Statistical signal processing (5CTA0)

Lecture 1, part B

Lecturer: Simona Turco

Electrical Engingeering, Signal Processing Systems group

Part 1: Random variables and Random Signals

Part 1

Random Variables and Random Signals

Lecture 1: Probability and Random Variables

Part A: Probability
Part B: Random variables

Random variables

Lecture 1, Part B

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Lecture 1, part B: Random variables

- Introduction and basic definitions
- Discrete and continuous random variables:
  - Probability distributions
- Statistical description of random variables:
  - Expectation and moments
- Families of random variables

Random variables: Introduction

- Random variables are the outcome of a stochastic or random process

Random processes

Flipping of a fair coin

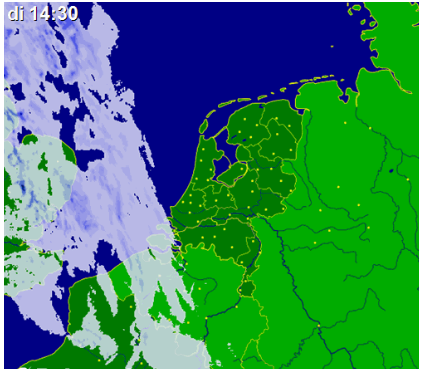
Rolling of a die

Random variables: Introduction

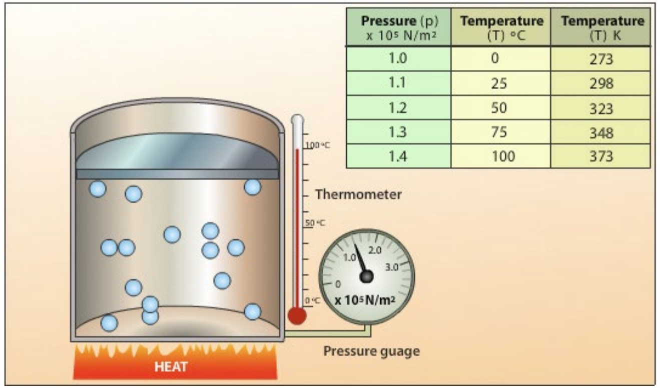
- Random variables are the outcome of a stochastic or random process

Random variables

Tomorrow's rainfall



Temperature of a gas



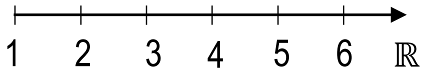
Random variables: introduction

Random variables (RV) are ways to map outcomes of a random process to numbers

Random process

- Rolling a dice
- Flipping a coin
- Measuring the temperature of a gas

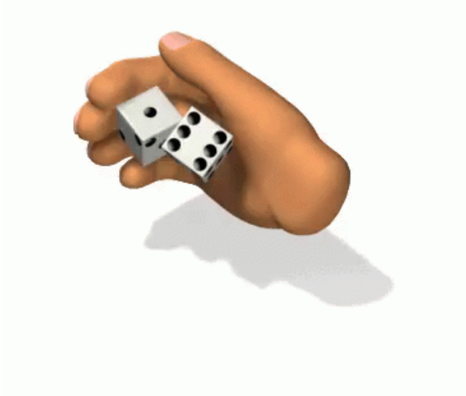
Random variable



While an event can be defined in many ways, a RV is always numerical!

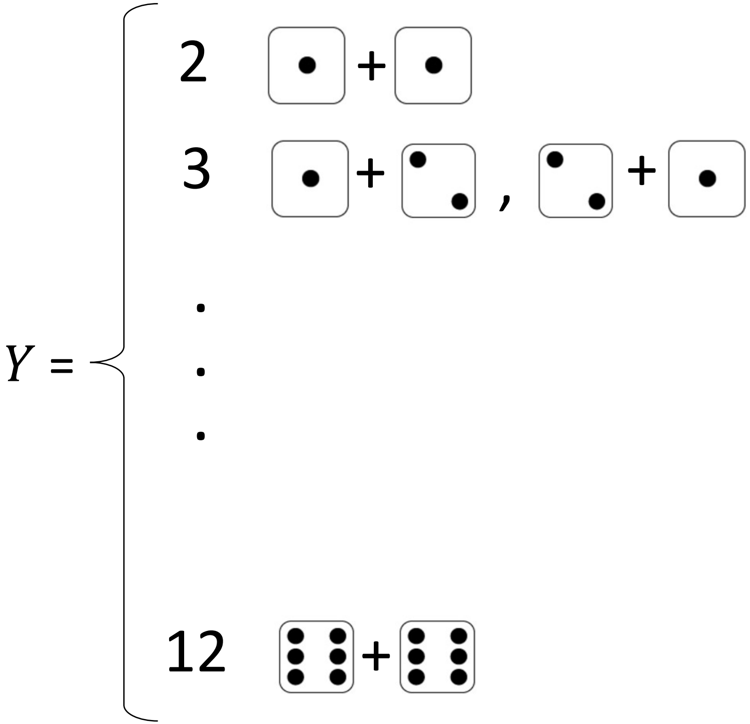
Random variables: examples

Random process: Rolling 2 dice



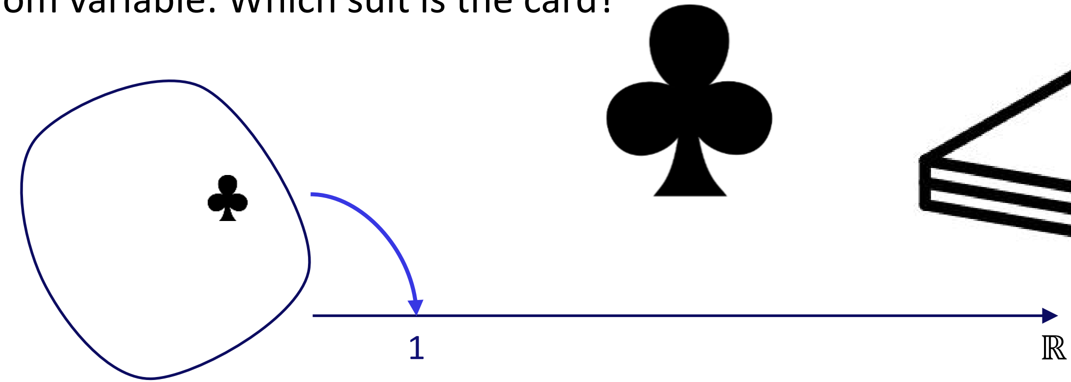
$D1$ = number on first die
$D2$ = number on second die
Y = [sum of the upward faces]

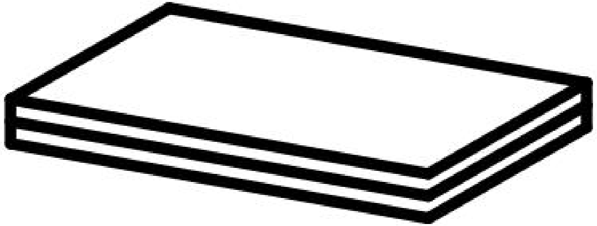
$$Y = D1 + D2$$



Random variables - examples

Random process: Draw a card from a poker deck.
Random variable: Which suit is the card?



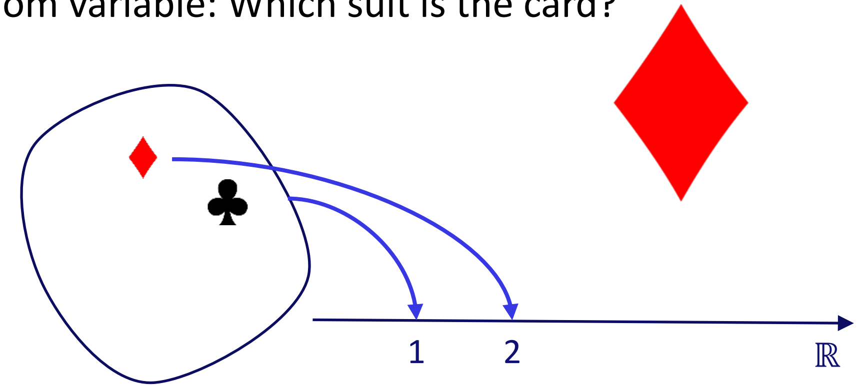


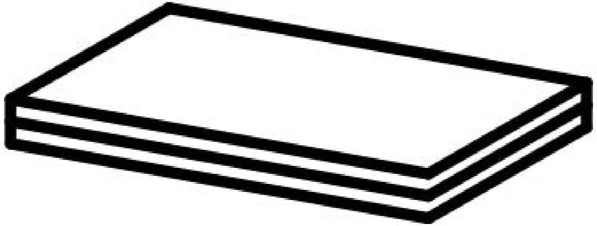
$$X(s) = \begin{cases} X(x=\clubsuit) = 1 \\ \\ \\ \end{cases}$$

Random variables - examples

Random process: Draw a card from a poker deck.

Random variable: Which suit is the card?

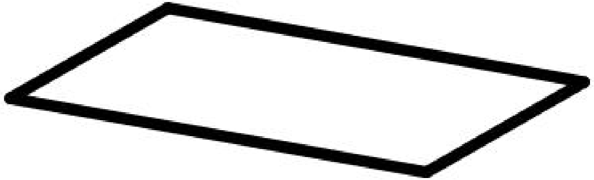


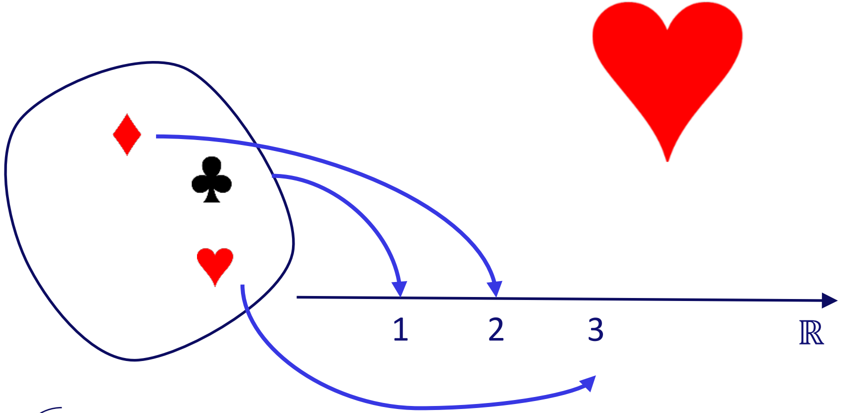


$$X(s) = \begin{cases} X(x = \clubsuit) = 1 \\ X(x = \diamondsuit) = 2 \\ \\ \end{cases}$$

Random variables - examples

Random process: Draw a card from a poker deck.
Random variable: Which suit is the card?

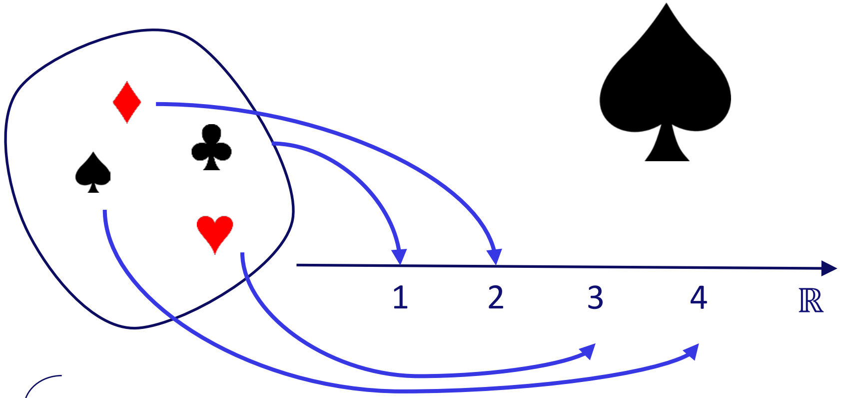




$$X(s) = \begin{cases} X(x = \clubsuit) = 1 \\ X(x = \diamondsuit) = 2 \\ X(x = \heartsuit) = 3 \end{cases}$$

Random variables - examples

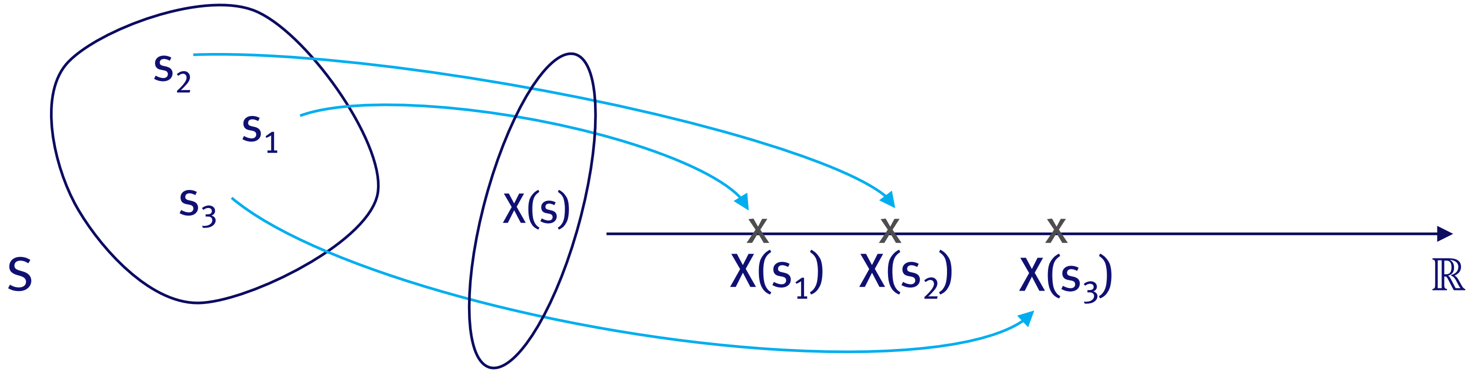
Random process: Draw a card from a poker deck.
Random variable: Which suit is the card?



$$X(s) = \begin{cases} X(x = \clubsuit) = 1 \\ X(x = \diamondsuit) = 2 \\ X(x = \heartsuit) = 3 \\ X(x = \spadesuit) = 4 \end{cases}$$

Random variables: formal definition

A random variable $X(s)$ is a function that maps all elements s of the sample space S into points on the real line (real numbers) of parts thereof



Notation:

- S, sample space
- $X(s)$ or X, random variable
- $s \in S$, value that the random variable X can take on

Deterministic vs random variables

x deterministic variables

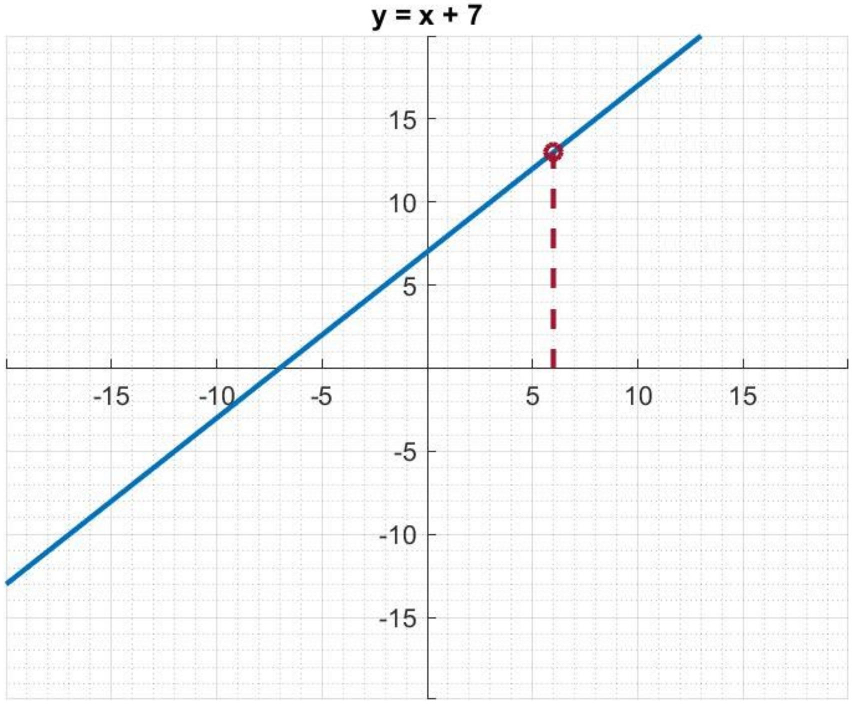
$$x + 5 = 7 \quad \Rightarrow \quad x = 2$$

x is fixed and it is always equal to 2

Deterministic vs random variables

x, y deterministic variables

$$y = x + 7$$

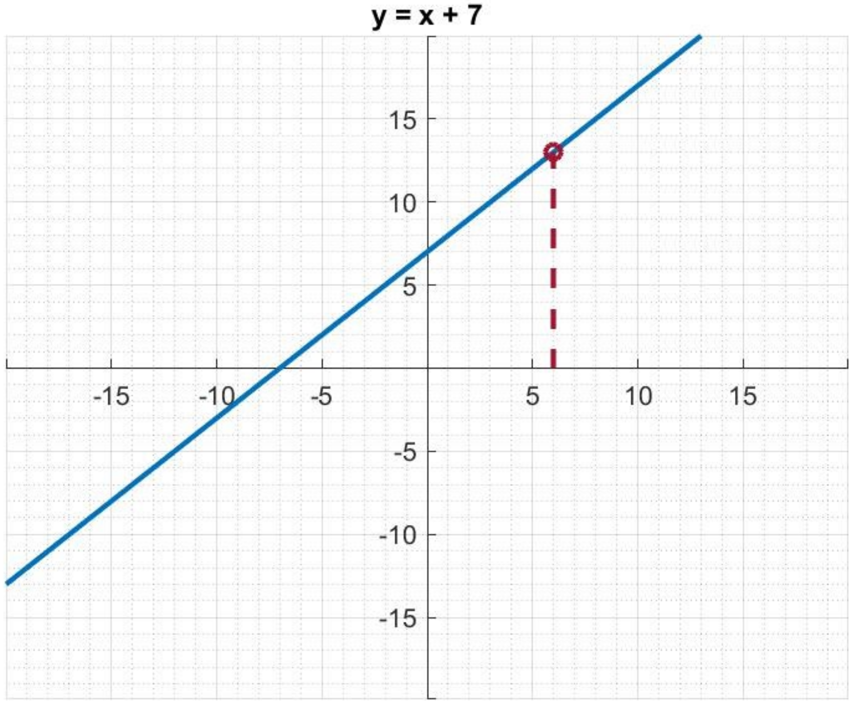


Deterministic vs random variables

x, y deterministic variables

$$y = x + 7$$

x	y
-15	-8
-10	-3
-5	2
0	7
5	5
10	17
15	22



Deterministic vs random variables

Ease of notation

- $\Pr[\text{sum of the upward faces of rolling two dice is smaller than 10}]$
- $\Pr[Y < 10]$

Mathematical expressions can be used, BUT…

- Random variables can take on many different values with different probabilities => we talk about the probability of taking on a certain value

Random variables

Y can take on many values!

$Y = 8, 10, 9, 8.5 \dots$ → $\Pr[Y = 10] = 0.025$

$\Pr[Y = 8] = 0.15$

Random variables: definitions and conditions

Notation and definitions

- A random variable (RV) is $X(s)$ or X;
- A ***realization*** or ***observation*** of RV is x;
- The sample space S is the **domain** of the RV;
- The **range** S_X of the RV is a collection of all the possible values x that X can take on

Conditions for a function to be a random variable

- **Single-valued**: every point in S corresponds to only one value of X;
- $[X \leq x]$ is an event for any real number x;
- $\Pr[X = -\infty] = 0$; $\Pr[X = \infty] = 0$;

Discrete and continuous random variables

- Random variables can be either **discrete**, **continuous,** or **mixed**

Discrete random variables

- Random variables can be either **discrete**, **continuous,** or **mixed**
- A discrete random variable can take any of a countable list of distinct values (discrete range)

Examples of discrete RVs

- Random variables can be either **discrete, continuous,** or **mixed**
- A discrete random variable can take any of a <u>countable</u> list of distinct values (discrete range)

X = {number on upward face of a die}
X can only take values 1, 2, 3, 4, 5, 6

Finite discrete random variable

Examples of discrete RVs

- Random variables can be either **discrete**, **continuous,** or **mixed**
- A discrete random variable can take any of a countable list of distinct values (discrete range)

Z = {length of straw}
Take straws from 4 different brands
Z can only take values 18.7; 19.2; 19.5; 18.5

Finite discrete random variable

Examples of discrete RVs

- Random variables can be either **discrete**, **continuous,** or **mixed**
- A discrete random variable can take any of a <u>countable</u> list of distinct values (discrete range)

Y = {number of years before the end of the world}
Y can take any discrete value from 0 to ∞

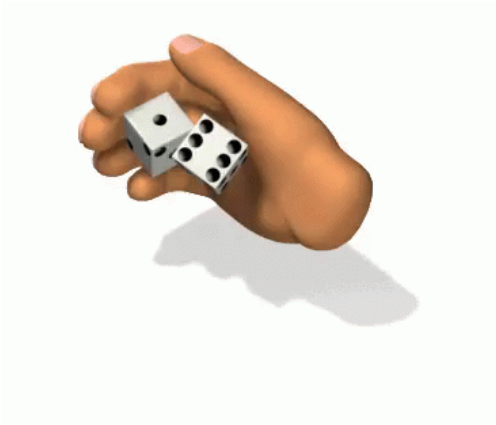
Infinite discrete random variable

Probability distributions

- The **probability distribution** of a random variable X is a description of the probabilities associated with all the possible values of X

Probability distributions: example

Random process: Roll 2 dice
Y = [sum of the upward faces]
N = number of trials

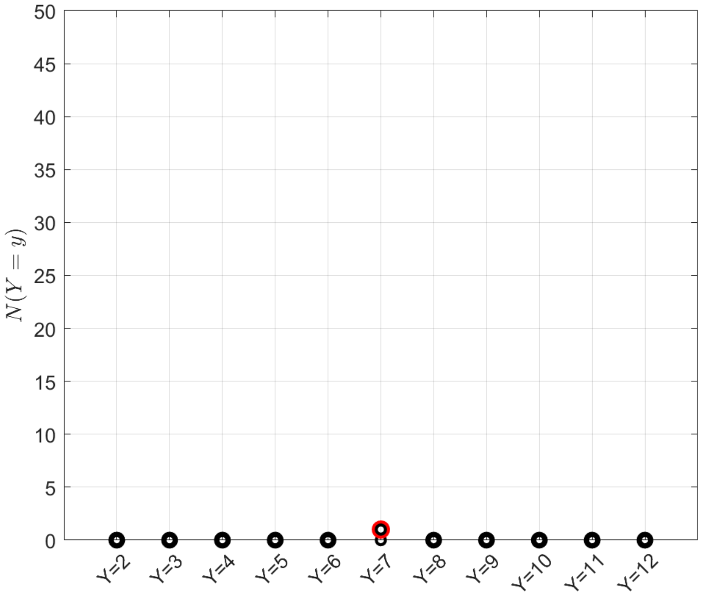


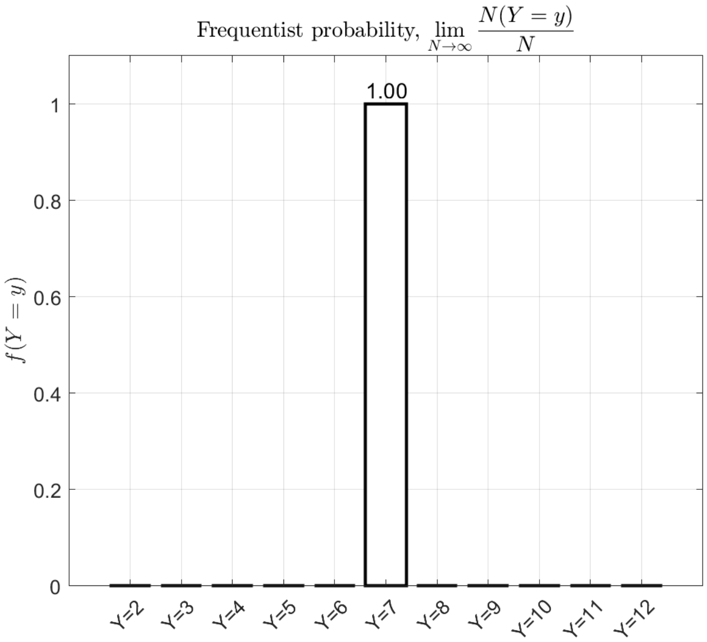
Frequentist definition of probability

$$\Pr[Y = y] = \lim_{N \to \infty} \frac{N(Y = y)}{N} = \lim_{N \to \infty} f(Y = y)$$

Probability distributions: example

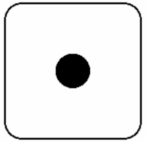
N = 1



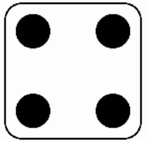


Probability distributions: example

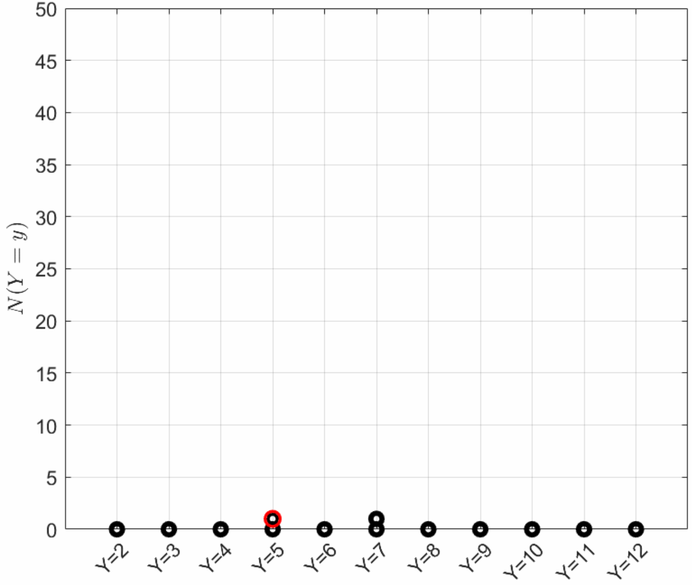
D1

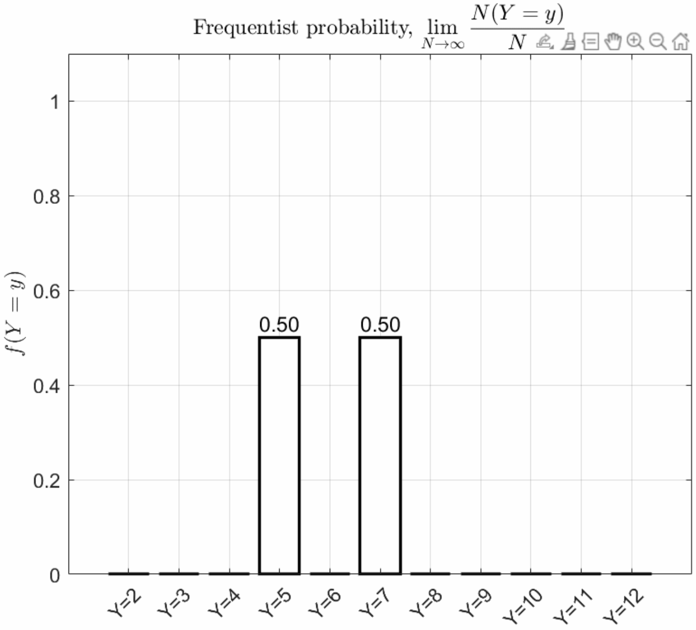


D2



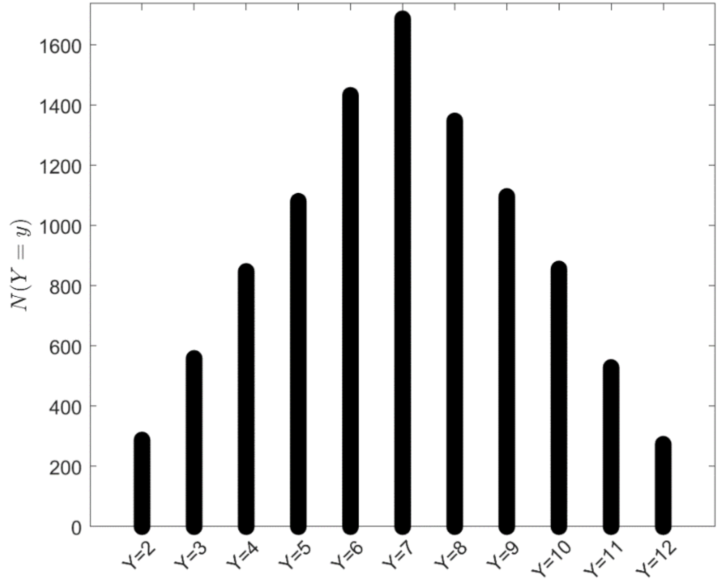
N = 2

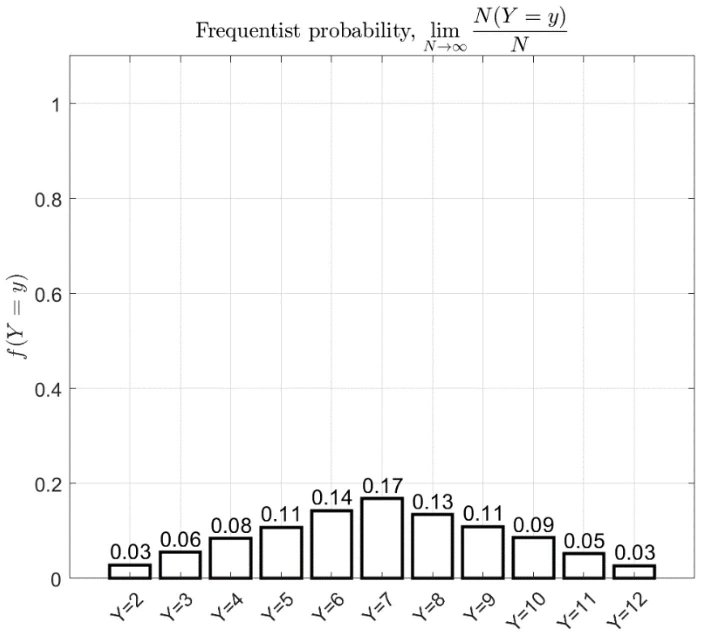




Probability distributions: example

N = 10000





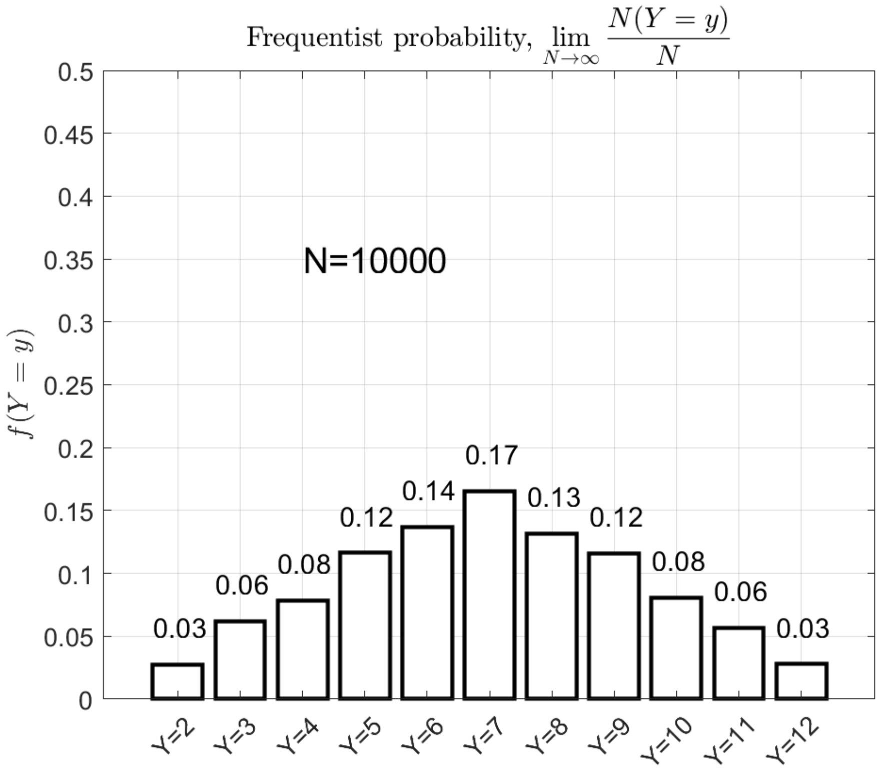
TU/e

Probability distributions: example

Random process: Roll 2 dice

Y = [sum of the upward faces]

{d1, d2}	y	Pr[Y=y]
{1,1}	2	0.028
{1,2}, {2,1}	3	0.056
{1,3}, {2,2}, {3,1}	4	0.083
{1,4}, {2,3}, {3,2}, {4,1}	5	0.11
{1,5}, {2,4}, {3,3}, {4,2}, {5,1}	6	0.14
{1,6}, {2,5}, {3,4}, {4,3}, {5,2}, {6,1}	7	0.17
{2,6}, {3,5}, {4,4}, {5,3}, {6,2}	8	0.14
{3,6}, {4,5}, {5,4}, {6,3}	9	0.11
{4,6}, {5,5}, {6,4}	10	0.083
{5,6}, {6,5}	11	0.056
{6,6}	12	0.028



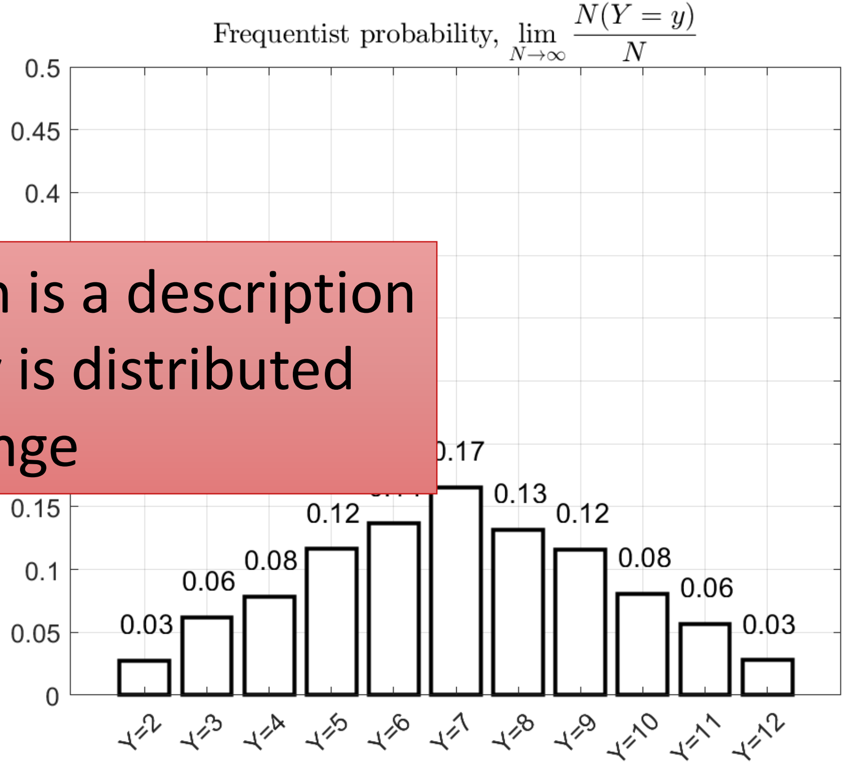
Probability distributions: example

Random process: Roll 2 dice

Y = [sum of the upward faces]

{d1, d2}	y	Pr[Y=y]
{1,1}	2	0.028
{1,2		
{1,3}, {		
{1,4}, {2,3		
{1,5}, {2,4}, {		
{1,6}, {2,5}, {3,4		
{2,6}, {3,5}, {4,4}, {5,3}, {6,2}	8	0.14
{3,6}, {4,5}, {5,4}, {6,3}	9	0.11
{4,6}, {5,5}, {6,4}	10	0.083
{5,6}, {6,5}	11	0.056
{6,6}	12	0.028

A probability distribution is a description of how the probability is distributed over the range



Probability distributions

- The **probability mass function (PMF)** of a **discrete** random variable X is a list of each possible value of X together with the probability that X takes that value in one trial of the experiment

$$p_X(x) = \Pr[X = x]$$

Properties

- $p_X(x) = 0$ if x is not one of the possible value of X
- $0 \leq p_X(x) \leq 1$
- $\sum_x p_X(x) = 1$

Probability Axioms:

1. $0 \leq \Pr[A] \leq 1$
2. $\Pr[S] = 1$
3. $\Pr[A_1 \cup A_2 \cup \ldots \cup A_M] = \Pr[A_1] + \Pr[A_2] + \ldots + \Pr[A_M]$

Probability distributions

- The **cumulative distribution function (CDF)** is defined as

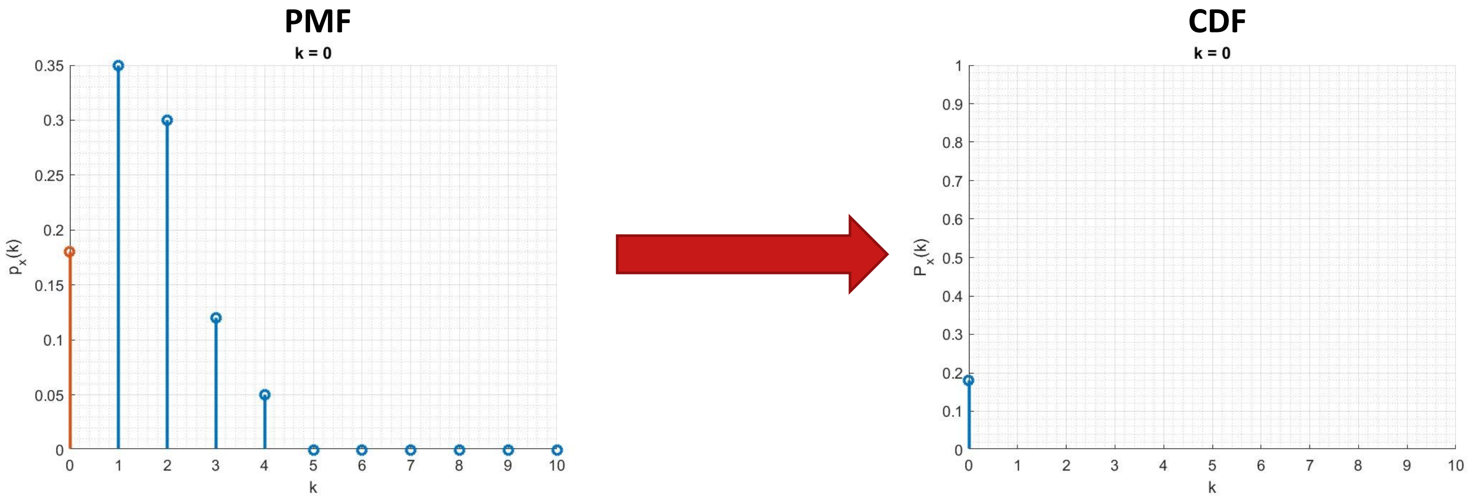
$$P_X(x) = \Pr[X \leq x]$$

Properties

- It can be defined for discrete (only if ordered), continuous and mixed random variables
- $0 \leq P_X(x) \leq 1$
- $P_X(-\infty) = 0$ and $P_X(\infty) = 1$
- $P_X(x') \geq P_X(x)$ for all $x' \geq x$ (increasing function)
- $P_X(x_1) - P_X(x_2) = \Pr[x_1 < X < x_2]$

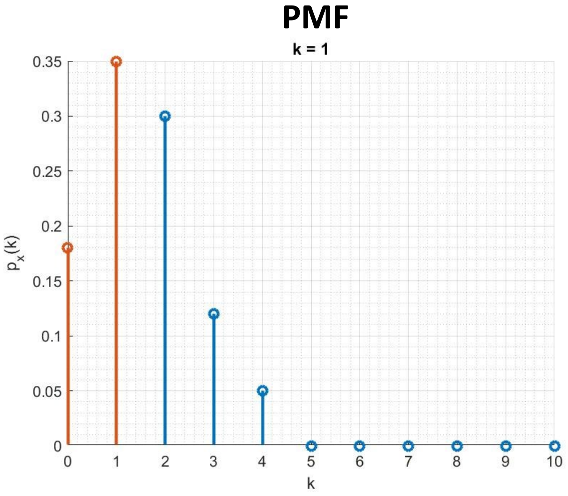
From PMF to CDF

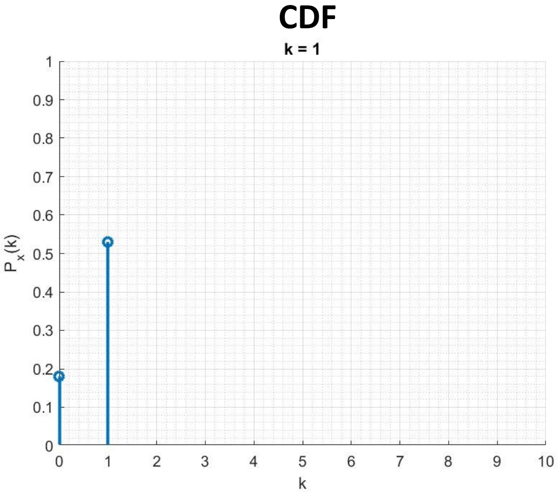
- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \leq x} p_X(x_i)$



From PMF to CDF

- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \leq x} p_X(x_i)$

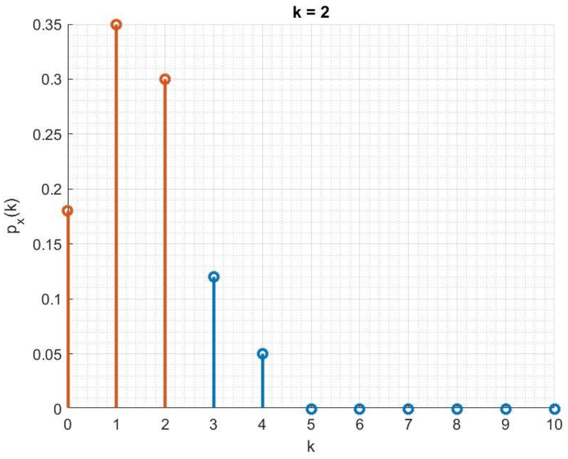




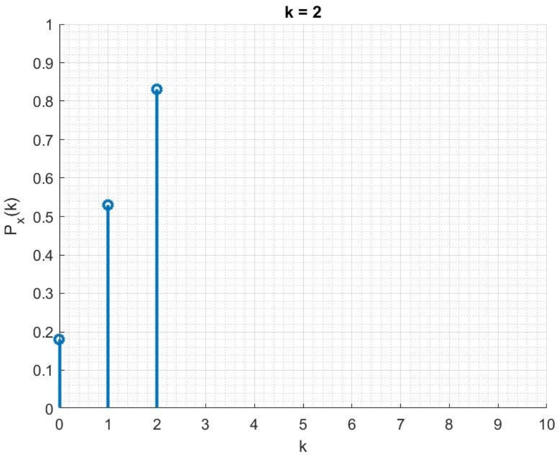
From PMF to CDF

- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \leq x} p_X(x_i)$

PMF

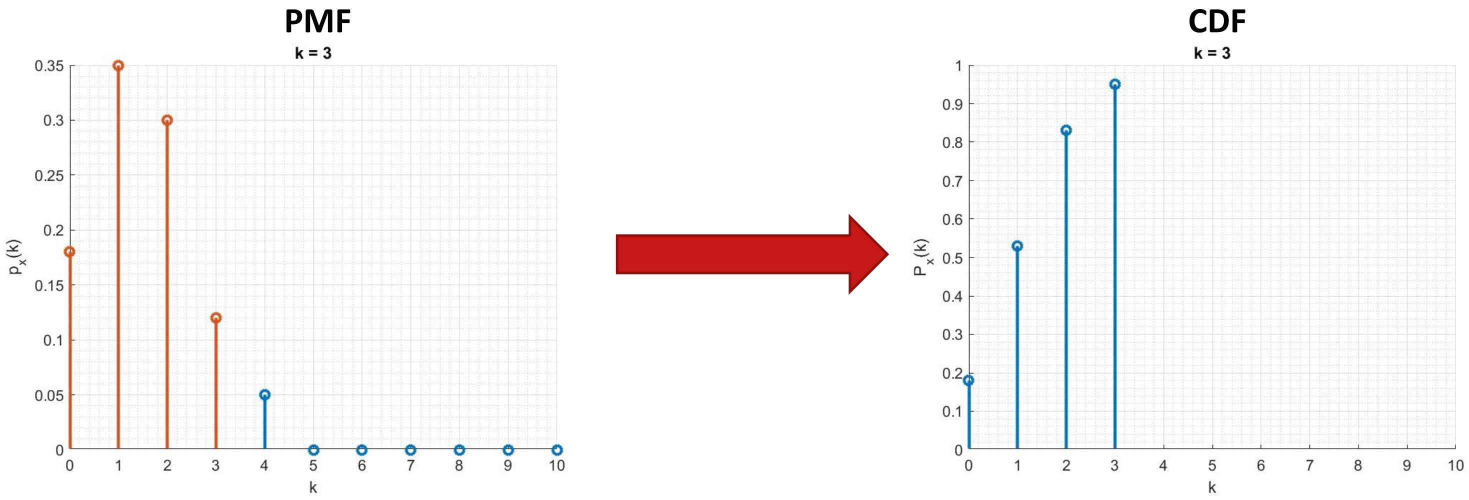


CDF



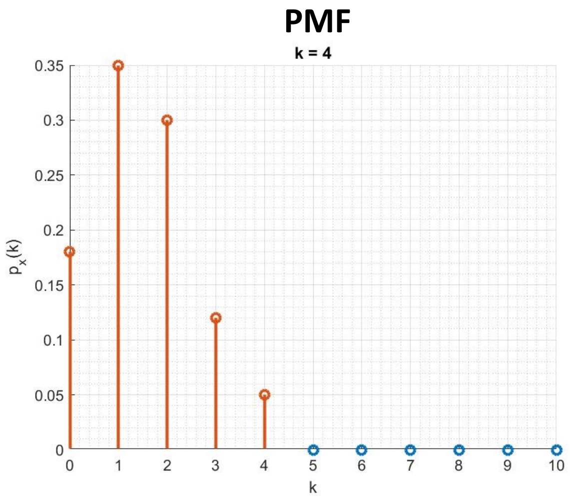
From PMF to CDF

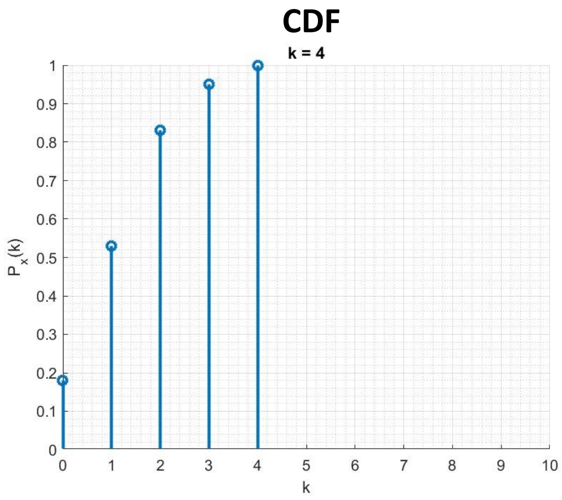
- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \leq x} p_X(x_i)$



From PMF to CDF

- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \leq x} p_X(x_i)$

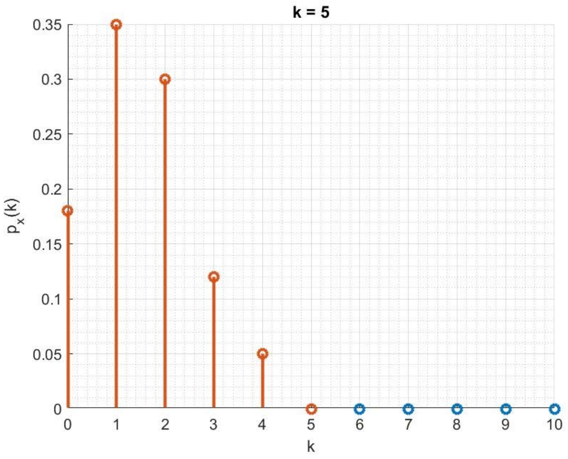




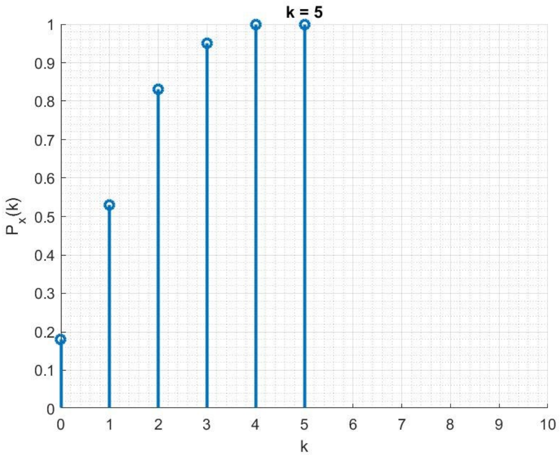
From PMF to CDF

- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \le x} p_X(x_i)$

PMF

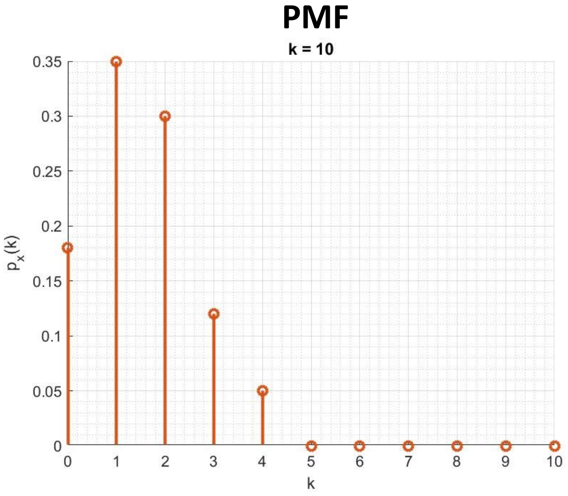


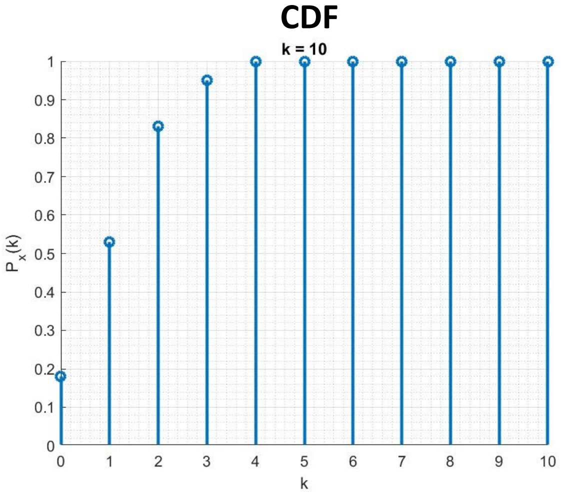
CDF



From PMF to CDF

- A CDF can be obtained from the PMF as $P_X(x) = \sum_{x_i \leq x} p_X(x_i)$



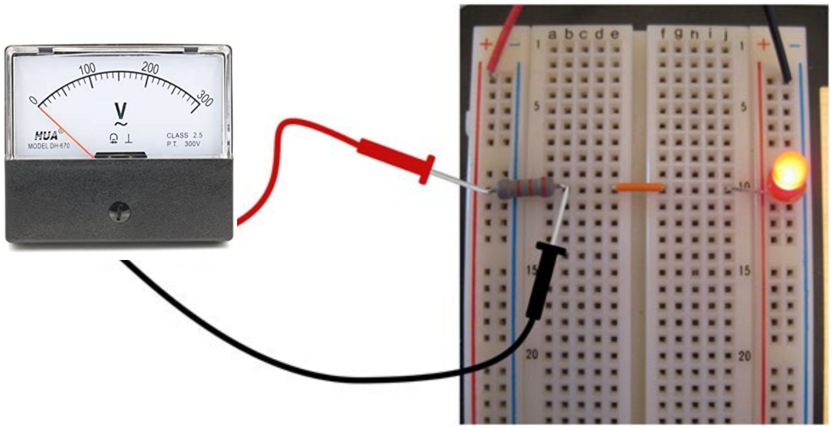


Continuous random variables

- A **continuous** random variable can take any value in an interval of collections of intervals (continuous range)

Continous random variables

- A **continuous** random variable can take any value in an interval of collections of intervals (continuous range)



X= {voltage across the resistor}
X can take any real value between 0 mV and 300mV

Continuous random variable

Continous random variables

- A **continuous** random variable can take any value in an interval of collections of intervals (continuous range)

X= {time of arrival of the professor in the classroom}

Continuous random variable

Probability distributions

The **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X

Discrete random variables

- Probability *mass* function (PMF)
- Cumulative distribution function (CDF)

⟺

Continuous random variables

- Probability **density** function (PDF)
- Cumulative distribution function (CDF)

Probability distributions

- The **probability density function (PDF)** of a continuous random variable

$$p_X(x) = \frac{dP_X(x)}{dx} \neq \Pr[X = x]$$

- Properties
  - $p_X(x) = 0$ if x is not one of the possible value of X
  - $0 \leq p_X(x) \leq 1$
  - $\int_{-\infty}^{\infty} p_X(x) = 1$

Probability distributions

- The **cumulative distribution function (CDF)** ...

$$P_X(x) = \Pr[X \leq x] = \int_{-\infty}^{x} p_X(x)dx$$

- Properties of CDF
  - $0 \leq P_X(x) \leq 1$
  - $P_X(-\infty) = 0$ and $P_X(\infty) = 1$
  - $P_X(x') \geq P_X(x)$ for all $x' \geq x$ (increasing function)
  - $P_X(x_1) - P_X(x_2) = \Pr[x_1 < X < x_2]$

Probability distributions

For continuous random variables, the probability of each individual outcome is exactly zero

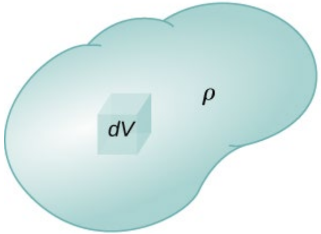
- Professor typically arrives between 8.55 and 9.05
- Model the arrival time as T ={difference in minutes from 9.00}
- $\Pr[-1 < \mathrm{T} < 1] > \Pr[-0.5 < \mathrm{T} < 0.5] > \Pr[-10^{-6} < \mathrm{T} < 10^{-6}]$
- $\Pr[\mathrm{T} = 0] = 0$

"Amount" of probability gets smaller as the interval gets smaller

Probability distributions

For continuous random variables, the probability of each individual outcome is exactly zero.

Analogy: mass in continuous volume



- Finite volume has some mass
- No mass at a single point
- Density of matter

- Finite interval has some probability
- The probability at a single point is zero
- Probability density

Statistical characterization of a RV

When the probability distribution of a RV is known, we can characterize the RV by calculating moments by the expectation operator $E[\cdot]$

- The expected value (first moment) of a random variable X is

DISCRETE RVs:

$$E[X] = \mu_X = \sum_{x \epsilon S_X} x p_X(x)$$

CONTINOUS RVs:

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

- The expected value describes the center of gravity of the probability distribution

Second moment: variance

- The variance $\mathrm{Var}[X]$ (second *central* moment) of a random variable X, with expectation μ_x

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

$$\mathrm{Std}[X] = \sqrt{\sigma_X^2} = \sigma_X$$

- The variance describes the spread of the probability distribution

Properties of expectation

a, b deterministic constant

- The expectation is a linear operator

  $$E[aX + bY] = aE[X] + bE[Y]$$

- The expectation is a positive operator

  If $X \geq Y$ then $E[X] \geq E[Y]$

- For a deterministic variable z,

  $$E[z] = z$$

- $E[X - \mu_x] = 0$
- $E[aX + b] = aE[X] + b$
- $\mathrm{var}[X] = E[X^2] - \mu_X^2$
- $\mathrm{var}[aX + b] = a^2 \mathrm{var}[X]$

Moments: definition

For any integer $m > 0$, the m-th moment of a random variable X is

$$E[X^m] = \sum_{x \epsilon S_X} x^m p_X(x)$$

For any integer $m > 0$, the m-th central moment of a random variable X is

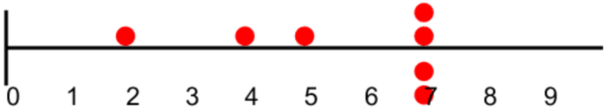
$$E[(X - \mu_X)^m] = \sum_{x \epsilon S_X} (x - \mu_X)^m p_X(x)$$

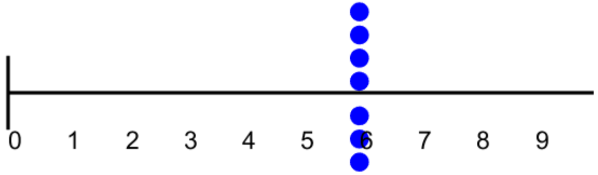
For any integer $m > 0$, the m-th central normalized moment of a random variable X is

$$\frac{E[(X - \mu_X)^m]}{\sigma^m} = \frac{1}{\sigma^m} \sum_{x \epsilon S_X} (x - \mu_X)^m p_X(x)$$

Moments and descriptive statistics

- In practice we don't have probability distribution
- *Descriptive statistics*: moments provide information about data distribution





$$m^{(1)} = \frac{1}{n}\sum_{i=1}^{n} x_i = 6$$

$$m^{(1)} = \frac{1}{n}\sum_{i=1}^{n} x_i = 6$$

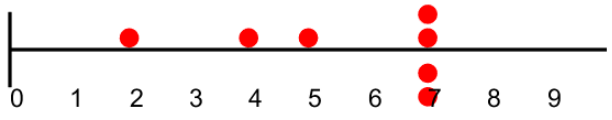
First moment: average **DISTANCE** from zero

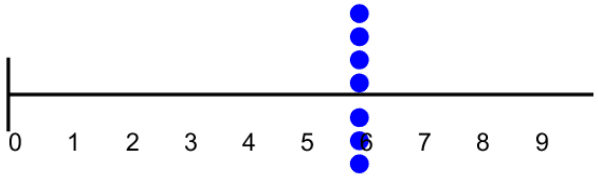
$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Mean

Moments and descriptive statistics

- In practice we don't have probability distribution
- *Descriptive statistics:* moments provide information about data distribution





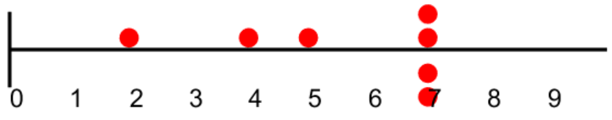
$$m^{(2)} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 288$$

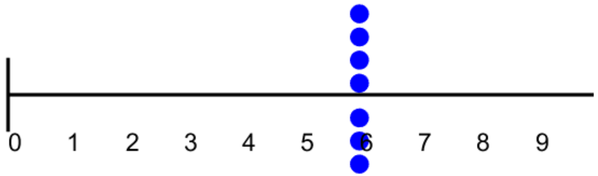
$$m^{(2)} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 255$$

Second moment: average **SQUARED DISTANCE** from zero

Moments and descriptive statistics

- In practice we don't have probability distribution
- *Descriptive statistics*: moments provide information about data distribution





$$\frac{\sum_{i=1}^{n}\left(x_i - m^{(1)}\right)^2}{n} = 36$$

$$\frac{\sum_{i=1}^{n}\left(x_i - m^{(1)}\right)^2}{n} = 0$$

Second centered moment:
average **SQUARED DISTANCE** from *the mean*

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Variance

Moments: statistical approach

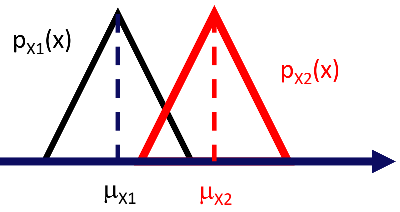
		Centered	Standardized
1st	$\frac{1}{n}\sum_{i=1}^{n} x_i$		
2nd	$\frac{1}{n}\sum_{i=1}^{n} x_i^2$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$	
3rd	$\frac{1}{n}\sum_{i=1}^{n} x_i^3$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{n}$	$\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3}$
4th	$\frac{1}{n}\sum_{i=1}^{n} x_i^4$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{n}$	$\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4}$

Moments: statistical approach

	MEAN	Centered	Standardized	
1^{st}	$\boxed{\frac{1}{n}\sum_{i=1}^{n} x_i}$			
		VARIANCE		
2^{nd}	$\frac{1}{n}\sum_{i=1}^{n} x_i^2$	$\boxed{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$		
3^{rd}	$\frac{1}{n}\sum_{i=1}^{n} x_i^3$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{n}$	$\boxed{\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3}}$	*SKEWNESS*
4^{th}	$\frac{1}{n}\sum_{i=1}^{n} x_i^4$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{n}$	$\boxed{\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4}}$	*KURTOSIS*

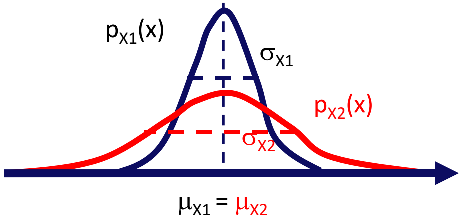
Mean (m=1)

Center of gravity



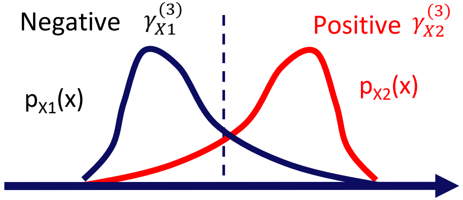
Variance (m=2, central)

Spread



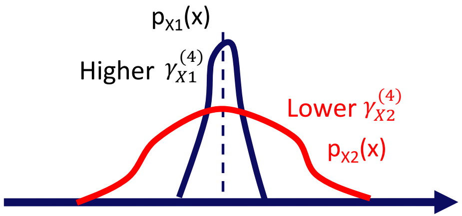
Skewness (m=3, normalized)

Simmetry



Kurtosis (m=4, normalized)

Tailedeness



Moments: statistical approach

	MEAN	Centered	Standardized	
1^{st}	$\boxed{\frac{1}{n}\sum_{i=1}^{n} x_i}$			
		VARIANCE		
2^{nd}	$\frac{1}{n}\sum_{i=1}^{n} x_i^2$	$\boxed{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$		
3^{rd}	$\frac{1}{n}\sum_{i=1}^{n} x_i^3$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{n}$	$\boxed{\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3}}$	*SKEWNESS*
4^{th}	$\frac{1}{n}\sum_{i=1}^{n} x_i^4$	$\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{n}$	$\boxed{\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4}}$	*KURTOSIS*

Good approximation only for $n \longrightarrow \infty$

Sample Moments

	MEAN		***SAMPLE MEAN***
1st	$\frac{1}{n}\sum_{i=1}^{n} x_i$	$\Rightarrow$	$\frac{1}{n}\sum_{i=1}^{n} x_i$
	VARIANCE		***SAMPLE VARIANCE***
2nd	$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$	$\Rightarrow$	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
	SKEWNESS		***SAMPLE SKEWNESS***
3rd	$\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3}$	$\Rightarrow$	$\frac{n}{(n-1)(n-2)}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$
	KURTOSIS		***SAMPLE KURTOSIS***
4th	$\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4}$	$\Rightarrow$	$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{s^4}-\frac{3(n-1)^2}{(n-2)(n-3)}$

Good approximation only for $n \longrightarrow \infty$

Expectation vs (sample) mean

Example: we wonder how many goals does the PSV scores per match. We take as a sample the last 10 games and calculate some statistics.

$G = \{1, 2, 2, 3, 0, 1, 0, 2, 1, 4\}$

g	N_g
0	2
1	3
2	3
3	1
4	1

$$m_g = \frac{g_1+g_2+\cdots+g_n}{n}x = \sum_{g\epsilon S_G} \frac{N_g g}{n} = \frac{2\cdot 0}{10} + \frac{3\cdot 1}{10} + \frac{3\cdot 2}{10} + \frac{1\cdot 3}{10} + \frac{1\cdot 4}{10}$$

Expectation vs (sample) mean

- While the sample mean is a *statistic* of a set of experimental outcomes, the expectation is a parameter of a probability model

Sample mean

$$m_g = \sum_{g\epsilon S_G} \frac{N_g g}{n}$$

$$p_G(g) = \Pr[G = g] = \lim_{n\to\infty} \frac{N_g}{n}$$

Expected value (mean)

$$E[G] = \mu_G = \sum_{g\epsilon S_g} g p_g(g) = \sum_{g\epsilon S_G} g \lim_{n\to\infty} \frac{N_g}{n} = \lim_{n\to\infty} \sum_{g\epsilon S_G} \frac{N_g}{n} g = \lim_{n\to\infty} m_g$$

Moments vs descriptive statistics

When the probability distribution of a random variable X is available, moments can be calculated by the expectation operator $E[\cdot]$

When we only have a sample of possible values of X, moments can be approximated from the sample data

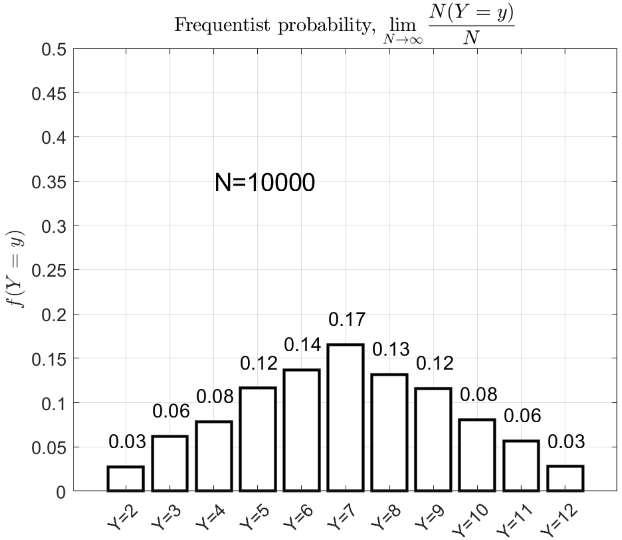
Parameters of the probability model describing X

Descriptive statistics of the random variable X

Families of random variables

- Families of random variables are sets of random variables that can be described by the same probability distributions

{d1, d2}	y	Pr[Y=y]
{1,1}	2	0.028
{1,2}, {2,1}	3	0.056
{1,3}, {2,2}, {3,1}	4	0.083
{1,4}, {2,3}, {3,2}, {4,1}	5	0.11
{1,5}, {2,4}, {3,3}, {4,2}, {5,1}	6	0.14
{1,6}, {2,5}, {3,4}, {4,3}, {5,2}, {6,1}	7	0.17
{2,6}, {3,5}, {4,4}, {5,3}, {6,2}	8	0.14
{3,6}, {4,5}, {5,4}, {6,3}	9	0.11
{4,6}, {5,5}, {6,4}	10	0.083
{5,6}, {6,5}	11	0.056
{6,6}	12	0.028



Families of random variables

- Families of random variables are sets of random variables that can be described by the same probability distributions
- They are typically fully described by a mathematical formula governed by a set of parameters

Discrete RV: binomial random variable

- A binomial random variable is the result of an experiment for which the following 4 conditions apply

1. The experiment consists of a sequence of n trials, with n is fixed in advance
2. The trials are identical, and each trial can result in one of the same two possible outcomes, which are denoted by success (S) or failure (F)
3. The trials are independent
4. The probability of success is constant from trial to trial and it is denoted by p.

Discrete RV: binomial random variable

- The binomial random variable depends only on the parameters *n* and *p* and it denotes as Binomial(*n*, *p*)

- The PMF and CDF of a binomial random variable are given by

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P_X(x) = \sum_{m=-\infty}^{x} \binom{n}{m} p^m (1-p)^{n-m}$$

With parameter p

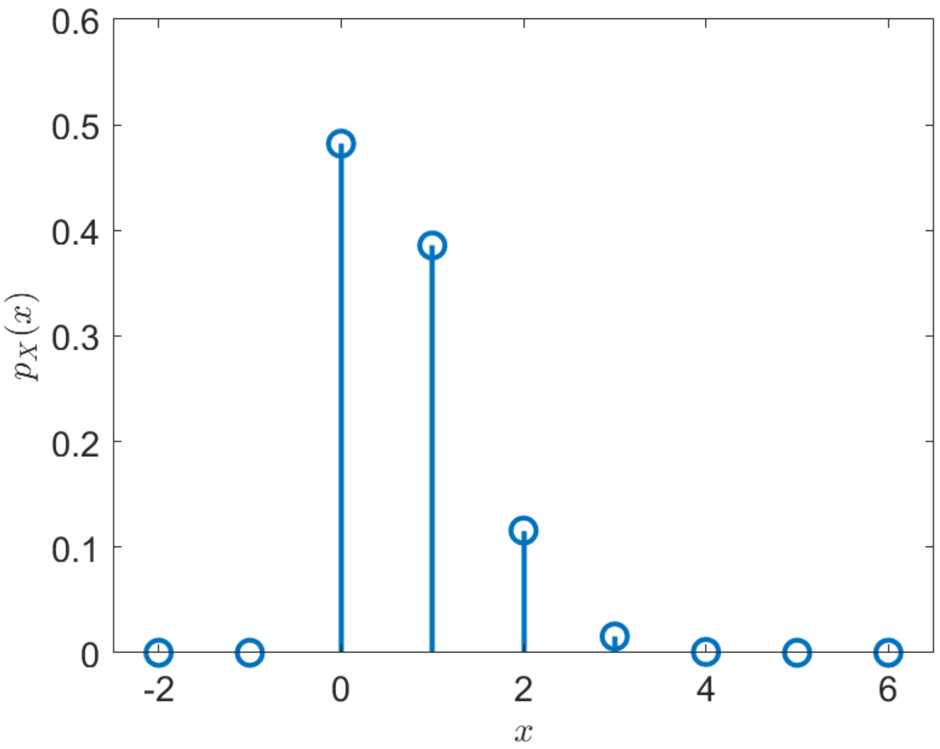
$$0 < p < 1$$

and n integer so that

$$n \geq 1$$

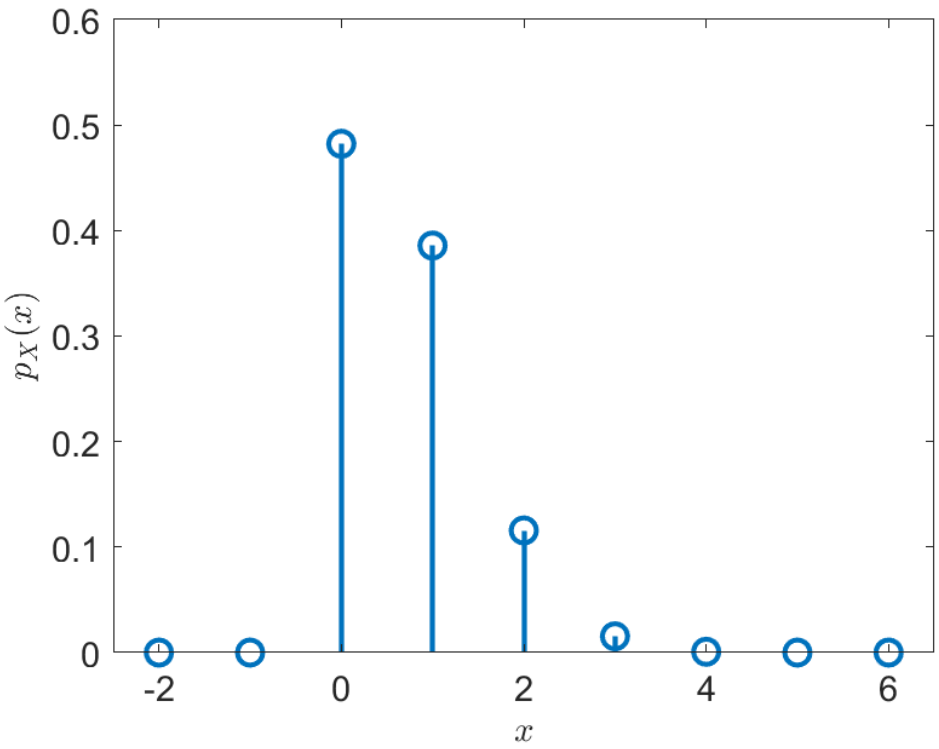
- It describes the probability of observing *x* successes in *n* independent trials

Discrete RV: binomial random variable



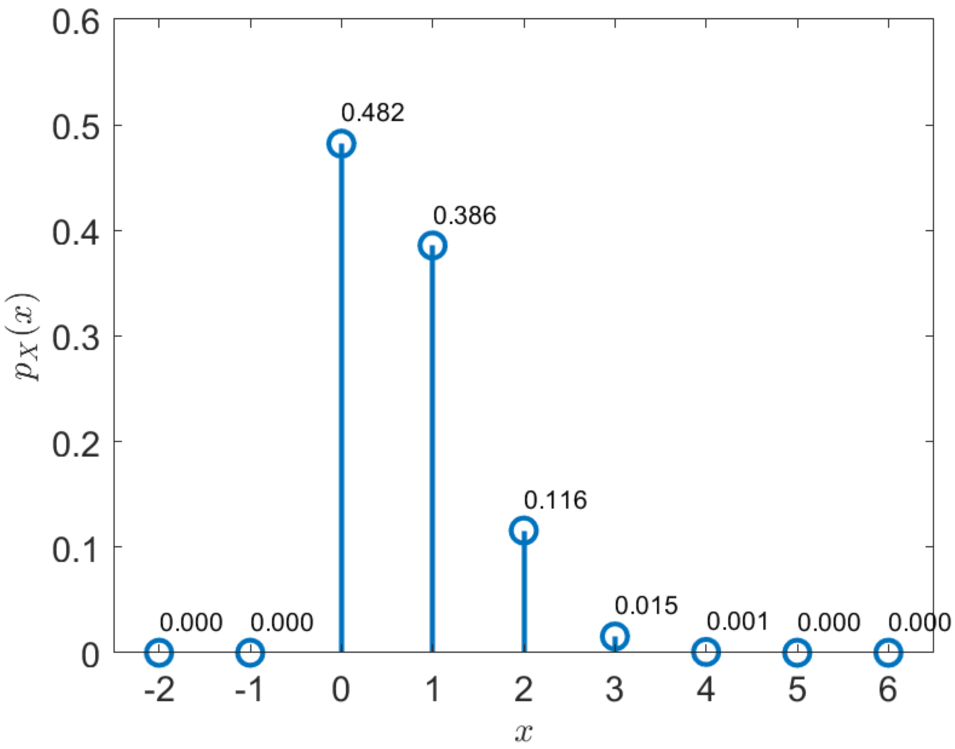
X~Binomial(n, p)

Discrete RV: binomial random variable



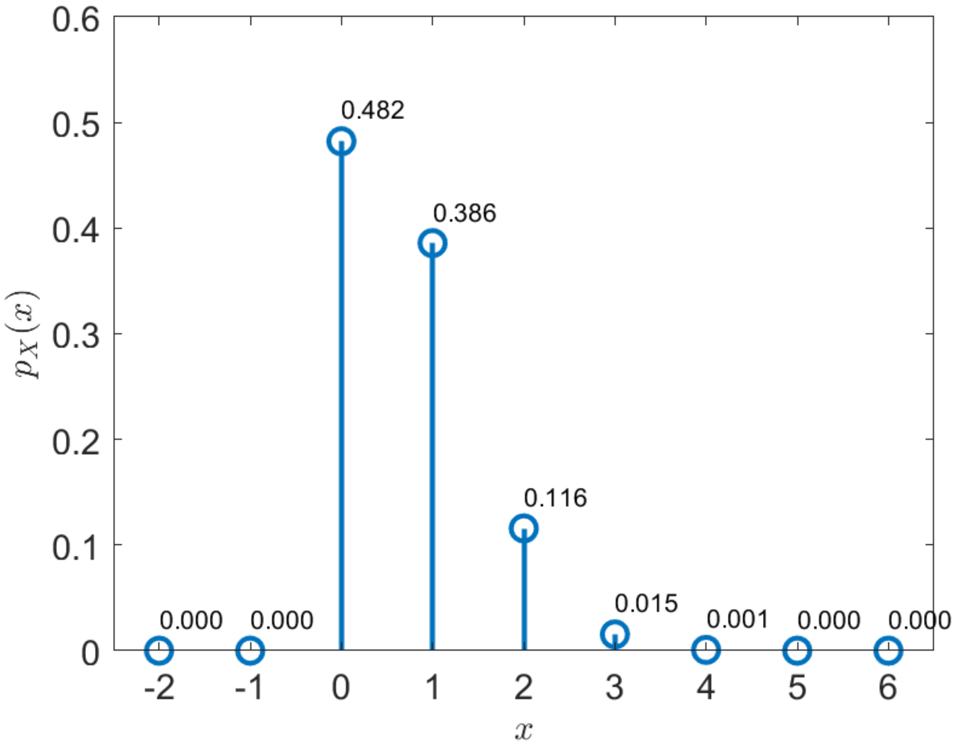
X~ Binomial(*n*, *1/6*)

Discrete RV: binomial random variable



X~ Binomial(*n*, *1/6*)

Discrete RV: binomial random variable



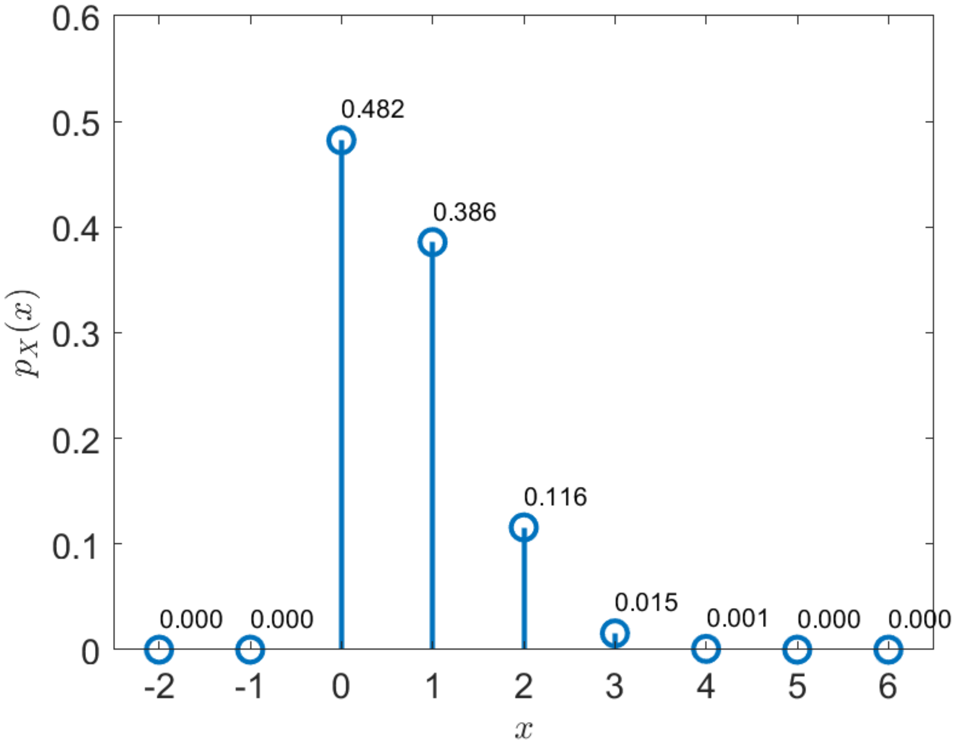
X~ Binomial(*4*, *1/6*)

Roll the dice *4* times and let X be the number of 2s

Example

We perform genetic testing on 4 embryos to test whether the newborn will have green eyes. Suppose the probability for a newborn of carrying the genes for green eyes is 1/6, and define the random variable X as the number of embryos carrying green-eyes genes. How is X distributed?

Example: binomial random variable

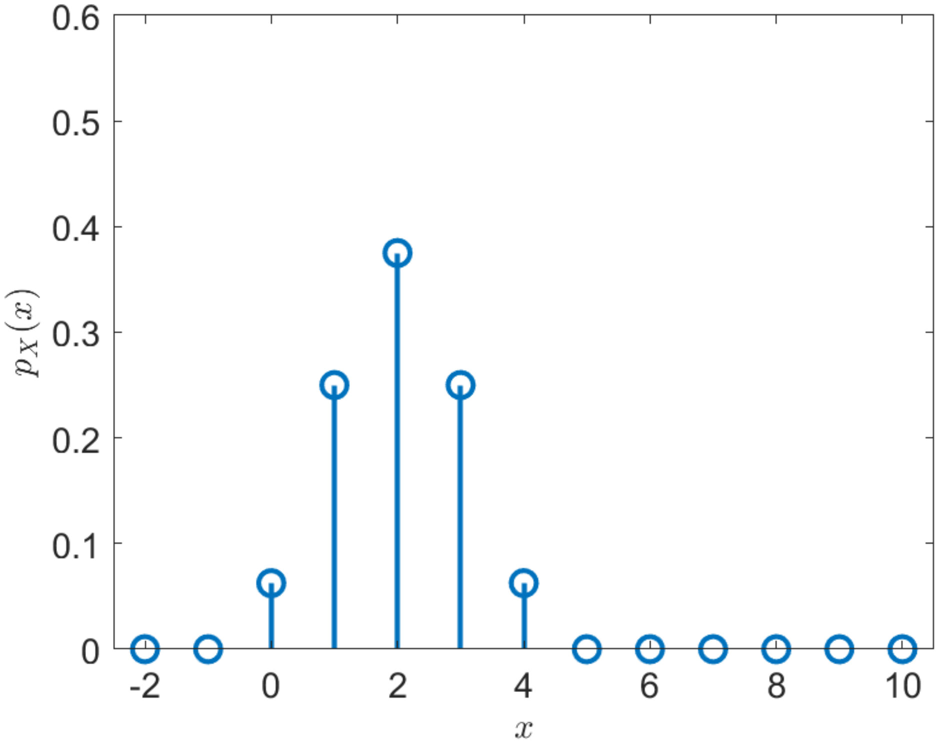


X~ Binomial(*4*, *1/6*)

Example

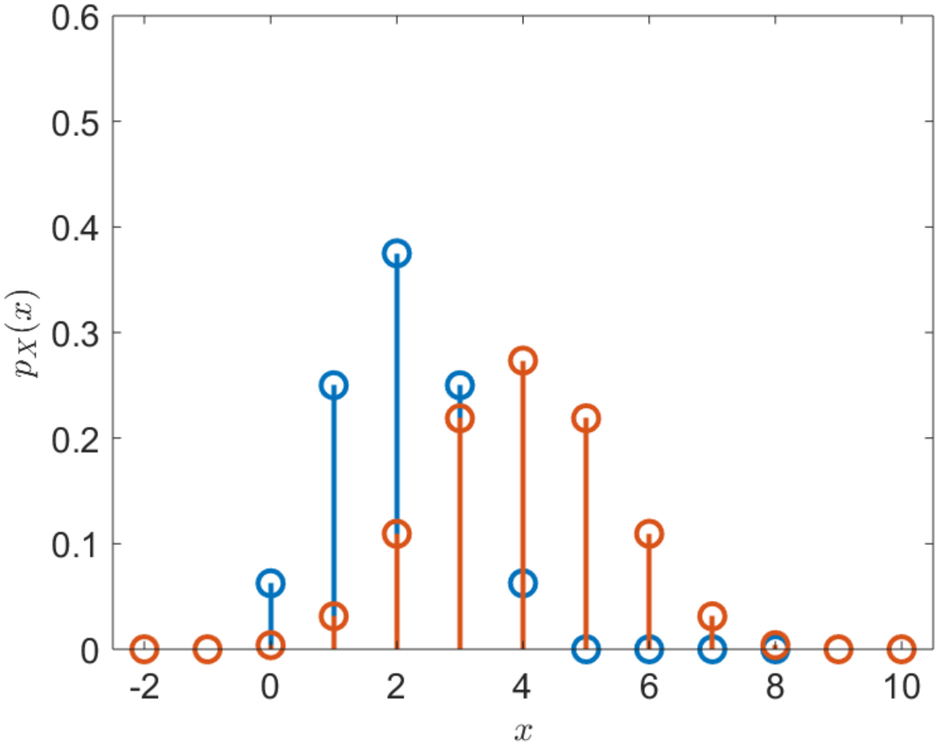
We repeat the previous test, but now for brown-eyes genes, which have a probability of 1/2. How is X distributed?

Example: binomial random variable



Binomial(*4*, *1/2*)

Example: binomial random variable



X~ Binomial(*8*, *1/2*)

Example continuous RV: Gaussian random variable

Arrival time of professor in the classroom. Typically, the professor arrives within 5 minutes from 9.00 o’clock.

T = {arrival time of the professor}

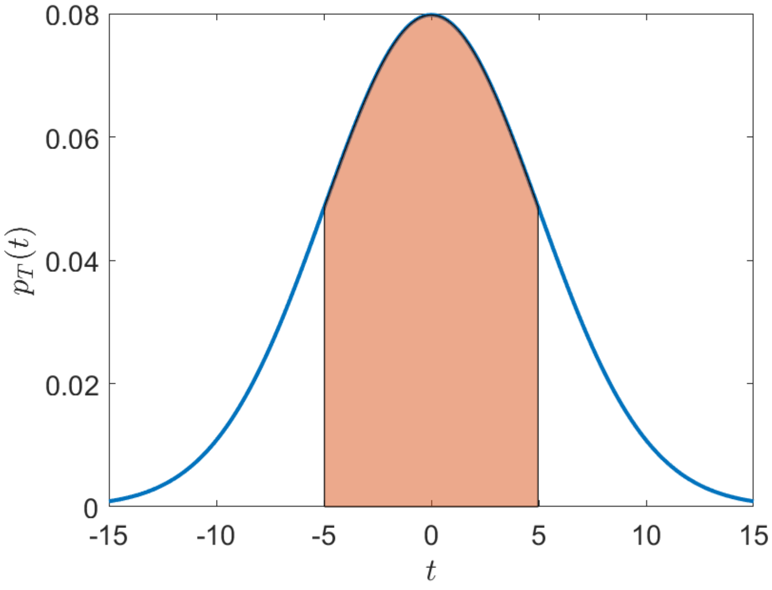
- t given as difference in minutes from a 9.00
- Assumption of Gaussian distribution

Example: Gaussian (or normal) random variable

The random variable T is normally distributed as T~ $N\ (0, 5^2)$

Gaussian distribution $N\ (\mu, \sigma^2)$

$$p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

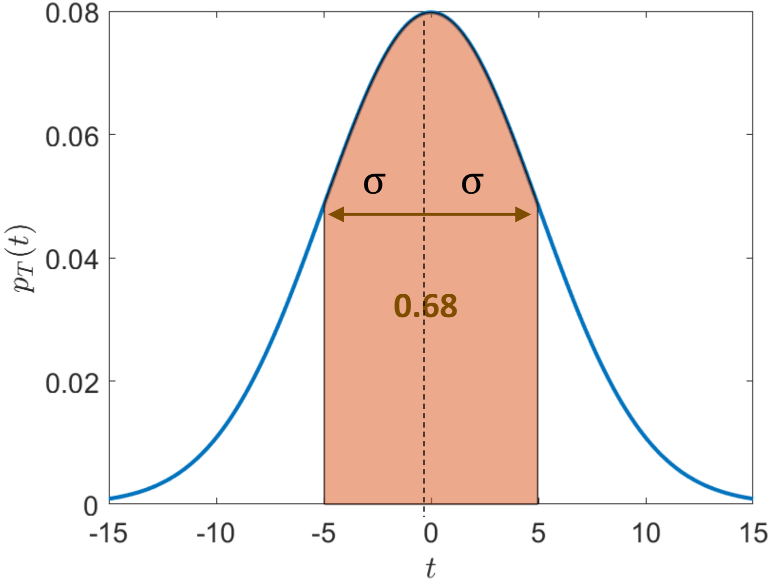


Example: Gaussian (or normal) random variable

The random variable T is normally distributed as T~ $N\ (0, 5^2)$

- $\Pr[-5 < \mathrm{T} < 5] = \Pr[-\sigma < \mathrm{T} < \sigma] \sim 68.2\%$
- $\Pr[-10 < \mathrm{T} < 10] = \Pr[-2\sigma < \mathrm{T} < 2\sigma] \sim 95.4\%$
- $\Pr[-15 < \mathrm{T} < 15] = \Pr[-3\sigma < \mathrm{T} < 3\sigma] \sim 99.7\%$

$$\Pr[x_1 < X < x_2] = \int_{x_1}^{x_2} p_X(x)dx$$

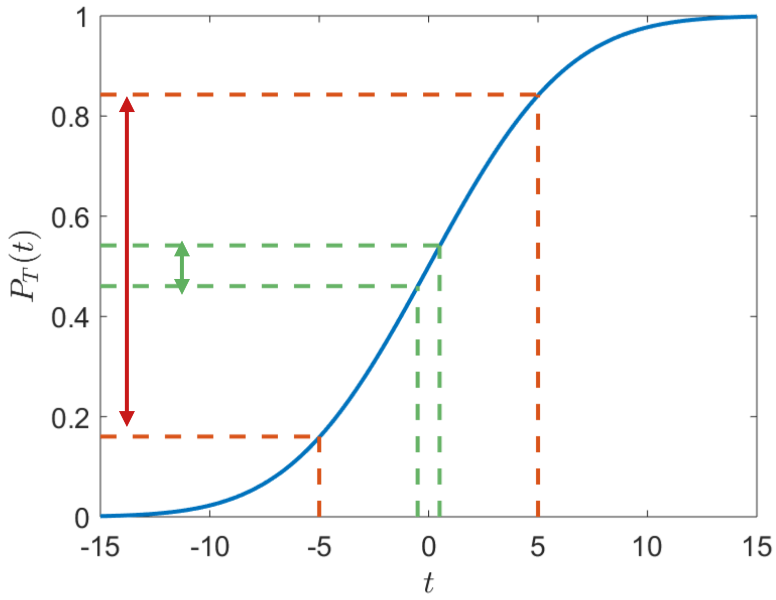


Example: Gaussian (or normal) random variable

The random variable T is normally distributed as T~ $N\ (0, 5^2)$

- $P_X(5) - P_X(-5) \sim 0.68$
- $P_X(5) - P_X(-5) > P_X(0.5) - P_X(-0.5)$

$$\Pr[x_1 < X < x_2] = \int_{x_1}^{x_2} p_X(x)dx = P_X\ (x_2) - P_X(x_1)$$



TU/e

Families of random variables

An overview of the most common families of discrete and continous RVs is available at:

- https://spseducation.tue.nl/courses/5cta0/mathematicalbackground_probability_families/ [no longer mantained]
- Lecture notes

***Note**: There is no need to learn all the formulas, but important to understand what are families of random variables and how they are described and used*

Wrap up (I)

- Random variables are way to map outcomes of random processes into **real numbers**
- **Discrete** random variables can take on a countable list of values (discrete range)
- **Continuous** random variables can take on any value in a continuous range
- **Probability distributions** describe the probability associated to each value that a random variable can take

Wrap up (II)

- **Moments** of a random variables are obtained from the probability distribution and are thus ***parameters of the probability model*** describing the random variable
- **Sample moments** are obtained from a sample of the random variable and are ***descriptive statistics*** of the random variable
- Sample moments are a good **approximation** of the moments only for a very large number of samples

Wrap up (III)

- **Families of random variables** are powerful to describe random variables that are very different in nature but behave statistically in a similar way
- Changing the ***parameters*** of a family of random variables we can describe the same experiment under ***different conditions***

Statistical signal processing (5CTA0)

Lecture 1, part B

Lecturer: Simona Turco

Electrical Engingeering, Signal Processing Systems group