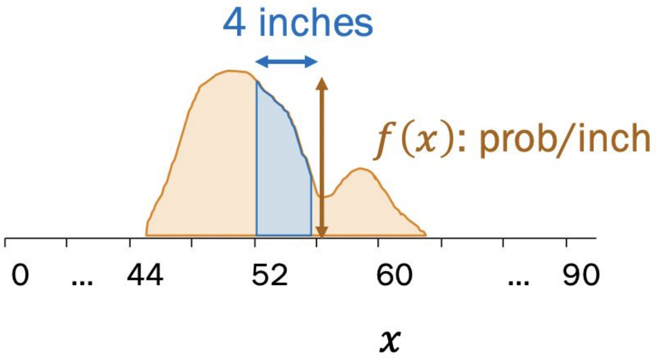
PDF Units: probability per units of X

Integrating a PDF must always yield a valid probability, no matter the values of a and b. The PDF must also satisfy:

$$\int_{-\infty}^{\infty} f(x)\,dx = P(-\infty < X < \infty) = 1$$

Main Takeaway # 1

Integrate $f(x)$ to get probabilities.



PDF Units: probability per units of X

$$P(52 \leq X \leq 56) = \int_{52}^{56} f(x)\,dx$$

Comparing probability MASS vs probability DENSITY functions

PMF vs PDF

Discrete random variable X

Probability mass function (PMF):

$$p(x)$$

To get probability:

$$P(X = x) = p(x)$$

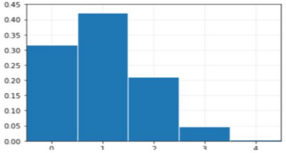
Continuous random variable X

Probability density function (PDF):

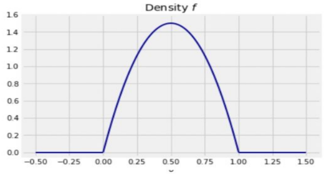
$$f(x)$$

To get probability:

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



Both are measures of how likely X is to take on a value or some range of values.

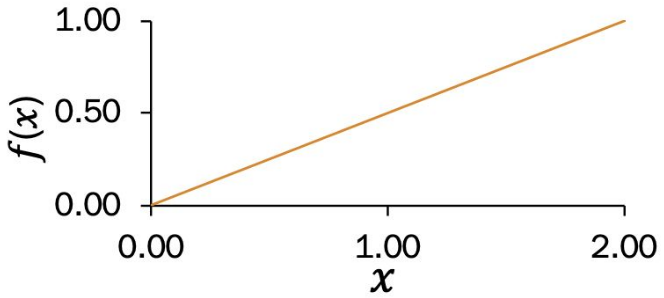


Computing probability

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

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$



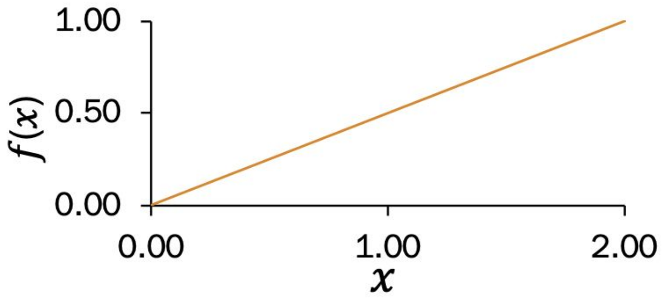
What is $P(X \ge 1)$?

Computing probability

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

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$



What is $P(X \ge 1)$?

Strategy 1: Integrate

$$P(1 \le X < \infty) = \int_1^\infty f(x)dx = \int_1^2 \frac{1}{2}x dx$$

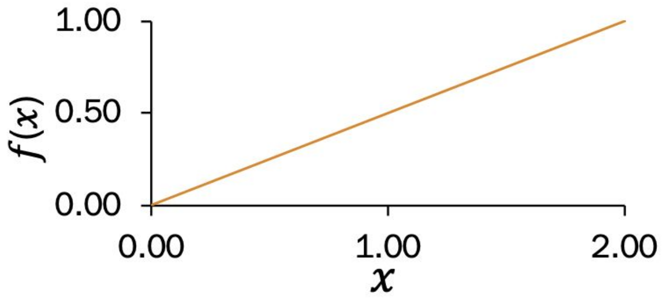
$$= \frac{1}{2}\left(\frac{1}{2}x^2\right)\Big|_1^2 = \frac{1}{2}\left[2 - \frac{1}{2}\right] = \frac{3}{4}$$

Computing probability

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

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$



What is $P(X \ge 1)$?

Strategy 1: Integrate

$$P(1 \le X < \infty) = \int_1^\infty f(x)dx = \int_1^2 \frac{1}{2}x dx$$

$$= \frac{1}{2}\left(\frac{1}{2}x^2\right)\Big|_1^2 = \frac{1}{2}\left[2 - \frac{1}{2}\right] = \frac{3}{4}$$

Strategy 2: Know triangles

$$1 - \frac{1}{2}\left(\frac{1}{2}\right) = \frac{3}{4}$$

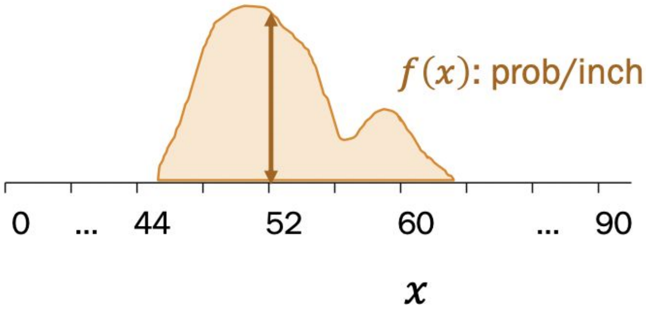
Wait! Is this even legal?

$$P(0 \le X < 1) = \int_0^1 f(x)dx\,??$$

Main Takeaway #2

For a continuous random variable X with PDF $f(x)$,

$$P(X = c) = \int_c^c f(x)dx = 0.$$

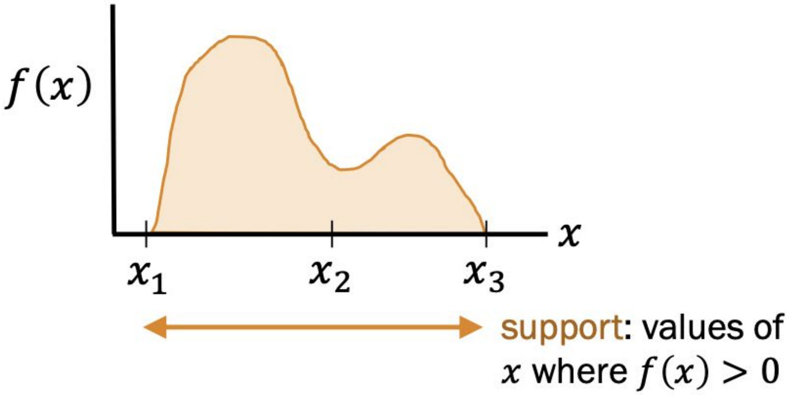


Contrast with PMF in discrete case: $P(X = c) = p(c)$

PDF Properties

For a continuous RV X with PDF f,

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



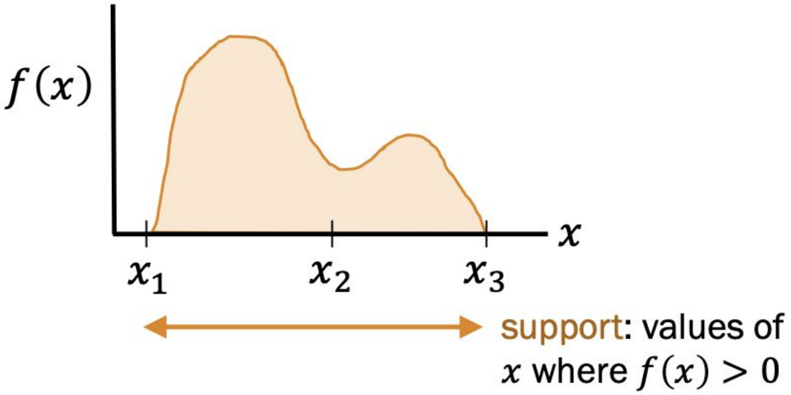
True/False:

1. $P(X = c) = 0$
2. $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$
3. $f(x)$ is a probability
4. In the graphed PDF above, $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

PDF Properties

For a continuous RV X with PDF f,

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



Interval width $dx \to 0$

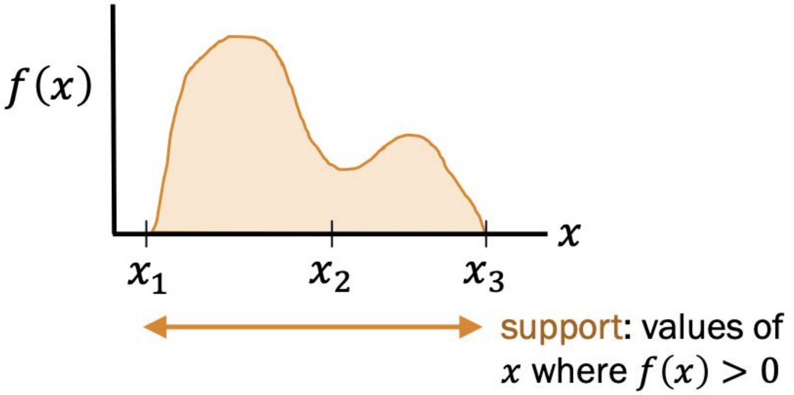
True/False:

1. $P(X = c) = 0$
2. $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$
3. $f(x)$ is a probability
4. In the graphed PDF above, $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

PDF Properties

For a continuous RV X with PDF f,

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



True/False:

1. $P(X = c) = 0$

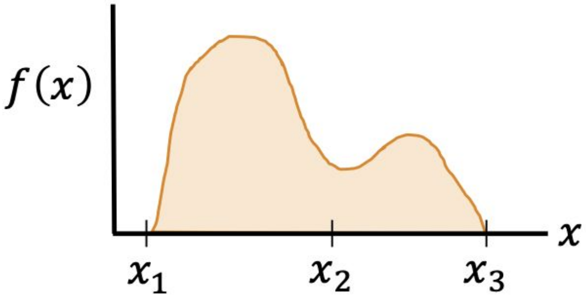
Interval width $dx \to 0$

2. $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$
3. $f(x)$ is a probability
4. In the graphed PDF above,
   $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

PDF Properties

For a continuous RV X with PDF f,

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



support: values of x where $f(x) > 0$

True/False:

1. $P(X = c) = 0$

   Interval width $dx \to 0$

2. $P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b)$

3. $f(x)$ is a probability

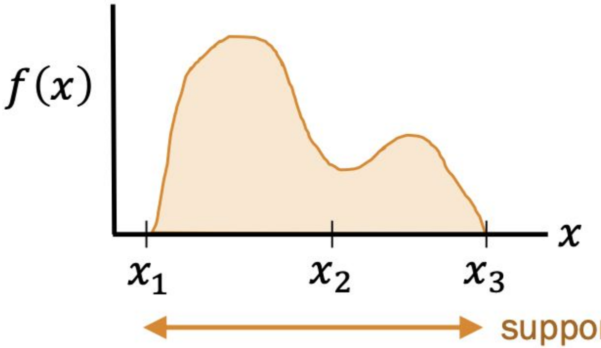
   It's a probability density!

4. In the graphed PDF above, $P(x_1 \le X \le x_2) > P(x_2 \le X \le x_3)$

PDF Properties

For a continuous RV X with PDF f,

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



support: values of x where $f(x) > 0$

True/False:

1. $P(X = c) = 0$ — Interval width $dx \to 0$

2. $P(a \le X \le b) = P(a < X < b) = P(a \le X < b) = P(a < X \le b)$

3. $f(x)$ is a probability — It's a probability density!

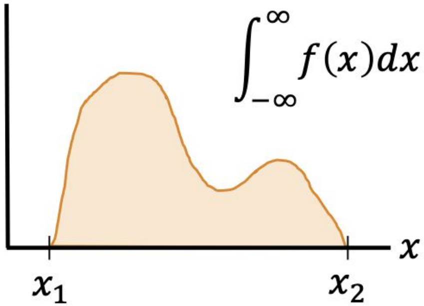
4. In the graphed PDF above, $P(x_1 \le X \le x_2) > P(x_2 \le X \le x_3)$ — Compare area under the curve

Determining valid PDFs

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

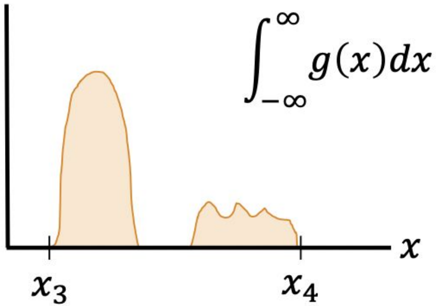
Which of the following functions are valid PDFs?

1. $f(x)$



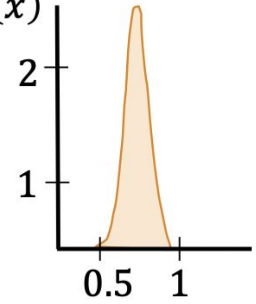
$$\int_{-\infty}^{\infty} f(x)dx = 0.5$$

2. $g(x)$



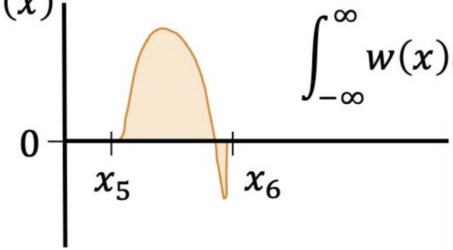
$$\int_{-\infty}^{\infty} g(x)dx = 1$$

3. $h(x)$



$$\int_{-\infty}^{\infty} h(x)dx = 1$$

4. $w(x)$



$$\int_{-\infty}^{\infty} w(x)dx = 1$$

Expected Value & Variance

Introduction to Continuous RVs

- Probability Density Functions (PDF)
- **Expected Value and Variance of Continuous RV**
- Cumulative Distribution Functions (CDF)

Discrete vs Continuous Variables:

Expectation and Variance

Discrete RV X

$$E[X] = \sum_x x\,p(x)$$

$$E[g(X)] = \sum_x g(x)\,p(x)$$

Continuous RV X

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

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\ dx$$

Both continuous and discrete RVs

$$E[aX + b] = aE[X] + b$$

Linearity of Expectation

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

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

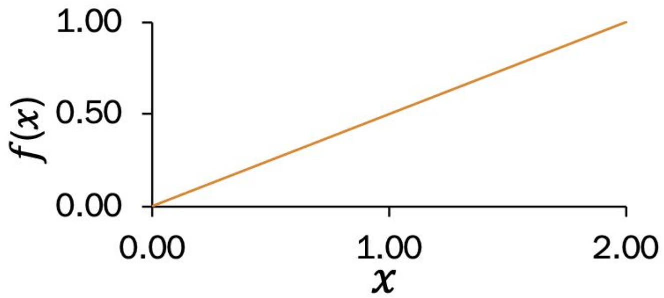
Properties of variance

TL;DR: $\sum_{x=a}^{b} \Rightarrow \int_a^b$

Expected Value & Variance of Continuous Random Variables

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$



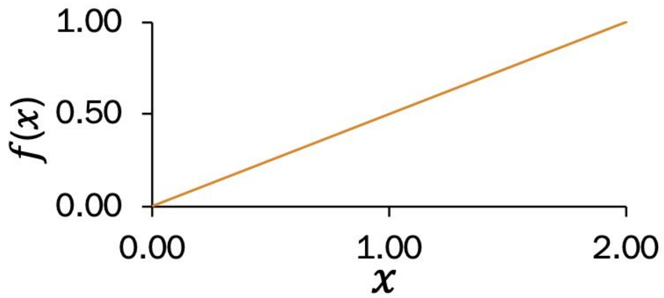
a). What is E[X]?

b). What is Var[X]?

Expected Value & Variance of Continuous Random Variables

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$



a). What is E[X]?

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

$$= \int_0^2 x \frac{x}{2}\, dx \qquad = \int_0^2 \frac{x^2}{2}\, dx$$

$$= \frac{x^3}{6}\Big|_0^2 = \frac{2^3}{6} - \frac{0^3}{6} = \frac{8}{6} = \boxed{\frac{4}{3}}$$

b). What is Var[X]?

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

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

$$= \int_0^2 x^2 \frac{x}{2}\, dx = \int_0^2 \frac{x^3}{2}\, dx$$

$$= \frac{x^4}{8}\Big|_0^2 = \frac{2^4}{8} - \frac{0^4}{7} = \frac{16}{8} = 2$$

$$Var[X] = 2 - (\tfrac{4}{3})^2 = \boxed{\frac{2}{9}}$$

CDFs

Introduction to Continuous RVs

Probability Density Functions (PDF)

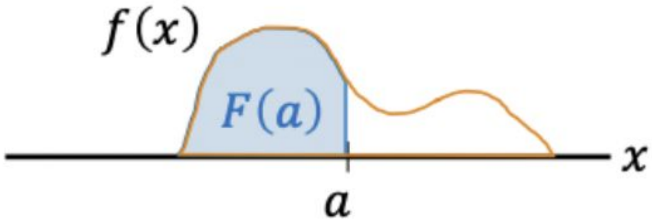
- Expected Value and Variance of Continuous RV

Cumulative Distribution Functions (CDF)

Cumulative Distribution Functions (CDFs)

For a continuous random variable X with PDF $f(x)$, the CDF of X is

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

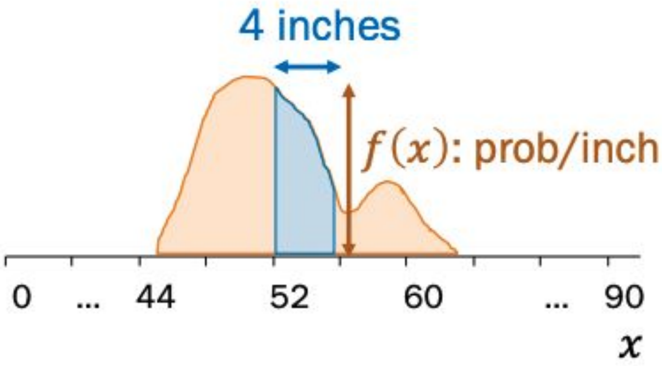


CDF is a probability, though PDF is not.

If you learn to use CDFs, you can avoid integrating the PDF.

Addendum to main takeaway #1

Integrate $f(x)$ to get probabilities.*



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

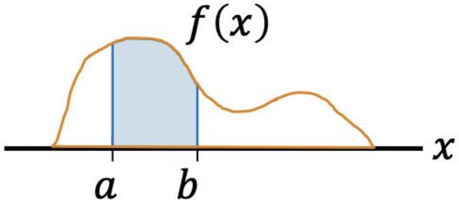
*If you have $F(a)$, you already have probabilities, since $F(a) = \int_{-\infty}^{a} f(x)dx$

Cumulative Distribution Functions (CDFs)

For a continuous random variable X with PDF $f(x)$, the CDF of X is

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

THUS: $P(a \leq X \leq b) =$

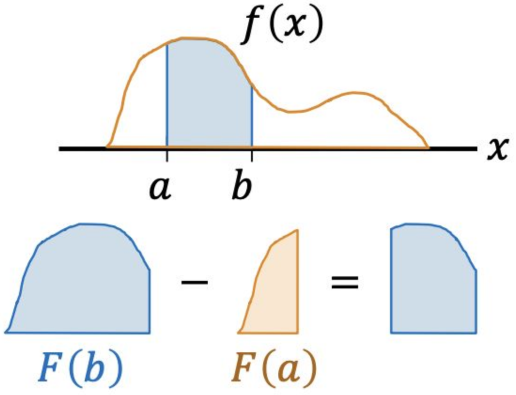


Cumulative Distribution Functions (CDFs)

For a continuous random variable X with PDF $f(x)$, the CDF of X is

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

$$P(a \leq X \leq b) = F(b) - F(a)$$



Proof:

$$F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$

$$= \left(\int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx\right) - \int_{-\infty}^{a} f(x)dx$$

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

Cumulative Distribution Functions (CDFs)

For a continuous random variable X with PDF $f(x)$, the CDF of X is

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

Matching (choices are used 0/1/2 times)

1. $P(X < a)$
2. $P(X > a)$
3. $P(X \geq a)$
4. $P(a \leq X \leq b)$

A. $F(a)$

B. $1 - F(a)$

C. $F(b) - F(a)$

D. $F(a) - F(b)$

Cumulative Distribution Functions (CDFs)

For a continuous random variable X with PDF $f(x)$, the CDF of X is

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

Matching (choices are used 0/1/2 times)

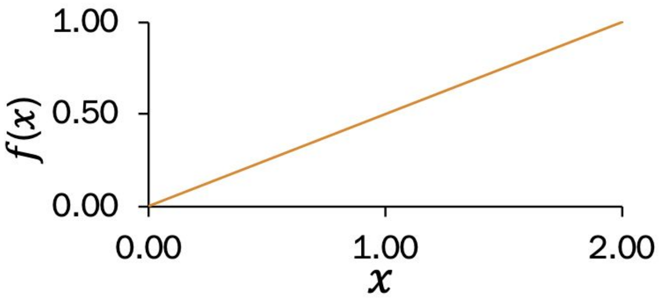
1. $P(X < a)$
2. $P(X > a)$
3. $P(X \geq a)$
4. $P(a \leq X \leq b)$

A. $F(a)$
B. $1 - F(a)$
C. $F(b) - F(a)$ (next slide)
D. $F(a) - F(b)$

CDF Continuous Random Variables

Let X be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$



c). What is the CDF of X?

PMF, PDF and CDFs in Python

```
from scipy import stats

X=stats.distribution
```

Distribution Names and Reference Pages in Scipy:

[bernoulli]()	[norm]()
[binom]()	[expon]()
[poisson]()	[uniform]()

Function	Defined for…	Description
`X.pmf(k, params)`	DISCRETE RV (Bernoulli, Binom, etc))	$P(X = k)$
`X.pdf(k, params)`	CONTINUOUS RV (Uniform, Expon, Norm, etc)	Probability Density, i.e. probability ***per unit x***, when X=k.
`X.cdf(k, params)`	BOTH Discrete and Continuous	$P(X \leq k)$