Continuous random variables

Andrew Thangaraj

IIT Madras

Subsection 1

Introduction

Why continuous random variables?

- Consider a discrete random variable X taking values in an alphabet $\mathcal{X}$
- Suppose that $|\mathcal{X}|$ is growing very large and unwieldy for calculations

Why continuous random variables?

- Consider a discrete random variable X taking values in an alphabet $\mathcal{X}$
- Suppose that $|\mathcal{X}|$ is growing very large and unwieldy for calculations
- Example 1
  - Met~~oe~~ eo rite weight
    - 0.01 grams to 60 tons
    - Data available for 45000+ meteorites
    - Weights spread out over a large range

Why continuous random variables?

- Consider a discrete random variable X taking values in an alphabet $\mathcal{X}$
- Suppose that $|\mathcal{X}|$ is growing very large and unwieldy for calculations

- Example 1
  - Metoerite weight
    - 0.01 grams to 60 tons
    - Data available for 45000+ meteorites
    - Weights spread out over a large range

- Example 2
  - Binomial(n, p)
    - n grows very large
    - p is a constant
    - can happen in Bernoulli trials

Why continuous random variables?

- Consider a discrete random variable X taking values in an alphabet $\mathcal{X}$
- Suppose that $|\mathcal{X}|$ is growing very large and unwieldy for calculations
- Example 1
  - Metoerite weight
    - 0.01 grams to 60 tons
    - Data available for 45000+ meteorites
    - Weights spread out over a large range
- Example 2
  - Binomial(n, p)
    - n grows very large
    - p is a constant
    - can happen in Bernoulli trials
- Is there a way to simplify the descriptions of these random-like phenomena?
  - Yes! But we need to give up something...

Meteorite data

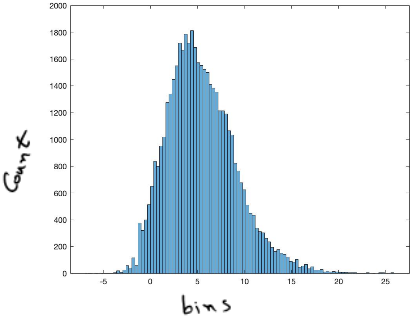
- Preprocessing: Take logarithms ($\log_2$)
  - Range: from $[0.01, 60000000]$ to $[-6.6, 25.8]$
  - Still, we have 45000+ data

Meteorite data

- Preprocessing: Take logarithms ($\log_2$)
  - Range: from $[0.01, 60000000]$ to $[-6.6, 25.8]$
  - Still, we have 45000+ data
- Main idea: Move from individual values to intervals of values
  - Divide $[-6.6, 25.8]$ into ≈ 100 intervals
    - $[-6.6, -6.3], [-6.3, -6], \ldots, [25.5, 25.8]$
  - Count the number of values falling inside each interval

Meteorite data

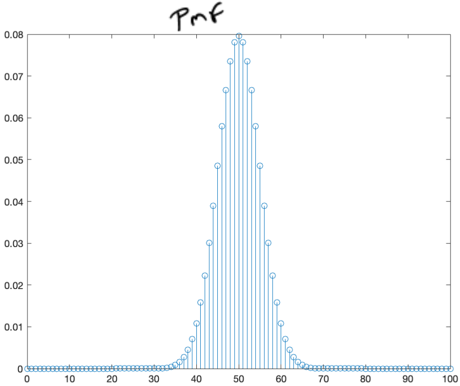
- Preprocessing: Take logarithms ($\log_2$)
  - Range: from $[0.01, 60000000]$ to $[-6.6, 25.8]$
  - Still, we have 45000+ data
- Main idea: Move from individual values to intervals of values
  - Divide $[-6.6, 25.8]$ into ≈ 100 intervals
    - ★ $[-6.6, -6.3], [-6.3, -6], \ldots, [25.5, 25.8]$
  - Count the number of values falling inside each interval



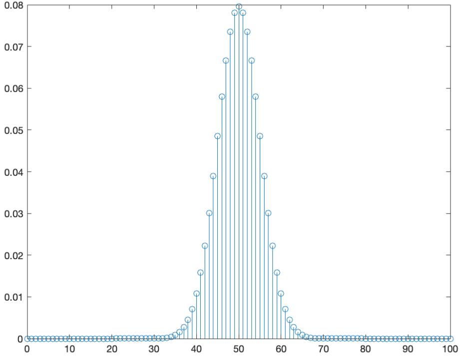
Histogram of log of weights

- Describe shape of the histogram instead of individual values
- What do we give up? Precision is reduced

Bernoulli trials, $n = 100$, $p = 0.5$



Bernoulli trials, $n = 100$, $p = 0.5$



- Calculations with the PMF is not very easy even in this case
- Working with intervals and histograms can be much simpler

Subsection 2

~~Cumulative~~ distribution function

CDF of a random variable

Definition (CDF of a random variable)

The Cumulative Distribution Function (CDF) of a random variable X, denoted $F_X(x)$, is a function from $\mathbb{R}$ to $[0,1]$, defined as

$$F_X(x) = P(X \leq x).$$

CDF of a random variable

Definition (CDF of a random variable)

The Cumulative Distribution Function (CDF) of a random variable X, denoted $F_X(x)$, is a function from $\mathbb{R}$ to $[0, 1]$, defined as

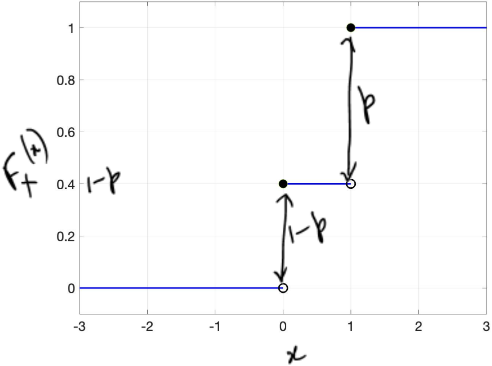
$$F_X(x) = P(X \leq x).$$

Properties

- $F_X(b) - F_X(a) = P(a < X \leq b)$
- F_X: non-decreasing function taking non-negative values
- As $x \to -\infty$, F_X goes to 0.
- As $x \to \infty$, F_X goes to 1.

Example: Bernoulli random variable

$$X \sim \{\overset{1-p}{\underline{\underline{0}}}, \overset{p}{\underline{\underline{1}}}\}$$



$$F_X(x) = \begin{cases} 0 & x < 0 \checkmark \\ \underline{1-p} & 0 \le x < 1 \checkmark \\ 1 & 1 \le x \checkmark \end{cases}$$

$F_X(-1) = P(X \le -1) = 0$

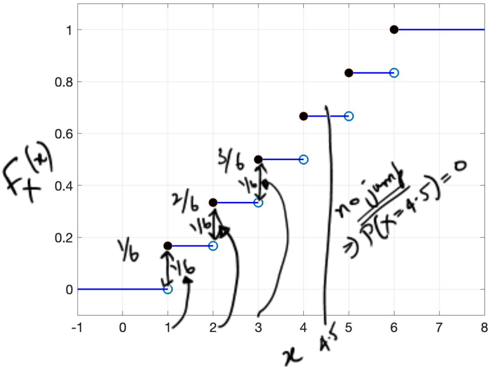
$F_X(x) = P(X \le \underline{\underline{0.1}}) = P(X = 0) = 1-p$

$x = 0.1$ any x: $0 \le x \underline{\underline{\le}} 1$

Example: Throw a die

$$X \sim \left\{ \overset{1/6}{1}, \overset{1/6}{2}, \overset{1/6}{3}, \overset{1/6}{4}, \overset{1/6}{5}, \overset{1/6}{6} \right\}$$ PMF

CDF



$$F_X(x) = \begin{cases} 0 & x < 1 \\ 1/6 & 1 \le x < 2 \\ 2/6 & 2 \le x < 3 \\ 3/6 & 3 \le \underline{x} < 4 \\ 4/6 & 4 \le x < 5 \\ 5/6 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$$

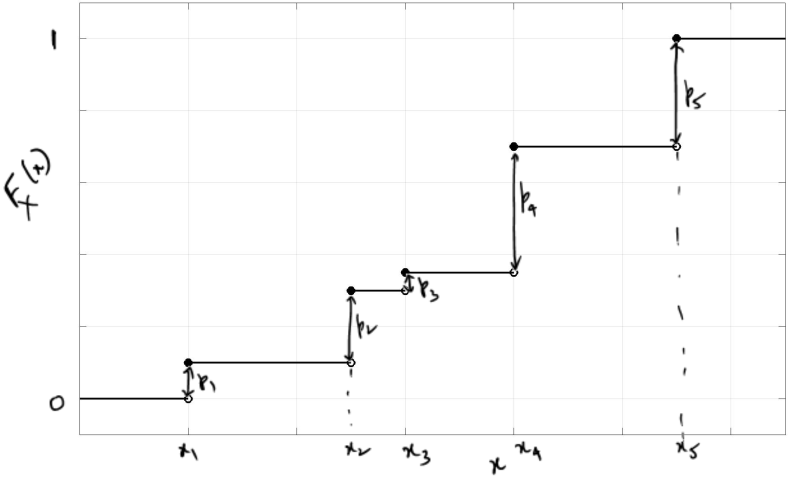
$X \in \{1, 2, 3\}$

CDF of a discrete random variable

$$X \sim \{\overset{p_1}{x_1}, \overset{p_2}{x_2}, \overset{p_3}{x_3}, \overset{p_4}{x_4}, \overset{p_5}{x_5}\}$$

CDF of a discrete random variable

$$X \sim \{\overset{p_1}{x_1}, \overset{p_2}{x_2}, \overset{p_3}{x_3}, \overset{p_4}{x_4}, \overset{p_5}{x_5}\}$$



Computing probability of intervals using CDF

$$X \sim \text{Uniform}\{1, 2, \ldots, 100\}$$

$\frac{1}{100}$ $1 \leq x < 2$

$\frac{2}{100}$ $2 \leq x < 3$

$\vdots$

$$F_X(x) = \begin{cases} 0 & \cancel{x \leq 0} \; x < 1 \\ k/100 & k \leq x < k+1, k = 1, 2, \ldots, 99 \\ 1 & x \geq 100 \end{cases}$$

Computing probability of intervals using CDF

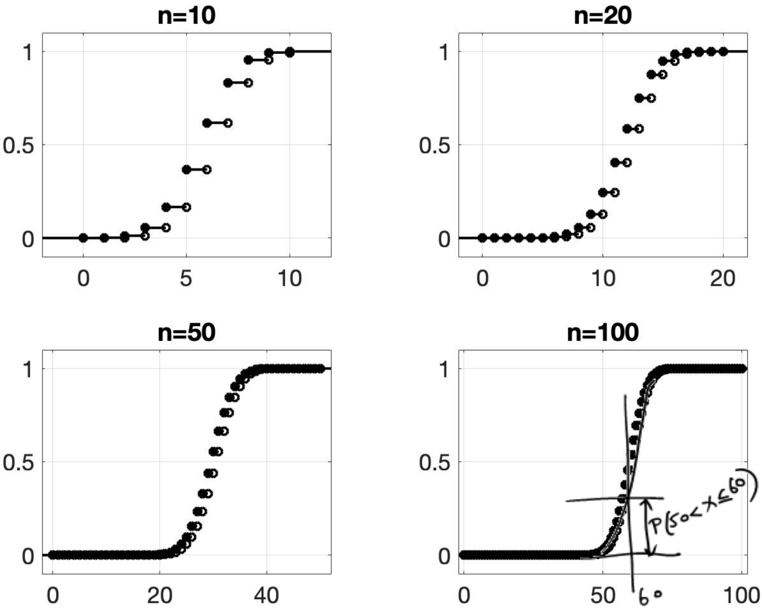
$$X \sim \text{Uniform}\{1, 2, \ldots, 100\}$$

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ k/100 & k \leq x < k+1, k = 1, 2, \ldots, 99 \\ 1 & x \geq 100 \end{cases}$$

- $P(3 < X \leq 10) = \overset{10/100}{F_X(10)} - \overset{3/100}{F_X(3)} = 7/100$
- $P(3.2 < X \leq 10.6) = \overset{10/100}{F_X(10.6)} - \overset{3/100}{F_X(3.2)} = 7/100$
- $P(X \leq 17) = F_X(17) = 17/100$
- $P(X \leq 17.3) = F_X(17.3) = 17/100$
- $P(\underline{X > 87}) = 1 - \underline{F_X(87)} = 13/100$
- $P(X > 87.4) = 1 - F_X(87.4) = 13/100$

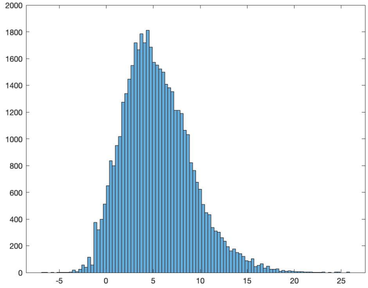
$P(X > 87) = 1 - P(X > 87)$
$= 1 - P(X \leq 87)$
$= 1 - F_X(87)$

Large alphabet: Binomial(n, 0.6)

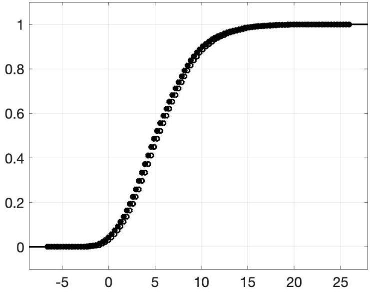


Large alphabet: Meteorite data

Histogram of weights



CDF of weight distribution



Cumulative Distribution Functions

Definition (CDF)

A function $F : \mathbb{R} \to [0, 1]$ is said to be a Cumulative Distribution Function (CDF) if

1. F is a non-decreasing function taking values between 0 and 1.
2. As $x \to -\infty$, F[redacted] goes to 0.
3. As $x \to \infty$, F[redacted] goes to 1.
4. Technical: F is continuous from the right.

Cumulative Distribution Functions

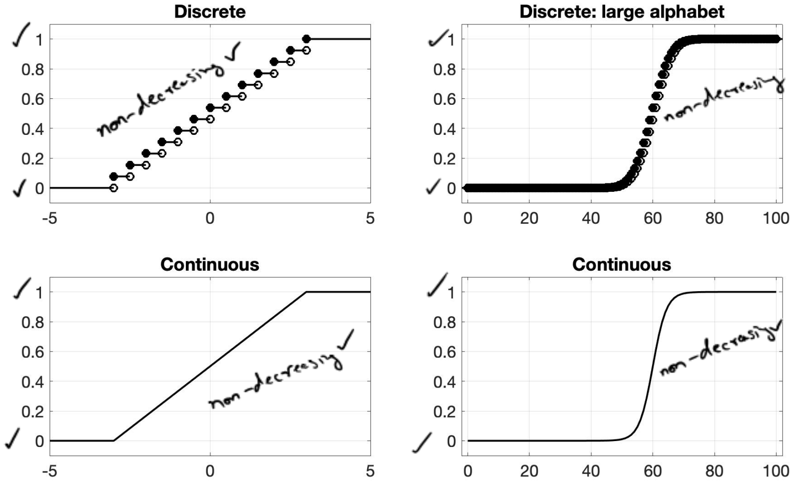
Definition (CDF)

A function $F : \mathbb{R} \to [0, 1]$ is said to be a Cumulative Distribution Function (CDF) if

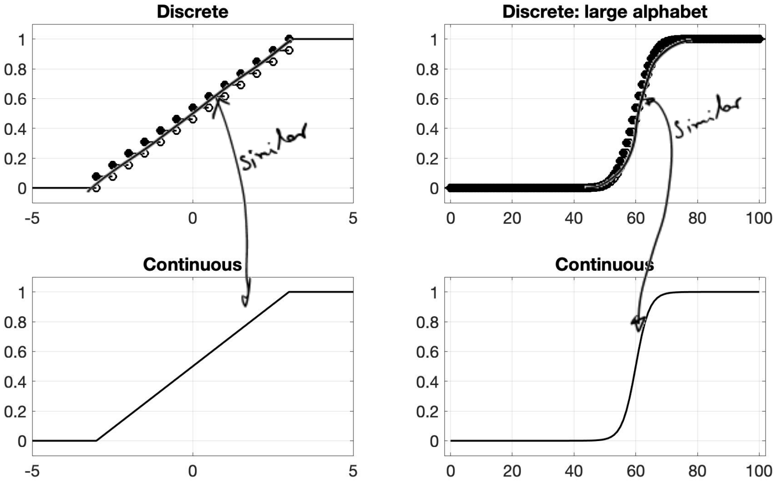
1. F is a non-decreasing function taking values between 0 and 1.
2. As $x \to -\infty$, F_x goes to 0.
3. As $x \to \infty$, F_x goes to 1.
4. Technical: F is continuous from the right.

- Definition motivated by CDF of a random variable defined earlier
- A general CDF need not be like a CDF of a discrete random variable
  - No need for a step-like structure
  - Can be smooth and continuous

Examples of valid CDFs



Examples of valid CDFs



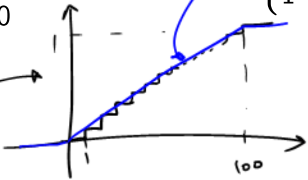
- Continuous CDFs appear to be close approximations to CDFs of discrete random variables, particularly when alphabet grows.

Probability of intervals using continuous CDF

$$X \sim \text{Uniform}\{1, 2, \ldots, 100\}$$

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ k/100 & k \le x < k+1 \\ & k = 1, \ldots, 99 \\ 1 & x \ge 100 \end{cases}$$

$$F(x) = \begin{cases} 0 & x \le 0 \\ x/100 & 0 \le x \le 100 \\ 1 & x \ge 100 \end{cases}$$



Probability of intervals using continuous CDF

$$X \sim \text{Uniform}\{1, 2, \ldots, 100\}$$

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ k/100 & k \leq x < k+1 \\ & k = 1, \ldots, 99 \\ 1 & x \geq 100 \end{cases} \qquad F(x) = \begin{cases} 0 & x \leq 0 \\ x/100 & 0 \leq x \leq 100 \\ 1 & x \geq 100 \end{cases}$$

- $P(3 < X \leq 10) = F_X(10) - F_X(3) = 7/100 = F(10) - F(3)$
- $P(3.2 < X \leq 10.6) = F_X(10.6) - F_X(3.2) = 7/100 \approx F(10.6) - F(3.2) = 7.4/100$
- $P(X \leq 17.3) = F_X(17.3) = 17/100 \approx F(17.3) = 17.3/100$

Probability of intervals using continuous CDF

$$X \sim \text{Uniform}\{1, 2, \ldots, 100\}$$

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ k/100 & k \le x < k+1 \\ & k = 1, \ldots, 99 \\ 1 & x \ge 100 \end{cases} \qquad F(x) = \begin{cases} 0 & x \le 0 \\ x/100 & 0 \le x \le 100 \\ 1 & x \ge 100 \end{cases}$$

- $P(3 < X \le 10) = F_X(10) - F_X(3) = 7/100 = F(10) - F(3)$
- $P(3.2 < X \le 10.6) = F_X(10.6) - F_X(3.2) = 7/100 \approx F(10.6) - F(3.2) = 7.4/100$
- $P(X \le 17.3) = F_X(17.3) = 17/100 \approx F(17.3) = 17.3/100$

Simpler continuous CDF approximates prob of intervals!

Binomial using continuous CDF

$$X \sim \text{Binomial}(100, 0.6)$$

$$F_X(k) = \sum_{j=0}^{k} \binom{100}{j} (0.6)^j (0.4)^{n-j}$$

$k = 0, 1, \ldots, 100$

$$F(x) = \frac{1}{1 + \exp\left(\frac{-1.65451(x - \overset{np}{60})}{\sqrt{\underset{np(1-p)}{24}}}\right)}$$

Binomial using continuous CDF

$$X \sim \text{Binomial}(100, 0.6)$$

$$F_X(k) = \sum_{j=0}^{k} \binom{100}{j} (0.6)^j (0.4)^{n-j} \qquad F(x) = \frac{1}{1 + \exp\left(\frac{-1.65451(x-60)}{\sqrt{24}}\right)}$$

- $P(40 < X \leq 50) = 0.0271,\ F(50) - F(40) = 0.0318$
- $P(50 < X \leq 60) = 0.5108,\ F(60) - F(50) = 0.4670$
- $P(60 < X \leq 70) = 0.4473,\ F(70) - F(60) = 0.4670$
- $P(70 < X \leq 80) = 0.0148,\ F(80) - F(70) = 0.0318$

Binomial using continuous CDF

$$X \sim \text{Binomial}(100, 0.6)$$

$$F_X(k) = \sum_{j=0}^{k} \binom{100}{j} (0.6)^j (0.4)^{n-j} \qquad F(x) = \frac{1}{1 + \exp\left(\frac{-1.65451(x-60)}{\sqrt{24}}\right)}$$

- $P(40 < X \leq 50) = 0.0271,\ F(50) - F(40) = 0.0318$
- $P(50 < X \leq 60) = 0.5108,\ F(60) - F(50) = 0.4670$
- $P(60 < X \leq 70) = 0.4473,\ F(70) - F(60) = 0.4670$
- $P(70 < X \leq 80) = 0.0148,\ F(80) - F(70) = 0.0318$

Better approximations are possible, partcularly as n grows!

Subsection 3

General random variables and Continuous random variables

CDFs and random variables

Theorem (Random variable with CDF $F(x)$)

Given a valid CDF $F(x)$, there exists a random variable X taking values in $\mathbb{R}$ such that

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

CDFs and random variables

Theorem (Random variable with CDF $F(x)$)

Given a valid CDF $F(x)$, there exists a random variable X taking values in $\mathbb{R}$ such that

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

Properties

- $P(a < X \leq b) = \underbrace{F(b)}_{}^{P(X\leq b)} - \overbrace{F(a)}^{P(X\leq a)}$

CDFs and random variables

Theorem (Random variable with CDF $F(x)$)

Given a valid CDF $F(x)$, there exists a random variable X taking values in $\mathbb{R}$ such that

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

Properties

- $P(a < X \leq b) = F(b) - F(a)$
- If $F(x)$ rises from F_1 to F_2 at x_1, $P(X = x_1) = F_2 - F_1$

CDFs and random variables

Theorem (Random variable with CDF $F(x)$)

Given a valid CDF $F(x)$, there exists a random variable X taking values in $\mathbb{R}$ such that

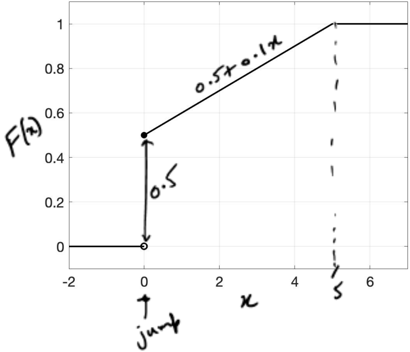
$$P(X \leq x) = F(x).$$

Properties

- $P(a < X \leq b) = F(b) - F(a)$
- If $F(x)$ rises from F_1 to F_2 at x_1, $P(X = x_1) = F_2 - F_1$
- If $F(x)$ is continuous at x_0, $P(X = x_0) = 0$ (non-intuitive!)

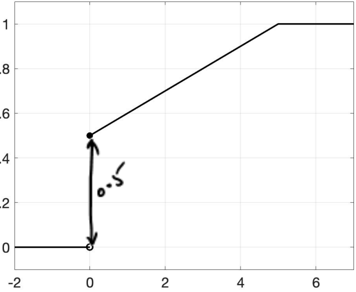
Example: Random variable X with CDF $F(x)$

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \le x \le 5 \\ 1 & x > 5 \end{cases}$$



Example: Random variable X with CDF $F(x)$

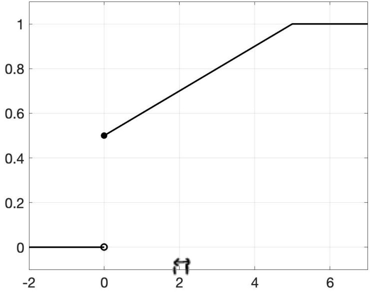
$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$

Example: Random variable X with CDF $F(x)$

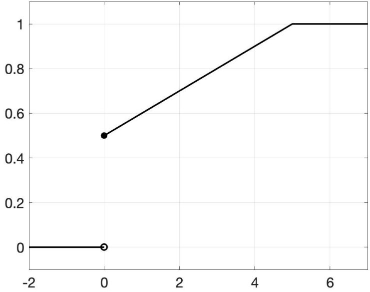
$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \le x \le 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$
- $P(1.99 < X \le 2.01) = F(2.01) - F(1.99) = 0.002$

Example: Random variable X with CDF $F(x)$

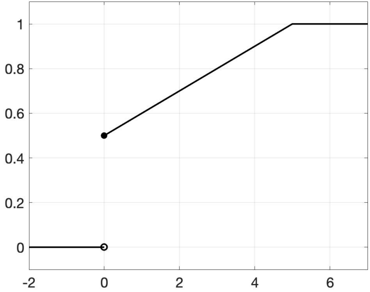
$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \le x \le 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$
- $P(1.99 < X \le 2.01) = F(2.01) - F(1.99) = 0.002$
  - Value with finite precision taken with positive probability

Example: Random variable X with CDF $F(x)$

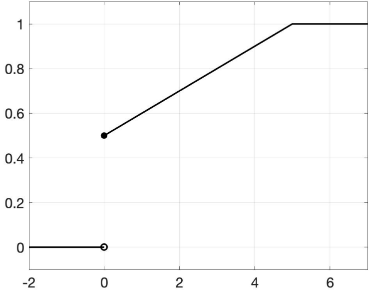
$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \le x \le 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$
- $P(1.99 < X \le 2.01) = F(2.01) - F(1.99) = 0.002$
  - Value with finite precision taken with positive probability
- $P(1.9999999 < X \le 2.0000001) = 0.00000002$

Example: Random variable X with CDF $F(x)$

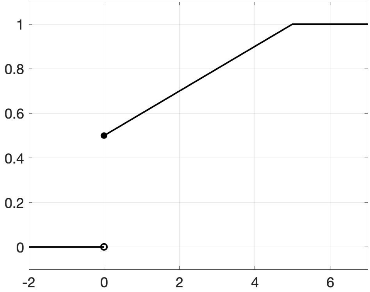
$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$
- $P(1.99 < X \leq 2.01) = F(2.01) - F(1.99) = 0.002$
  - Value with finite precision taken with positive probability
- $P(1.9999999 < X \leq 2.0000001) = 0.00000002$
  - As precision increases, probability decreases

Example: Random variable X with CDF $F(x)$

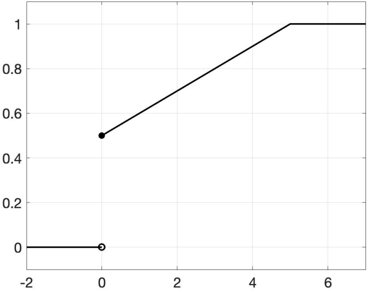
$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$
- $P(1.99 < X \leq 2.01) = F(2.01) - F(1.99) = 0.002$
  - Value with finite precision taken with positive probability
- $P(1.9999999 < X \leq 2.0000001) = 0.00000002$
  - As precision increases, probability decreases
- $P(X = 2.00000\cdots) = 0$

Example: Random variable X with CDF $F(x)$

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.1x & 0 \leq x \leq 5 \\ 1 & x > 5 \end{cases}$$



- $P(X = 0) = 0.5$
- $P(1.99 < X \leq 2.01) = F(2.01) - F(1.99) = 0.002$
  - Value with finite precision taken with positive probability
- $P(1.9999999 < X \leq 2.0000001) = 0.00000002$
  - As precision increases, probability decreases
- $P(X = 2.00000\cdots) = 0$
  - Values with infinte precision cannot be taken, when $F(x)$ is continuous at that point

Continuous random variable

Definition (Continuous random variable)

A random variable X with CDF $F_X(x)$ is said to be a continuous random variable if $F_X(x)$ is continuous at every x.

Continuous random variable

Definition (Continuous random variable)

A random variable X with CDF $F_X(x)$ is said to be a continuous random variable if $F_X(x)$ is continuous at every x.

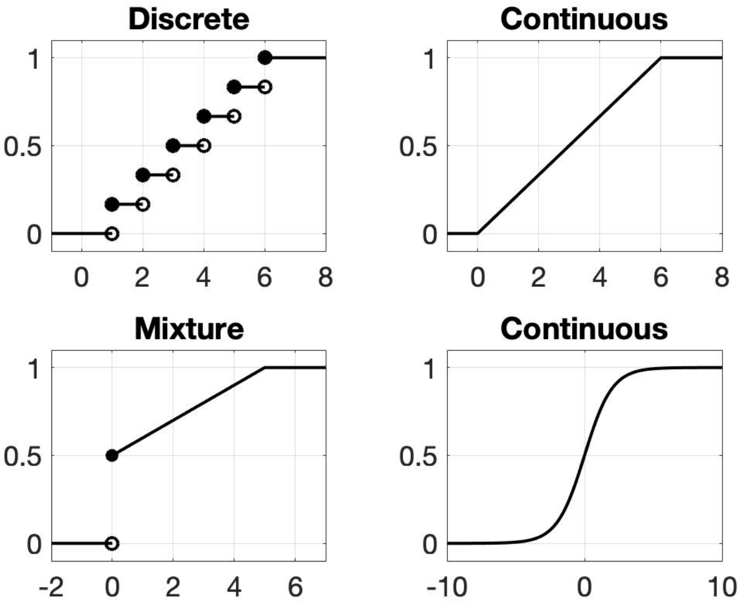
- CDF has no jumps or steps
- So, $P(X = x) = 0$ for all x
- Probability of X falling in an interval will be nonzero

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

- Since $P(X = a) = 0$ and $P(X = b) = 0$, we have

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Examples



Discrete
1
0.5
0
0
2
4
6
8
Continuous
1
0.5
0
0
2
4
6
8
Mixture
1
0.5
0
-2
0
2
4
6
Continuous
1
0.5
0
-10
0
10

Some scenarios for continuous models

- Throw a dart onto a circular board - distance of the point of impact from the center of the board.
- Weight of a metoerite hitting earth
- Weight of a human being, height of a human being
- Speed of a delivery in cricket
- Price of a stock
- Many discrete random variables are well-approximated by continuous random variables that are much simpler to describe

Problem

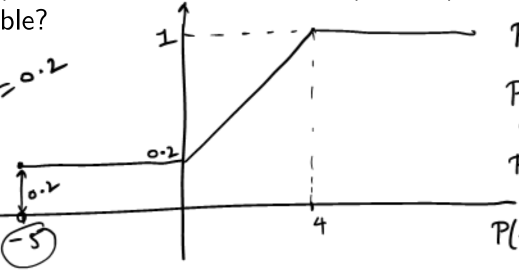
Consider a random variable X with CDF

$$F_X(x) = \begin{cases} 0 & x < -5 \\ 0.2 & -5 \leq x < 0 \\ 0.2 + 0.2x & 0 \leq x < 4 \\ 1 & x \geq 4. \end{cases}$$

Find $P(X < -3)$, $P(-3 < X < -1)$, $P(-1 < X < 1)$, $P(X \leq -3)$, $P(X > -2)$, $P(X \geq 3)$, $P(0 \leq X < 3)$. Is there an x_0 for which $P(X = x_0) > 0$? Is X a continuous random variable?

No

$P(X=-5)=0.2$



$P(X \leq -3) = F(-3)$
$= 0.2$

$P(X=-3) = 0$

$P(X<-3) = 0.2$

$P(-3<X<-1) = F(-1) - F(-3)$
$= 0$

$P(-1<X<1) = F(1) - F(-1)$
$= 0.4 - 0.2 = 0.2$

Working

$$P(X > -2) = 1 - P(X \le -2) = 1 - F(-2)$$
$$= 1 - 0.2 = 0.8$$

$$P(X \geq 3) = P(X > 3) = 1 - P(X \le 3) = 1 - F(3) = 1 - 0.8$$
$$= 0.2$$

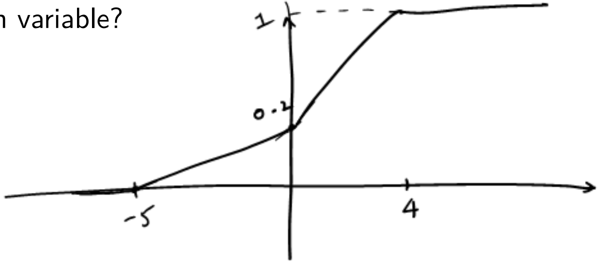
Problem

Consider a random variable X with CDF

$$F_X(x) = \begin{cases} 0 & x < -5 \\ 0.04x + 0.2 & -5 \le x < 0 \\ 0.2 + 0.2x & 0 \le x < 4 \\ 1 & x \ge 4. \end{cases}$$

Find $P(X < -3)$, $P(-3 < X < -1)$, $P(-1 < X < 1)$, $P(X \le -3)$, $P(0 \le X < 3)$. Is there an x_0 for which $P(X = x_0) > 0$? Is X a continuous random variable?

Yes



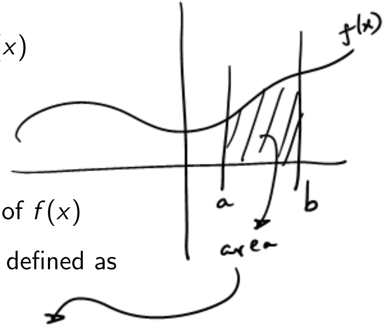
Working

Subsection 4

Probability density function and common continuous distributions

Refresher on integration

- Indefinite integral of a function $f(x)$
  - A function $F(x)$ such that $\dfrac{dF(x)}{dx} = f(x)$
  - Denoted as $F(x) = \int f(x)dx$
- Definite integral of a function $f(x)$
  - Suppose $F(x)$ is the indefinite integral of $f(x)$
  - Definite integral of $f(x)$ from a to b is defined as



$$\int_a^b f(x)dx = F(b) - F(a)$$

  - Definite integral equals the area under the curve $f(x)$ from a to b

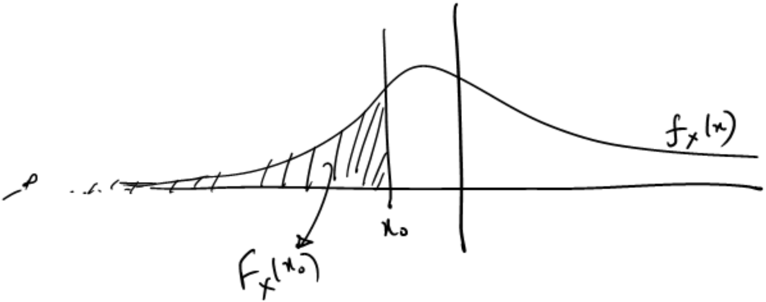
Tables of integrals: https://en.wikipedia.org/wiki/Lists_of_integrals

Probability density function (PDF)

Definition (PDF)

A continuous random variable X with CDF $F_X(x)$ is said to have a PDF $f_X(x)$ if, for all x_0,

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx.$$



Probability density function (PDF)

Definition (PDF)

A continuous random variable X with CDF $F_X(x)$ is said to have a PDF $f_X(x)$ if, for all x_0,

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx.$$

- CDF is the integral of the PDF
  - Derivative of the CDF (wherever it exists) is usually taken as the PDF

PDF: $f_X(x)$ Anti-derivative: $F_X(x)$

$$\int_{-\infty}^{x_0} f_X(x)\,dx = F_X(x_0) - \underbrace{F_X(-\infty)}_{=0}$$

Probability density function (PDF)

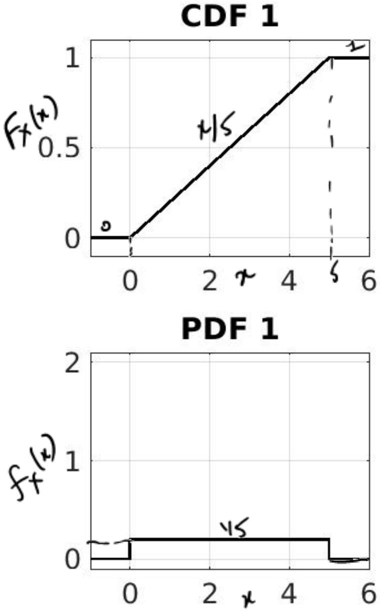
Definition (PDF)

A continuous random variable X with CDF $F_X(x)$ is said to have a PDF $f_X(x)$ if, for all x_0,

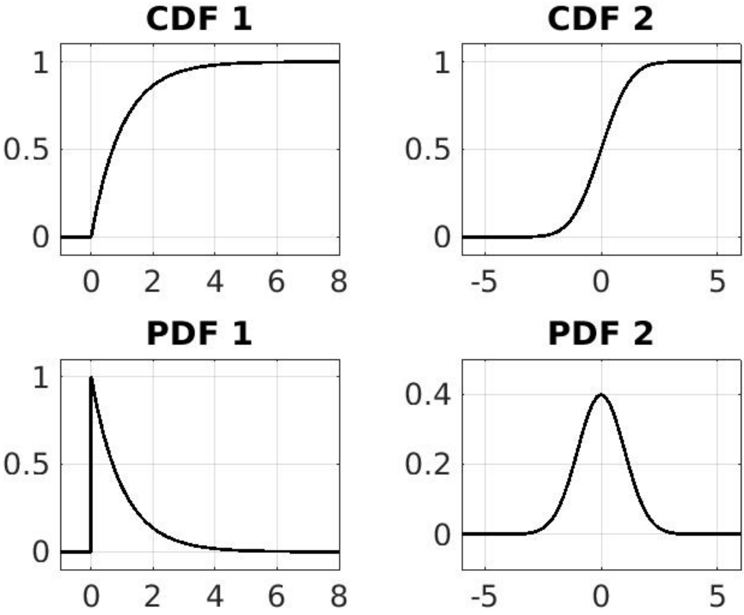
$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx.$$

- CDF is the integral of the PDF
  - Derivative of the CDF (wherever it exists) is usually taken as the PDF
- Why PDF?
  - Value of PDF around $f_X(x_0)$ is related to X taking a value around x_0
    - Higher the PDF, higher the chance that X lies there
    - Contrast with value of CDF at x_0, $F_X(x_0)$
    - PDF is much easier in probability computations

Examples: Uniform distribution



Examples: Exponential and normal distribution



Properties of PDF

Definition (Density function)

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be a density function if

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $f(x)$ is piecewise continuous

Properties of PDF

Definition (Density function)

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be a density function if

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $f(x)$ is piecewise continuous

- Given a density function f, there is a continuous random variable X with PDF as f
- Support of the random variable X with PDF f_X is $\text{supp}(X) = \{x : f_X(x) > 0\}$

  $\int_{\text{Supp}(x)} f_X(x)dx = 1$

  - $\text{supp}(X)$ contains intervals in which X can fall with positive probability
  - Remember: $P(X = x) = 0$ for a continuous random variable
- For a random variable X with PDF f_X, an event A is a subset of the real line and its probability is computed as $P(A) = \int_A f(x)dx$

Problem

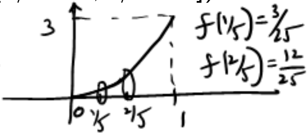
$f(x) \geq 0$

$\int_0^1 3x^2 \, dx = x^3 \Big|_0^1 = 1$

Consider the function

$$f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Show that f is a ✓ density function. Consider a random variable X with density f. Find $P(X = 1/5)$, $P(X = 2/5)$, $P(X \in [1/5 - \epsilon, 1/5 + \epsilon])$, $P(X \in [2/5 - \epsilon, 2/5 + \epsilon])$.



$P(X = 1/5) = 0$, $P(X = 2/5) = 0$

$P\left(\frac{1}{5} - \varepsilon < X < \frac{1}{5} + \varepsilon\right) = \int_{\frac{1}{5}-\varepsilon}^{\frac{1}{5}+\varepsilon} 3x^2 \, dx = x^3 \Big|_{1/5-\varepsilon}^{1/5+\varepsilon}$

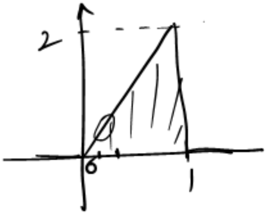
$= \left(\frac{1}{5} + \varepsilon\right)^3 - \left(\frac{1}{5} - \varepsilon\right)^3 = \frac{6}{25}\varepsilon + 2\varepsilon^3$ $\ll \varepsilon$

4 times

$P\left(\frac{2}{5} - \varepsilon < X < \frac{2}{5} + \varepsilon\right) = \left(\frac{2}{5} + \varepsilon\right)^3 - \left(\frac{2}{5} - \varepsilon\right)^3 = \frac{24}{25}\varepsilon + 2\varepsilon^3$ $\ll \varepsilon$

Problem

Consider a random variable X with density

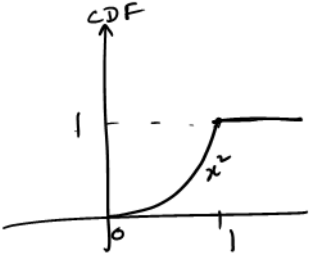


$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $P(X \in [0.1, 0.3])$, $P(X \in (0.1, 0.03])$, $P(X \in [0.1, 0.03))$, $P(X \in (0.1, 0.03))$.

$$P(0.1 \le X \le 0.3) = \int_{0.1}^{0.3} 2x\,dx = x^2 \Big|_{0.1}^{0.3} = 0.09 - 0.01 = 0.08$$

$$P(0.8 \le X \le 1) = 1^2 - 0.8^2 = 0.36$$

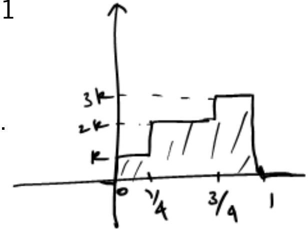


Problem

Consider the function

$$f(x) = \begin{cases} k & 0 \leq x < 1/4 \\ 2k & 1/4 \leq x < 3/4 \\ 3k & 3/4 \leq x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find k such that $f(x)$ is a valid density function.



$$\int_0^1 f(x)\,dx = 1$$

$$\int_0^{1/4} k\,dx + \int_{1/4}^{3/4} 2k\,dx + \int_{3/4}^{1} 3k\,dx = 1$$

$$\frac{1}{4}k + 2k \cdot \frac{1}{2} + 3k \cdot \frac{1}{4} = 1$$

$$k = 1/2$$

Common distributions: Uniform

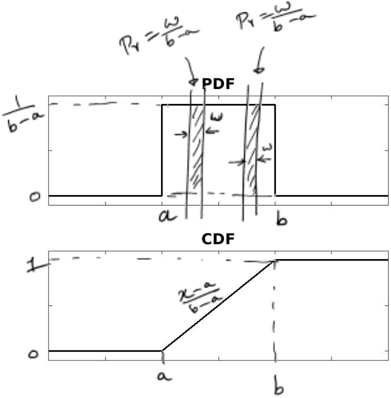
$$X \sim \text{Uniform}[a, b]$$

- PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

- CDF

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$



Probability computations with uniform distribution

Suppose $X \sim \text{Uniform}[-10, 10]$. Find $P(-3 \le X \le 2)$, $P(5 < |X| < 7)$, $P(1-\epsilon < X < 1+\epsilon)$, $P(9-\epsilon < X < 9+\epsilon)$, $P(X > 7 | X > 3)$.

$$f_X(x) = \begin{cases} \frac{1}{20}, & -10 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

$P(-3 \le X \le 2) = \frac{5}{20}$, $P(5 < |X| < 7) = P(5 < X < 7) + P(-7 < X < -5)$

$= \frac{2}{20} + \frac{2}{20} = \frac{4}{20}$

$$P(x_0 - \varepsilon < X < x_0 + \varepsilon) = \frac{2\varepsilon}{20}$$

x_0 inside $[-9, 9]$

ε: small

$$P(X > 7 | X > 3) = \frac{P(X > 7 \text{ and } X > 3)}{P(X > 3)} = \frac{P(X>7)}{P(X>3)} = \frac{3/20}{7/20} = \frac{3}{7}$$

f

3+4

0+4

$$P(X > 4 | X > 0) = \frac{P(X>4)}{P(X>0)} = \frac{6/20}{10/20} = \frac{3}{5}$$

Common distributions: Exponential

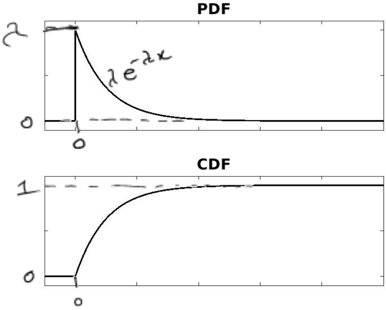
$$X \sim \text{Exp}(\lambda)$$

- PDF

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- CDF

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp(-\lambda x) & x > 0 \end{cases}$$



$$F_X(x_0) = \int_0^{x_0} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{x_0}$$
$$= 1 - e^{-\lambda x_0}$$

Probability computations with exponential distribution

Suppose $X \sim \text{Exp}(2)$. Find $P(5 < X < 7)$, $P(1-\epsilon < X < 1+\epsilon)$, $P(9-\epsilon < X < 9+\epsilon)$, $P(X > 4)$, $P(X > 7 | X > 3)$.

$$f_X(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$$P(5 < X < 7) = \int_5^7 2e^{-2x}\,dx = -e^{-2x}\Big|_5^7 = e^{-10} - e^{-14}$$

$e \approx 2.7\ldots$

$$P(1-\varepsilon < X < 1+\varepsilon) = e^{-2(1-\varepsilon)} - e^{-2(1+\varepsilon)} = e^{-2}\left(e^{2\varepsilon} - e^{-2\varepsilon}\right)$$

$$P(9-\varepsilon < X < 9+\varepsilon) = e^{-2(9-\varepsilon)} - e^{-2(9+\varepsilon)} = e^{-18}\left(e^{2\varepsilon} - e^{-2\varepsilon}\right) <<< P(1-\varepsilon < X < 1+\varepsilon)$$

$$P(X > 4) = 1 - P(X \le 4) = 1 - (1 - e^{-8}) = e^{-8}$$

$$P(X > 7 | X > 3) = \frac{P(X > 7)}{P(X > 3)} = \frac{e^{-14}}{e^{-6}} = e^{-8}$$

Unique to exponential distribution $\rightarrow$ $P(X > s+t | X > s) = e^{-t}$ independent of 's'

"memoryless"

Common distributions: Normal (or) Gaussian

μ : any real number
σ : positive real number

$$X \sim \text{Normal}(\mu, \sigma^2)$$

- PDF $\cdot \text{Supp}(X) = \mathbb{R}$

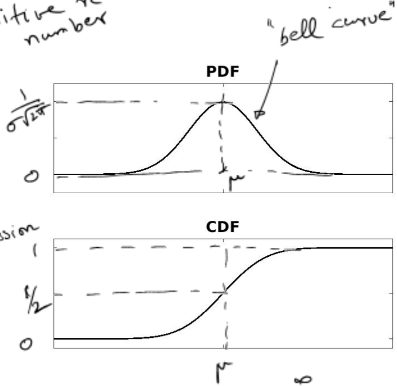
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- CDF   No closed form expression

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$$

- Standard normal: Normal$(0, 1)$

$\mu \quad \sigma$



Exercise: Check $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

$P(X<\mu) = P(X>\mu) = 1/2$

Probability computations with normal distribution

- CDF of $X \sim \text{Normal}(\mu, \sigma^2)$ does not have a closed form expression
- Standardization: If $X \sim \text{Normal}(\mu, \sigma^2)$, then
$Z = (X - \mu)/\sigma \sim \text{Normal}(0, 1)$
  - $Z \sim \text{Normal}(0, 1)$, PDF: $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$

$$\text{CDF: } F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du$$

  - Normal table: Tabulation of above function $F_Z(z)$
    - ★ Available on most computing systems
- How to compute probabilities for a normal distribution?
  - Convert probability computation to that of a standard normal
  - Use normal tables or computing systems

Problem

$\mu=2, \sigma^2=5$

Suppose $X \sim \text{Normal}(2,5)$. Find $P(X<5)$, $P(X<10)$, $P(X<-5)$, $P(X<-10)$, $P(X>5)$, $P(X>10)$.

$$Z = \frac{X-2}{\sqrt{5}} \sim \text{Normal}(0,1) \qquad X = \sqrt{5}\,Z + 2$$

Assume: $F_Z(z)$ (CDF of Z) is known

$$X<5 \Leftrightarrow \sqrt{5}\,Z+2<5 \Leftrightarrow Z < 3/\sqrt{5}$$

$$P(X<5) = P(Z<3/\sqrt{5}) = F_Z(3/\sqrt{5})$$

$$X>5 \Leftrightarrow \sqrt{5}\,Z+2>5 \Leftrightarrow Z > 3/\sqrt{5}$$

$$P(X>5) = P(Z>3/\sqrt{5}) = 1 - F_Z(3/\sqrt{5})$$

Problem

$\mu = 3, \sigma = 1$

Suppose $X \sim \text{Normal}(3, 1)$. Find $P(5 < X < 7)$, $P(-5 < X < 5)$, $P(1 - \epsilon < X < 1 + \epsilon)$, $P(9 - \epsilon < X < 9 + \epsilon)$, $P(X > 4)$, $P(X > 7 | X > 3)$.

$Z = X - 3 \sim \text{Normal}(0,1)$ $\qquad X = Z + 3$

$5 < X < 7 \Leftrightarrow 5 < Z + 3 < 7 \Leftrightarrow 2 < Z < 4$

$P(5 < X < 7) = P(2 < Z < 4) = F_Z(4) - F_Z(2)$

$$P(X > 7 / X > 3) = \frac{P(X > 7)}{P(X > 3)} = \frac{P(Z > 4)}{P(Z > 0)} = \frac{1 - F_Z(4)}{1/2}$$

$$= 2(1 - F_Z(4))$$

Subsection 5

Functions of a continuous random variable

Why functions?

- We may model one quantity as a random variable X. We may have to work with another closely related quantity
- Example 1
  - Length of a square: X
  - Area of the square: $Y = X^2$
- Example 2
  - Volume of a liquid: X
  - Density: ρ
  - Weight ~~Volume occupied~~: $Y = \rho X$
- Given the distribution of X, it is useful to have a method for finding the distribution of a function of X

Example

Suppose $X \sim$ Uniform$[0, 1]$

- $Y = 2X \in [0, 2]$ is clearly a random variable
- What is the distribution of Y?

Example

Suppose $X \sim$ Uniform$[0, 1]$

- $Y = 2X \in [0, 2]$ is clearly a random variable
- What is the distribution of Y?

For $y \in [0, 2]$, (Find CDF first)

$$F_Y(y) = P(Y \leq y) = P(2X \leq y)$$

$$= P(X \leq y/2) = \int_0^{\frac{y}{2}} \underbrace{f_X(x)}_{=1} dx = \frac{y}{2}.$$

Example

Suspose $X \sim$ Uniform$[0,1]$

- $Y = 2X \in [0,2]$ is clearly a random variable
- What is the distribution of Y?

For $y \in [0,2]$,

$$F_Y(y) = P(Y \le y) = P(2X \le y)$$

$$= P(X \le y/2) = \int_0^{\frac{y}{2}} f_X(x)dx = \frac{y}{2}.$$

PDF of Y, $f_Y(y) = \frac{dF_Y(Y)}{dy} = \frac{1}{2}$.

$$Y \sim \text{ Uniform}[0,2]$$

Example

$Y = X+5$?
(or) $Y = 2X+5$?

$Y = aX+b \sim$ Uniform$[b, b+a]$

Suppose $X \sim$ Uniform$[0,1]$

- $Y = 2X \in [0,2]$ is clearly a random variable
- What is the distribution of Y?

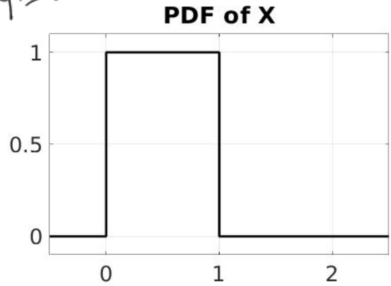
For $y \in [0,2]$,

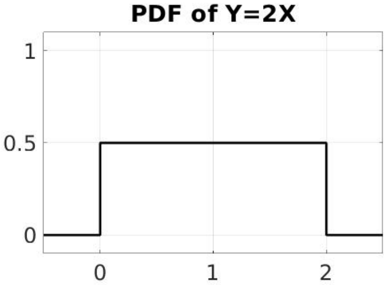
$$F_Y(y) = P(Y \leq y) = P(2X \leq y)$$

$$= P(X \leq y/2) = \int_0^{\frac{y}{2}} f_X(x)dx = \frac{y}{2}.$$

PDF of Y, $f_Y(y) = \frac{dF_Y(Y)}{dy} = \frac{1}{2}$.

$$Y \sim \text{Uniform}[0,2]$$





General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X

General case: CDF of $g(X)$

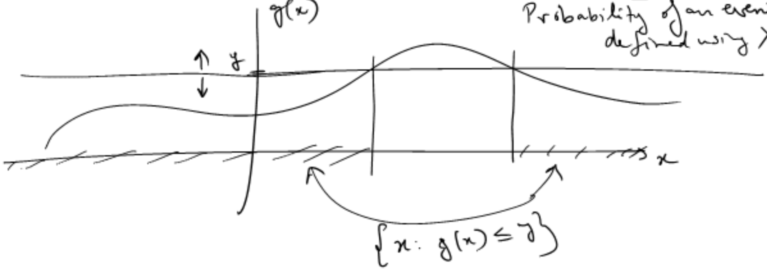
- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function

General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:

General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:
  - $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(\underbrace{X \in \{x : g(x) \leq y\}}_{\text{Probability of an event defined using } X})$



General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:
  - $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in \{x : g(x) \leq y\})$
- How to evaluate the above probability?

General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:
  - $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in \{x : g(x) \leq y\})$
- How to evaluate the above probability?
  - Convert the subset $A_y = \{x : g(x) \leq y\}$ into intervals in real line

General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:
  - $F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \in \{x : g(x) \le y\})$
- How to evaluate the above probability?
  - Convert the subset $A_y = \{x : g(x) \le y\}$ into intervals in real line
  - Find the probability that X falls in those intervals

General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:
    - $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in \{x : g(x) \leq y\})$
- How to evaluate the above probability?
    - Convert the subset $A_y = \{x : g(x) \leq y\}$ into intervals in real line
    - Find the probability that X falls in those intervals
    - $F_Y(y) = P(X \in A_y) = \int_{A_y} f_X(x)dx$

General case: CDF of $g(X)$

- Suppose X is a continuous random variable with CDF F_X and PDF f_X
- Suppose $g : \mathbb{R} \to \mathbb{R}$ is a (reasonable) function
- Then, $Y = g(X)$ is a random variable with CDF F_Y determined as follows:
  - $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in \{x : g(x) \leq y\})$
- How to evaluate the above probability?
  - Convert the subset $A_y = \{x : g(x) \leq y\}$ into intervals in real line
  - Find the probability that X falls in those intervals
  - $F_Y(y) = P(X \in A_y) = \int_{A_y} f_X(x)dx$
- If F_Y has no jumps, you may be able to differentiate and find a PDF

Monotonic, differentiable functions

Theorem

Suppose X is a continuous random variable with PDF f_X. Let $g(x)$ be monotonic for $x \in supp(X)$ with derivative $g'(x) = \frac{dg(x)}{dx}$. Then, the PDF of $Y = g(X)$ is

$$f_Y(y) = \underbrace{\frac{1}{|g'(g^{-1}(y))|}}_{g'(x),\ x = g^{-1}(y)} \underbrace{f_X(g^{-1}(y))}_{f_X(x),\ x = g^{-1}(y)}.$$

Monotonic, differentiable functions

Theorem

Suppose X is a continuous random variable with PDF f_X. Let $g(x)$ be monotonic for $x \in supp(x)$ with derivative $g'(x) = \frac{dg(x)}{dx}$. Then, the PDF of $Y = g(X)$ is

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)).$$

- Translation: $Y = \underbrace{X + a}$ $g(x) = x + a$, $g'(x) = 1$, $y = x + a$
  $x = y - a$
  $g^{-1}(y) = y - a$

$$f_Y(y) = f_X(y - a)$$

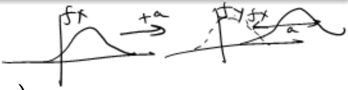
Monotonic, differentiable functions

Theorem

Suppose X is a continuous random variable with PDF f_X. Let $g(x)$ be monotonic for $x \in supp(X)$ with derivative $g'(x) = \frac{dg(x)}{dx}$. Then, the PDF of $Y = g(X)$ is

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)).$$

- Translation: $Y = X + a$



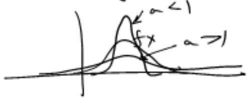
$$f_Y(y) = f_X(y - a)$$

- Scaling: $Y = aX$

$g(x) = ax$, $g'(x) = a$, $y = ax \Rightarrow x = y/a$

$g^{-1}(y) = y/a$

$$f_Y(y) = \frac{1}{|a|} f_X(y/a)$$



Monotonic, differentiable functions

Theorem

Suppose X is a continuous random variable with PDF f_X. Let $g(x)$ be monotonic for $x \in supp(X)$ with derivative $g'(x) = \frac{dg(x)}{dx}$. Then, the PDF of $Y = g(X)$ is

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)).$$

- Translation: $Y = X + a$

$$f_Y(y) = f_X(y - a)$$

- Scaling: $Y = aX$

$$f_Y(y) = \frac{1}{|a|} f_X(y/a)$$

- Affine: $Y = aX + b$

$$f_Y(y) = \frac{1}{|a|} f_X((y - b)/a)$$

Affine transformation of normal distributions

- $X \sim$ Normal$(0, 1)$ Standard normal

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

replace x with $\frac{y-\mu}{\sigma}$

  - $Y = \sigma X + \mu$ "Affine" $(\sigma > 0)$

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(y - \mu)^2/2\sigma^2)$$

: PDF of Normal(μ, σ^2)

$Y \sim$ Normal(μ, σ^2)

- $X \sim$ Normal(μ, σ^2) $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ (Exercise)

replace x with $\frac{y + \mu/\sigma}{1/\sigma} = \sigma y + \mu$

  - $Y = (X - \mu)/\sigma \sim \underline{\underline{\text{Normal}(0, 1)}}$

affine

$Y = \frac{1}{\sigma} X - \frac{\mu}{\sigma}$

Affine transformation of normal distributions

- $X \sim$ Normal$(0, 1)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

  - $Y = \sigma X + \mu$

$$f_Y(\cancel{x}\,y) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(\cancel{x}\,y - \mu)^2/2\sigma^2)$$

$Y \sim$ Normal(μ, σ^2)

- $X \sim$ Normal(μ, σ^2)
  - $Y = (X - \mu)/\sigma \ \sim$ Normal$(0, 1)$

Result

Affine transformation of a normal random variable is normal.

Problem

Let $X \sim \text{Exp}(\lambda)$. Find the PDF of X^2.

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

$$\text{Supp}(X) = \{x : x > 0\}$$

$Y = X^2$

$$f_Y(y) = \frac{1}{2\sqrt{y}} \lambda e^{-\lambda\sqrt{y}}, \quad y > 0$$

$$\frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y))$$

$g(x) = x^2$

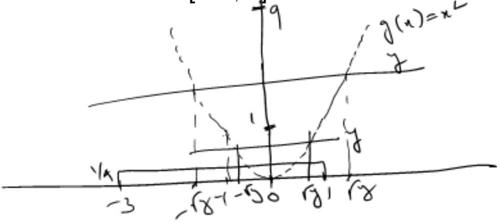
$g'(x) = 2x$

$y = x^2 \Rightarrow x = \sqrt{y}$

$g^{-1}(y) = \sqrt{y}$

Problem

Let $X \sim \text{Uniform}[-3,1]$. Find the PDF of X^2.



$Y = X^2 \in [0,9]$

$\text{supp}(X) = [-3,1]$ $\quad$ $g(x)=x^2$ is not monotonic in $\text{supp}(X)$

$y \in [0,9]$ $\quad F_Y(y) = P(Y \le y) = P(X^2 \le y)$

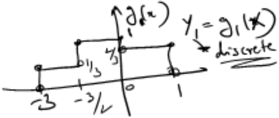
$\underline{y \in [0,1]}$: $(X^2 \le y) \Longleftrightarrow -\sqrt{y} < X < \sqrt{y} \Rightarrow F_Y(y) = \frac{2\sqrt{y}}{4}$

$y \in [1,9]$: $(X^2 \le y) \Longleftrightarrow -\sqrt{y} < X < 1 \Rightarrow F_Y(y) = \frac{1+\sqrt{y}}{4}$

$$f_Y(y) = \begin{cases} \frac{1}{2}\cdot\frac{1}{2\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{4}\cdot\frac{1}{2\sqrt{y}}, & 1 < y < 9 \end{cases}$$

Problem

Let $X \sim \text{Uniform}[-3, 1]$. Find the PDF of $\max(X, 0)$.



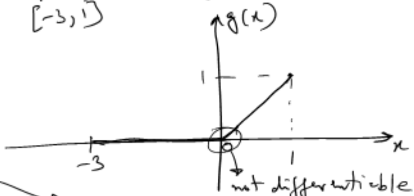
$Y = g(X) \in [0, 1]$

$$g(x) = \max(x, 0) = \begin{cases} 0, & \text{if } -3 < x < 0 \\ x, & \text{if } 0 < x < 1 \end{cases}$$

$x \in \text{Supp}(X)$
$[-3, 1]$

$y < 0$: $F_Y(y) = P(Y \leq \text{negative number}) = 0$

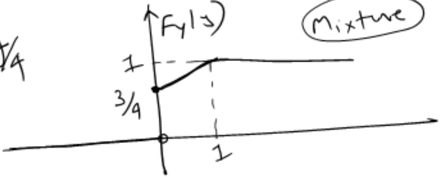
$y = 0$:
$F_Y(y) = P(Y \leq 0) = P(Y = 0)$
$= P(g(X) = 0)$
$= P(-3 < X < 0) = 3/4$



Y: not continuous

$0 < y < 1$: $F_Y(y) = \frac{3}{4} + y/4$
(exercise)

$y > 1$: $F_Y(y) = 1$



Mixture

Subsection 6

Continuous random variables: Expected value

Expected value: Function of a continuous random variable

Theorem

Let X be a continuous random variable with density $f_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a function. The expected value of $g(X)$, denoted $E[g(X)]$, is given by

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

whenever the above integral exists.

Expected value: Function of a continuous random variable

Theorem

Let X be a continuous random variable with density $f_X(x)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a function. The expected value of $g(X)$, denoted $E[g(X)]$, is given by

$$E[g(X)] = \underbrace{\int_{-\infty}^{\infty}}_{\text{supp}(X)} g(x) \underbrace{f_X(x)dx}_{\text{like PMF}},$$

whenever the above integral exists.

- If X is discrete with range T_X and PMF p_X,

$$E[g(X)] = \sum_{x \in T_X} g(x) p_X(x)$$

  - **Summation in discrete case is replaced by integration in coninuous case**

- The integral may diverge to $\pm\infty$ or may not exist in some cases

Mean and Variance

X: continuous random variable

- Mean, denoted $E[X]$ or μ_X or simply μ

$g(x) = x$

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

  - Mean is the average or expected value of X

- Variance, denoted $\text{Var}(X)$ or σ_X^2 or simply σ^2

$$\text{Var}(X) = E[\underbrace{(X - \mu_X)^2}_{g(x) = (x - \mu_X)^2}] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

  - Variance is a measure of spread of X about its mean
  - $\text{Var}(X) = E[X^2] - E[X]^2$

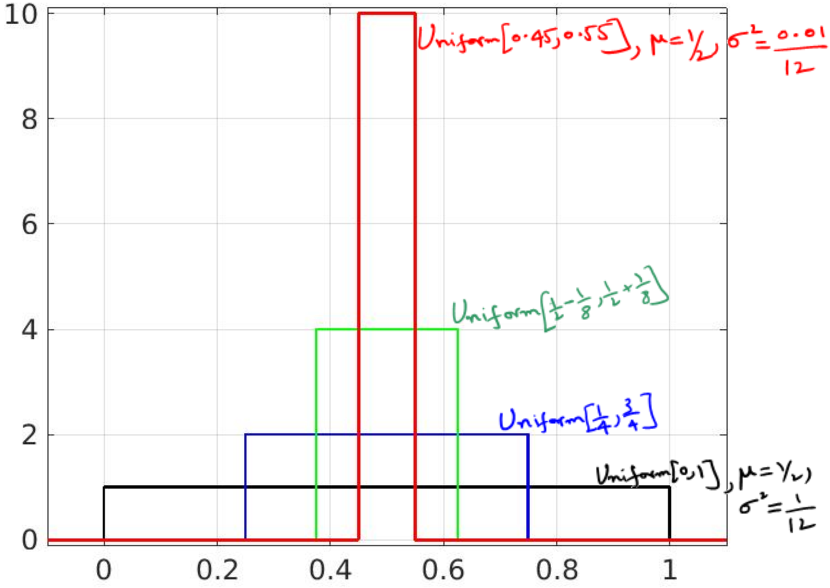
- Evaluating expected value needs good knowledge of integration
  - Formulae are available in numerous webpages and books

Examples of mean and variance

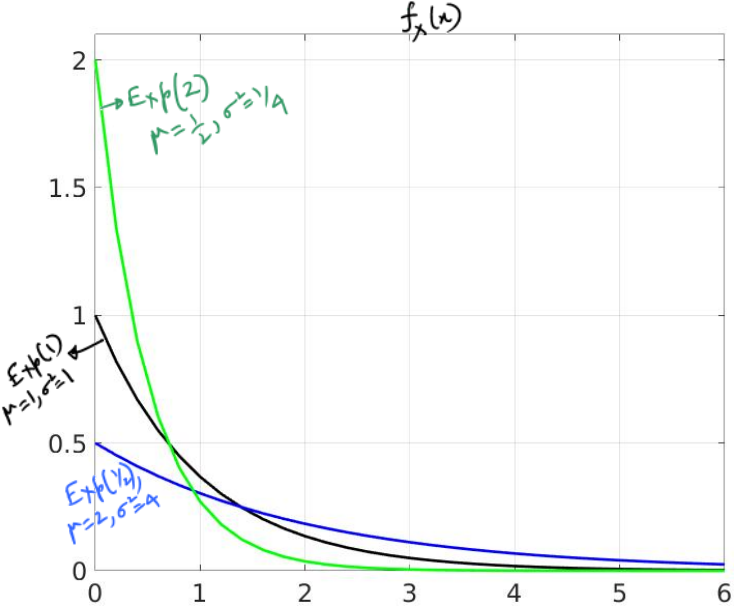
- $X \sim$ Uniform$[a, b]$, $f_X(x) = \dfrac{1}{b-a}$, $a < x < b$
  - $E[X] = \dfrac{a+b}{2}$, $\text{Var}(X) = \dfrac{(b-a)^2}{12}$
- $X \sim$ Exp(λ), $f_X(x) = \lambda \exp(-\lambda x)$, $x > 0$
  - $E[X] = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$
- $X \sim$ Normal(μ, σ^2), $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(x-\mu)^2/2\sigma^2)$
  - $E[X] = \mu$, $\text{Var}(X) = \sigma^2$

$$\left.\begin{aligned} \int_{-\infty}^{\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx &= \mu \\ \int_{-\infty}^{\infty} (x-\mu)^2 \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx &= \sigma^2 \end{aligned}\right\} \text{more involved integration}$$

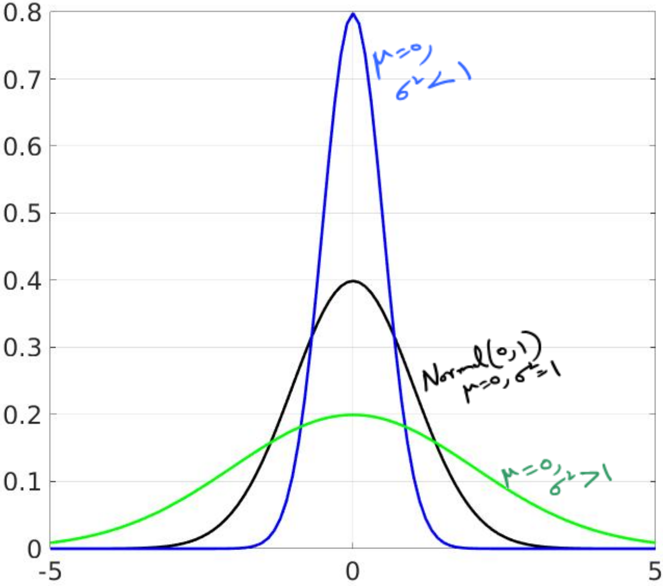
Uniform distribution with different variances



Exponential distribution with different λ



Normal distribution with different σ



Markov and Chebyshev inequalities

- Markov inequality
  - X: continuous random variable with mean μ
  - supp(X): non-negative, i.e. $P(X < 0) = 0$

$$P(X > c) \leq \frac{\mu}{c}$$

- Chebyshev inequality
  - X: continuous random variable with mean μ and variance σ^2

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

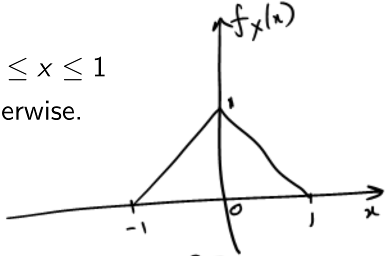
Probability space and its axioms

- Discrete case
  - Sample space: finite or countable set
  - Events: power set of sample space
  - Probability function: PMF
- Continuous case
  - Sample space: interval of real line
  - Events: intervals in the sample space along with their complements and countable unions
    - ⋆ This avoids some 'bizarre' subsets that defy our sense of measure
  - Probability function: function from intervals inside sample space to $[0, 1]$ satisfying the axioms
    - ⋆ Possible only if $P(X = x) = 0$
- Unified description of probability spaces: Measure-theoretic
  - NPTEL course: https://nptel.ac.in/courses/108/106/108106083/

Problem

A continuous random variable X has PDF

$$f_X(x) = \begin{cases} 1 - |x|, & -1 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$



Find the CDF of X, $E[X]$, $\mathrm{Var}(X)$.

$$F_X(x) = \begin{cases} 0, & x < -1 \\ ?, & -1 \le x < 0 \\ ?, & 0 \le x < 1 \\ 1, & x \geq 1 \end{cases}$$

$E[X] = 0$

$-1 \le x \le 0$

$$F_X(x) = \underbrace{\int_{-1}^{x} (1 - |u|)\, du}_{x=0} = \int_{-1}^{x} (1+u)\, du = u\Big|_{-1}^{x} + \frac{u^2}{2}\Big|_{-1}^{x} = (x - (-1)) + \left(\frac{x^2}{2} - \frac{(-1)^2}{2}\right)$$

$$= x + 1 + \frac{x^2}{2} - \frac{1}{2} = \frac{1}{2} + x + \frac{x^2}{2}$$

$0 \le x \le 1$

$$F_X(x) = \underbrace{\int_{-1}^{0} (1 - |u|)\, du}_{1/2} + \int_{0}^{x} (1 - |u|)\, du = \frac{1}{2} + u\Big|_{0}^{x} - \frac{u^2}{2}\Big|_{0}^{x} = \frac{1}{2} + x - \frac{x^2}{2}$$

$$E[X] = \int_{-1}^{1} x f_X(x)\,dx = \int_{-1}^{0} x(1+x)\,dx + \int_{0}^{1} x(1-x)\,dx$$

$$= \frac{x^2}{2}\Big|_{-1}^{0} + \frac{x^3}{3}\Big|_{-1}^{0} + \frac{x^2}{2}\Big|_{0}^{1} - \frac{x^3}{3}\Big|_{0}^{1} = -\frac{1}{2} + \frac{1}{3} + \frac{1}{2} - \frac{1}{3} = 0$$

$$Var(X) = E[X^2] = \int_{-1}^{0} x^2(1+x)\,dx + \int_{0}^{1} x^2(1-x)\,dx$$

$$= \frac{x^3}{3}\Big|_{-1}^{0} + \frac{x^4}{4}\Big|_{-1}^{0} + \frac{x^3}{3}\Big|_{0}^{1} - \frac{x^4}{4}\Big|_{0}^{1} = \frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4} = \frac{1}{6}$$

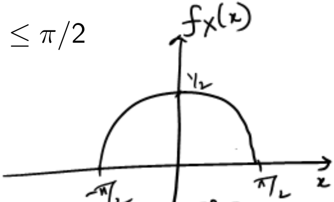
$$\int x^n\,dx = \frac{x^{n+1}}{n+1}$$

$$|x| = \begin{cases} -x, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases}$$

Problem

A continuous random variable X has PDF

$$f_X(x) = \begin{cases} \frac{1}{2}\cos x, & -\pi/2 \le x \le \pi/2 \\ 0, & \text{otherwise.} \end{cases}$$



Find the CDF of X, $E[X]$, $\text{Var}(X)$.

$E[X] = 0$

$$F_X(x) = \begin{cases} 0, & x < -\pi/2 \\ ?, & -\pi/2 \le x \le \pi/2 \\ 1, & x > \pi/2 \end{cases}$$

$$\int \cos x \, dx = \sin x$$

$-\pi/2 \le x \le \pi/2$

$$F_X(x) = \int_{-\pi/2}^{x} f_X(u)\,du = \frac{1}{2} \sin x \Big|_{-\pi/2}^{x} = \frac{1}{2}\Big(\sin x - \underbrace{\sin(-\pi/2)}_{-1}\Big) = \frac{1+\sin x}{2}$$

$$\int x\cos x\,dx = \cos x + x\sin x \qquad \int x^2\cos x\,dx = x^2\sin x + 2x\cos x - 2\sin x$$

$$E[X] = \int_{-\pi/2}^{\pi/2} x\,\tfrac{1}{2}\cos x\,dx = \frac{1}{2}\left(\cos x + x\sin x\right)\Big|_{-\pi/2}^{\pi/2} = \frac{1}{2}\Big[\big(\overset{0}{\cancel{\cos\tfrac{\pi}{2}}} + \tfrac{\pi}{2}\overset{1}{\cancel{\sin\tfrac{\pi}{2}}}\big) - \big(\underbrace{\cos(-\tfrac{\pi}{2})}_{0} - \tfrac{\pi}{2}\underbrace{\sin(-\tfrac{\pi}{2})}_{-1}\big)\Big]$$

$$= 0$$

$$\mathrm{Var}(X) = E[X^2] = \int_{-\pi/2}^{\pi/2} x^2\cdot\tfrac{1}{2}\cos x\,dx = \frac{1}{2}\left(x^2\sin x + 2x\cos x - 2\sin x\right)\Big|_{-\pi/2}^{\pi/2}$$

$$= \frac{1}{2}\left[\left(\frac{\pi^2}{4} - 2\right) - \left(-\frac{\pi^2}{4} + 2\right)\right]$$

$$= \frac{\pi^2}{4} - 2$$