Lecture 4 – Basic Statistics

Bulat Ibragimov

bulat@di.ku.dk

Department of Computer Science
University of Copenhagen

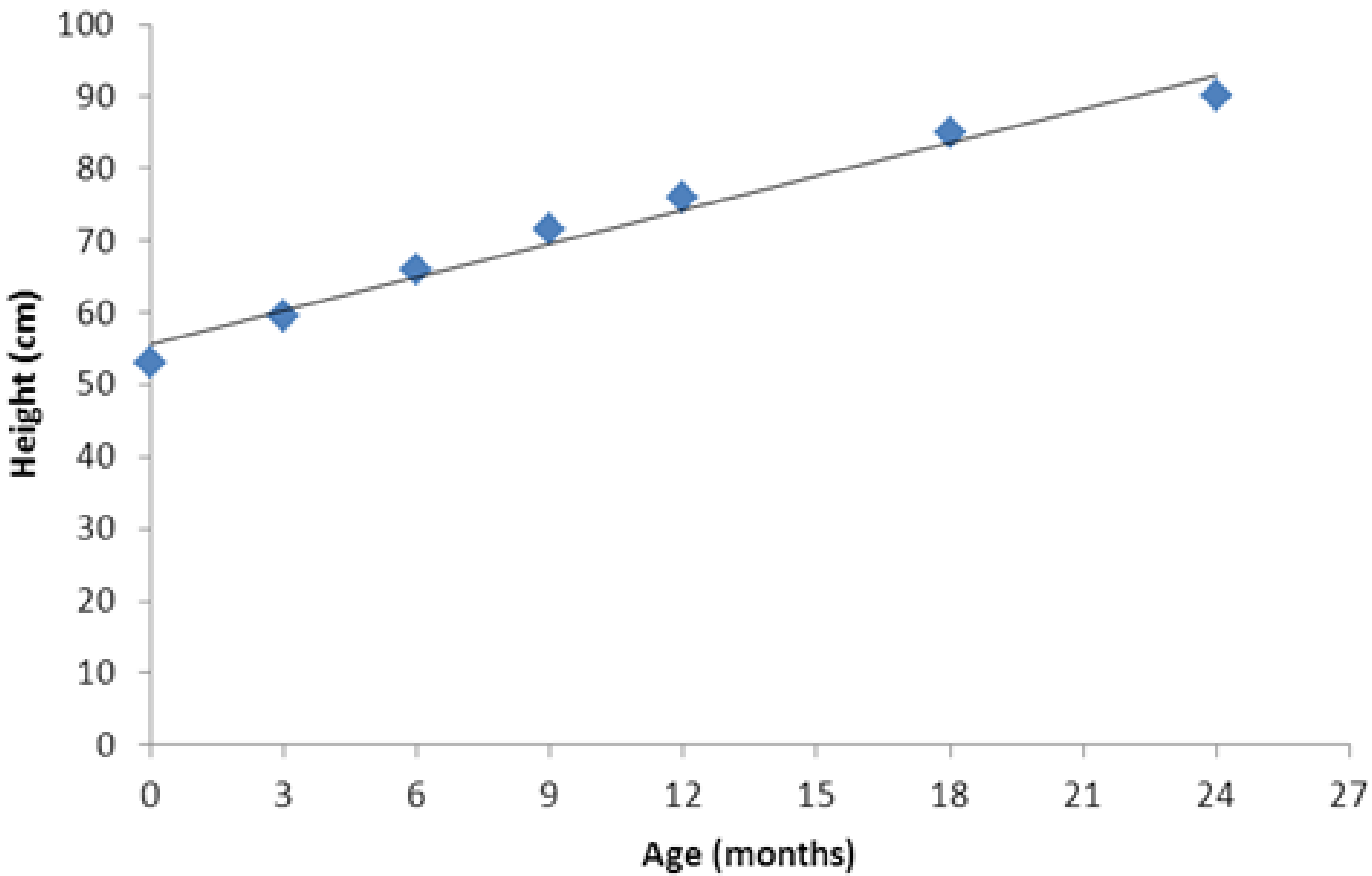
UNIVERSITY OF COPENHAGEN

Lecture X - today

Discrete random variables
Continues random variables
Mean and standard deviation
Bayes' rule
Simple distributions

Random events

- Height vs age in children
- The linear model does not fit perfectly. Why?



Random events

Children height depends on may factors:

- Genetics
- Diet
- Health
- etc.

Children height consists of deterministic and random parts

Discrete random variables

The total number of different outcomes is limited

Example:

- Tossing a coin

Outcomes:

- Ω ={Head (H), Tail (T)}

If the coin is fair, the probability of outcomes:

- P(Y = H) = 0.5
- P(Y = T) = 0.5

The sum probability of all outcomes is always one

Continues random variables

The total number of different outcomes is unlimited

Example:

- Throwing a coin on a round table to see how far from the center it will land

Outcomes:

- Ω =[0, R]

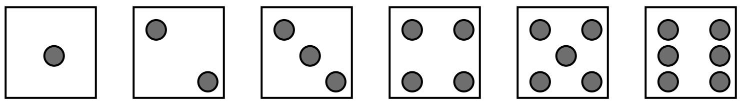
We cannot calculate probability for exact distance, but we can calculate probability for intervals

The probability for the complete interval [0, R] is again one

Adding probabilities

What is the probability of a die landing on x < 4?

Outcomes:

$\Omega = \{$  $\}$

Probability:

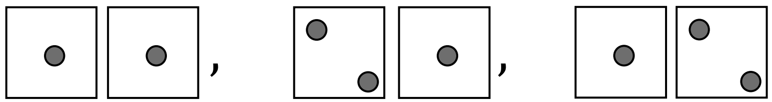
$$P(Y < 4) = P(Y = 1) + P(Y = 2) + P(Y = 3) = 1/6 + 1/6 + 1/6 = 0.5$$

Note that events should be **mutually exclusive**!

Adding probabilities

What is the probability of two dice landing on x < 4?

Outcomes – 36 combinations:

$\Omega = \{$  $\}$

Probability:

P(Y < 4) = P(Y = 1) + P(Y = 2) + P(Y = 3) = 0/36 + 1/36 + 2/36 = 0.08333

Note that outcome Y=3 can happen using different combinations of dice landing.

Conditional probabilities

Example:

- Probability that a person gets university degree is 0.6 ($X = 1$)
- Probability that a person with university degree gets a well-paid job is 0.7 $Y = 1$
- Probability that a person without university degree gets a well-paid job is 0.4

Conditional probabilities:

- $P(Y = y \mid X = x)$ – probability that $Y = y$ happens considering that $X = x$ happened
- $P(Y = 0 \mid X = 0) = 1 – 0.4 = 0.6$
- $P(Y = 0 \mid X = 1) = 1 - 0.7 = 0.3$
- $P(Y = 1 \mid X = 0) = 0.4$
- $P(Y = 1 \mid X = 1) = 0.7$

Joint probability

What is the probability that two random persons will get a university degree?

- These events are independent, so the joint probability is multiplication of individual probabilities:
  $$P(Y_1 = 1, Y_2 = 1) = P(Y_1 = 1) \cdot P(Y_2 = 1) = 0.6 \cdot 0.6 = 0.36$$

What is the probability that a person will get a university degree and a well-paid job?

- These events are dependent, we need to use conditional probabilities:
  $$P(Y = 1, X = 1) = P(Y = 1|X = 1) \cdot P(X = 1) = 0.7 \cdot 0.6 = 0.42$$

What are the probabilities of other scenarios for an arbitrary person?

Joint probability

$$P(Y = y, X = x) =$$
$$P(Y = y|X = x) \cdot P(X = x) = P(X = x|Y = y) \cdot P(Y = y)$$

Let's check what are the values of $P(Y = y|X = x)$ and $P(X = x|Y = y)$ for our example with university degrees and incomes?

	Degree	No degree
Well-paid	$0.6 \cdot 0.7$	$0.4 \cdot 0.4$
Low-paid	$0.6 \cdot 0.3$	$0.4 \cdot 0.6$

	Degree	No degree
Well-paid	0.42	0.16
Low-paid	0.18	0.24

Added, conditional and joint probability: example 1

Input:

- A woman was killed, and her husband is a suspect
- The husband was abusing the wife
- Defense attorney statement:
  - Only 0.01% of the men who abuse their wives end up murdering them
  - Therefore, the fact that Simpson abused his wife is irrelevant to the case (By irrelevant, he means that the probability of abusiveness importance is very low)
- Why is this a wrong use of conditional probability?

Added, conditional and joint probability: example 2

Input:

- You have a database of all life events for each Dane for this day
- You found a person who had a flat tire and lost 100DKK in the same day
- The probability of having a flat tire is 10^{-5}
- The probability of losing 100DKK is 10^{-7}
- The events are independent (joint probability = 10^{-12}), so you deduce that there is a conspiracy against this person
- Is this a correct deduction?

Added, conditional and joint probability: example 2

Issues:

- You did not check a specific person, but checked all Danes until you find a suitable one
- You first found a specific person, instead of first defining "bad"-events of interest (**Multiple comparisons problem**)
- Basically, you were looking for any person who experienced two arbitrary "bad" events in the database

- What is the probability of finding such a person by a coincidence?

Added, conditional and joint probability: example 2

Probability of >one "bad" events happening for a selected person:

- First, we need a table with statistics of all events we consider "bad":
  - Flat tire = 10^{-5}
  - Losing >=50DKK = 10^{-6}
  - Injury = $2 \cdot 10^{-5}$
  - Etc.
- Second, we compute the statistics that none of the events happen. The events are independent so their joint probability is the multiplication (we use the rule: $P(z) = 1 - P(\bar{z})$):

$$\text{P(nothing "bad")} = \underbrace{(1 - 10^{-5})}_{\text{Tire is alright}}\underbrace{(1 - 10^{-6})}_{\text{Did not lose any banknote}}\underbrace{(1 - 2 \cdot 10^{-5})}_{\text{No injury}}\cdots = 0.99$$

Added, conditional and joint probability: example 2

Probability of >one "bad" events happening for a selected person:

- Third, we compute that one "bad" event happen. Can we add probabilities of individual events?

$$\text{P(one "bad" event)} = 10^{-5} + 10^{-6} + 2 \cdot 10^{-5} + \cdots$$

- We cannot add probabilities, because the events are not **mutually exclusive**!

$$\begin{aligned}\text{P(one "bad" event)} = {} & 10^{-5}(1 - 10^{-6})(1 - 2 \cdot 10^{-5}) \ldots + \\ & (1 - 10^{-5})10^{-6}(1 - 2 \cdot 10^{-5}) \ldots + \\ & (1 - 10^{-5})(1 - 10^{-6})2 \cdot 10^{-5} \ldots + \cdots = 0.00998\end{aligned}$$

Added, conditional and joint probability: example 2

Probability of >one "bad" events happening for a selected person:

- Finally:

P(>one "bad" event) = 1 – (P(nothing "bad") + P(one "bad" event))=

1 – (0.99 + 0.00998) = **0.00002**

What is the probability that we will find at least one such a person in the database?

- We need to compute the probability that there is no such a person, and subtract it from 1. There are 5,814,461 people in the database, bad events occurring to them are independent:

P(at least one person found) =

$$1 - (1 - 0.00002)^{5,814,461} = 1 - 3 \cdot 10^{-51} \approx 1$$

- We will certainly find such a person in the database!

Bayes' rule

From joint probability formulation:

$$P(Y = y, X = 1) = P(Y = y|X = x) \cdot P(X = x) = P(X = x|Y = y) \cdot P(Y = y)$$

We can get Bayes' rule:

$$P(X = x|Y = y) = \frac{P(Y = y|X = x) \cdot P(X = x)}{P(Y = y)}$$

Bayes' rule

Returning to our example of university degrees and success (probability of university degree, on condition of well-paid job):

$$P(X = 1|Y = 1) = \frac{P(Y = 1|X = 1) \cdot P(X = 1)}{P(Y = 1)} = \frac{0.7 \cdot 0.6}{0.42 + 0.16} = 0.72$$

	Degree	No degree
Well-paid	$0.6 \cdot 0.7$	$0.4 \cdot 0.4$
Low-paid	$0.6 \cdot 0.3$	$0.4 \cdot 0.6$

	Degree	No degree
Well-paid	0.42	0.16
Low-paid	0.18	0.24

Bayes' rule: example

Input:

The probability of a certain medical test being positive is 90% if a patient has disease D. The 1% of the population have the disease and the test records a false positive 5% of the time. If a random person receives a positive test, what is the probability of D for him?

5 minutes to think

Bayes' rule: example

Input:

The probability of a certain medical test being positive is 90%, if a patient has disease D. 1% of the population have the disease, and the test records a false positive 5% of the time. If a random person receives a positive test, what is the probability of D for him?

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+)}$$

$$P(D) = 0.01; \qquad P(+|D) = 0.9$$

$$P(+) = P(+|D) \cdot P(D) + P(+|no\ D) \cdot P(no\ D) = 0.0585$$

0.9 0.01 0.05 0.99

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+)} = \frac{0.9 \cdot 0.01}{0.0585} \approx 0.15$$

Expectation: mean

A calculating student wants to find a most prosperous job and compares two education specialties A and B. He lives in a small city and there is only one company in A and one in B, so he will have to go to a specific company after finishing.

He has got an access to the database of salaries for people working in company A and B:

Salaries in A = [6800, 3150, 2700, 4700, 7100, 5800, 2000]

Salaries in B = [5500, 4500, 3900, 3800, 4800, 4500, 5900]

How to estimate the optimal strategy?

Mean value –

$$\mathbf{E}_{P(x)}\{X\} = \sum_x xP(x)$$

Expectation: mean

Salaries in A = [6800, 3150, 2700, 4700, 7100, 5800, 2000]
Salaries in B = [5500, 4500, 3900, 3800, 4800, 4500, 5900]

Mean value for company A:

$$\mathbf{E}_{P(x)}\{X\} = \sum_{x} xP(x) =$$

$$6800\frac{1}{7} + 3150\frac{1}{7} + 2700\frac{1}{7} + 4700\frac{1}{7} + 7100\frac{1}{7} + 5800\frac{1}{7} + 2000\frac{1}{7} = \mathbf{4607}$$

Mean value for company B:

$$\mathbf{E}_{P(y)}\{Y\} = \sum_{y} yP(y) =$$

$$5500\frac{1}{7} + 4500\frac{1}{7} + 3900\frac{1}{7} + 3800\frac{1}{7} + 4800\frac{1}{7} + 4500\frac{1}{7} + 5900\frac{1}{7} = \mathbf{4700}$$

Expectation: variance

Although the means for A and B are relatively similar, the actual salaries in A and B behave differently, in A salaries are in [2000:7100] while in B salaries are in [3800:5900] intervals

Salaries in A = [6800, 3150, 2700, 4700, 7100, 5800, 2000]
Salaries in B = [5500, 4500, 3900, 3800, 4800, 4500, 5900]

Variance value:

$$\mathrm{var}\{X\} = \mathbf{E}_{P(x)}\left\{\left(X - \mathbf{E}_{P(x)}(X)\right)^2\right\} =$$

$$\sum_{y}\left(x - \mathbf{E}_{P(x)}(X)\right)^2 P(x)$$

Expectation: variance

Salaries in A = [6800, 3150, 2700, 4700, 7100, 5800, 2000]
Salaries in B = [5500, 4500, 3900, 3800, 4800, 4500, 5900]

Variance value for company A:

$$\mathrm{var}\{X\} = \mathbf{E}_{P(x)}\left\{\left(X - \mathbf{E}_{P(x)}(X)\right)^2\right\} = 3573163$$

Variance value for company B:

$$\mathrm{var}\{Y\} = \mathbf{E}_{P(y)}\left\{\left(Y - \mathbf{E}_{P(y)}(Y)\right)^2\right\} = 517143$$

Standard deviation

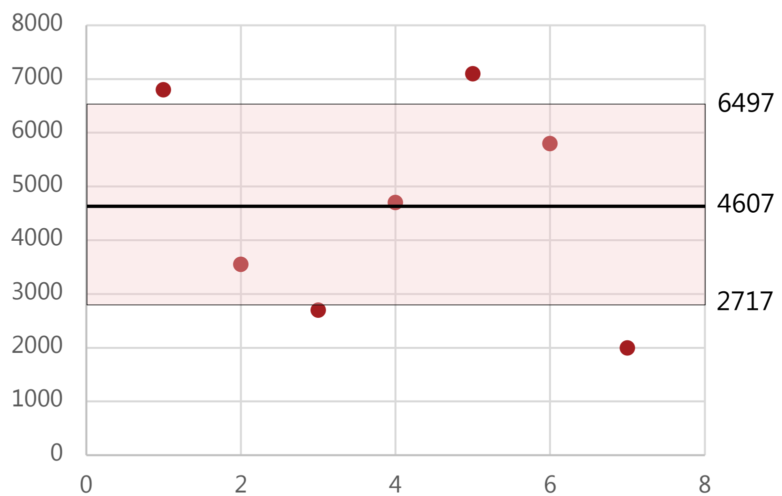
Standard deviation (SD) is the square root of variance.

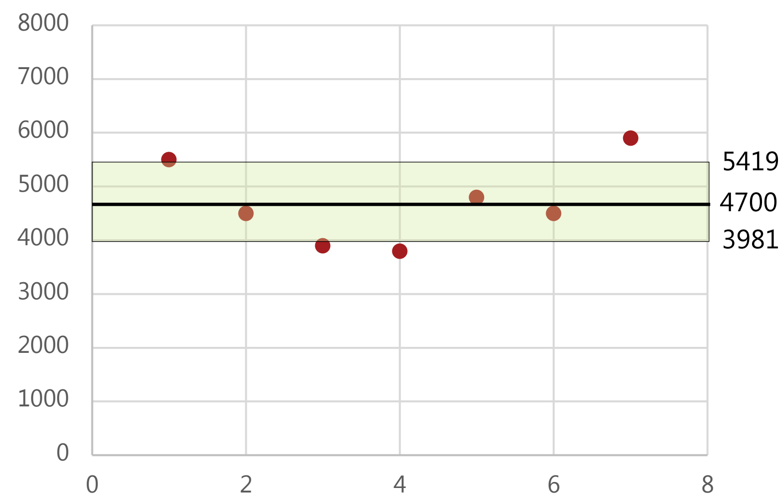
Standard deviation for company A:

$$\sigma_X = \mathbf{E}_{P(x)}\left\{\left(X - \mathbf{E}_{P(x)}(X)\right)^2\right\}^{0.5} = 1890$$

Standard deviation for company B:

$$\sigma_Y = \mathbf{E}_{P(y)}\left\{\left(Y - \mathbf{E}_{P(y)}(Y)\right)^2\right\}^{0.5} = 719$$





Expectation: vector form variables

Mean and variance can be computed for random variables in the vector form.

Let's say we have a set of students that passed math and English exams. How does a random student of such population look like?

	Math	English
Student 1	**80**	**40**
Student 2	**60**	**80**
Student 3	**50**	**70**
Student 4	**40**	**70**
Student 5	**20**	**90**
Student 6	**50**	**70**

Mean grades =
[50, 70]

Variance for individual grades =
[333.3, 233.3]

SD for individual grades =
[18.3, 15.3]

Expectation: vector form variables

For vector form random variables we can also compute covariance matrix (pairwise variances between induvial components):

$$\mathrm{cov}\{x\} = \mathbf{E}_{P(x)}\left\{\left(x - \mathbf{E}_{P(x)}\{x\}\right)\left(x - \mathbf{E}_{P(x)}\{x\}\right)^T\right\}$$

Matrix of [6x2] size

Mean grades = **[50, 70]**

	Math	English
Student 1	**80**	**40**
Student 2	**60**	**80**
Student 3	**50**	**70**
Student 4	**40**	**70**
Student 5	**20**	**90**
Student 6	**50**	**70**

Math	English
80-50	**40-70**
60-50	**80-70**
50-50	**70-70**
40-50	**70-70**
20-50	**90-70**
50-50	**70-70**

Math	English
30	**-30**
10	**10**
0	**0**
-10	**0**
-30	**20**
0	**0**

Expectation: vector form variables

For vector form random variables we can also compute covariance matrix (pairwise variances between induvial components):

$$\mathrm{cov}\{x\} = \mathbf{E}_{P(x)}\left\{\left(x - \mathbf{E}_{P(x)}\{x\}\right)\left(x - \mathbf{E}_{P(x)}\{x\}\right)^T\right\}$$

30	-30
10	10
0	0
-10	0
-30	20
0	0

X

30	10	0	-10	-30	0
-30	10	0	0	20	0

=

2000	-1400
-1400	1400

333	-233
-233	233

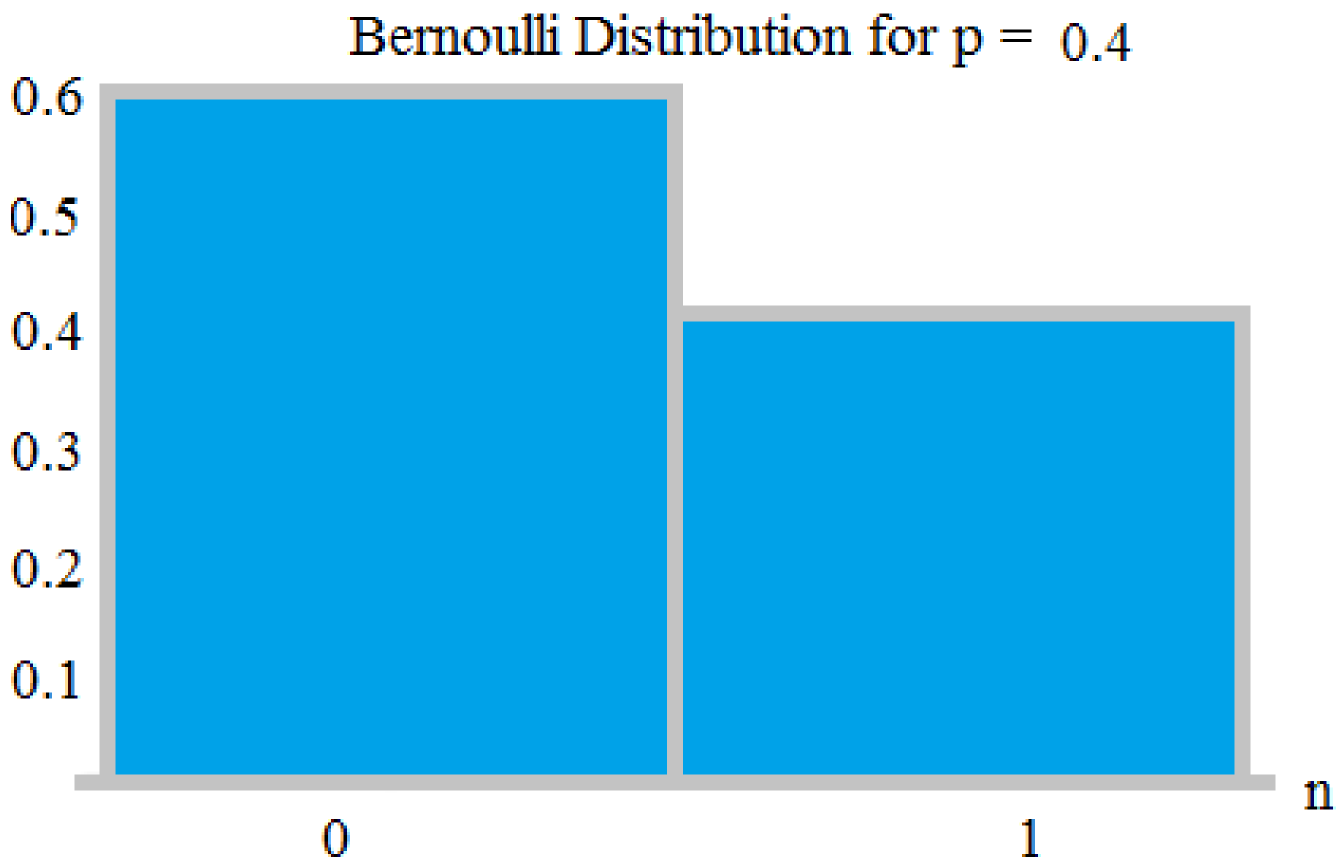
Normalized to # students

What is the meaning of the covariance matrix?

Simple distributions: Bernoulli distribution

Coin tossing is a good example

$$P(X = x) = p^x * (1-p)^{1-x}$$



Simple distributions: Binominal distribution

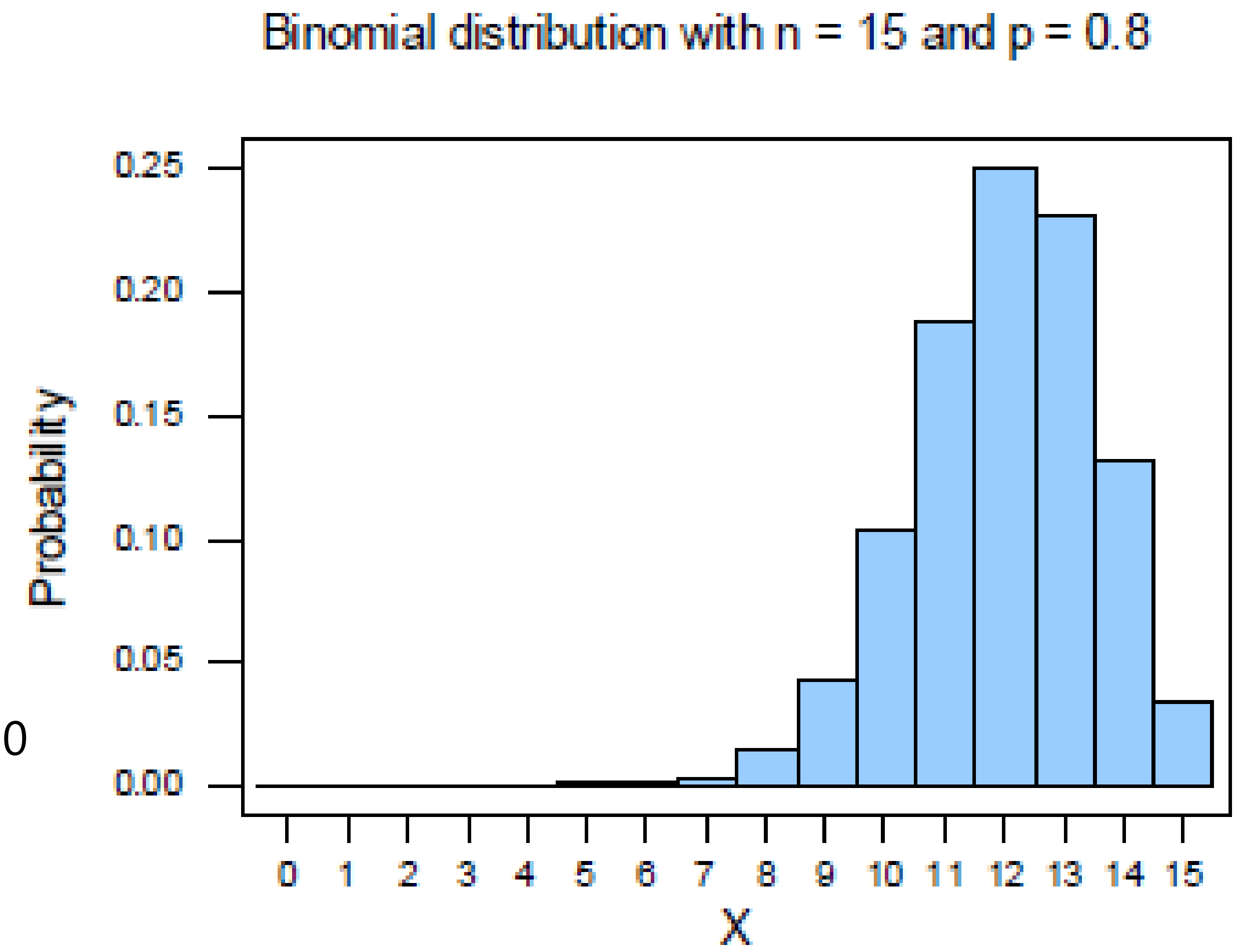
We toss a coin N times, what is the probability of getting x tails?

$P(X = x) =$

$$\binom{n}{x} p^x q^{n-x}$$

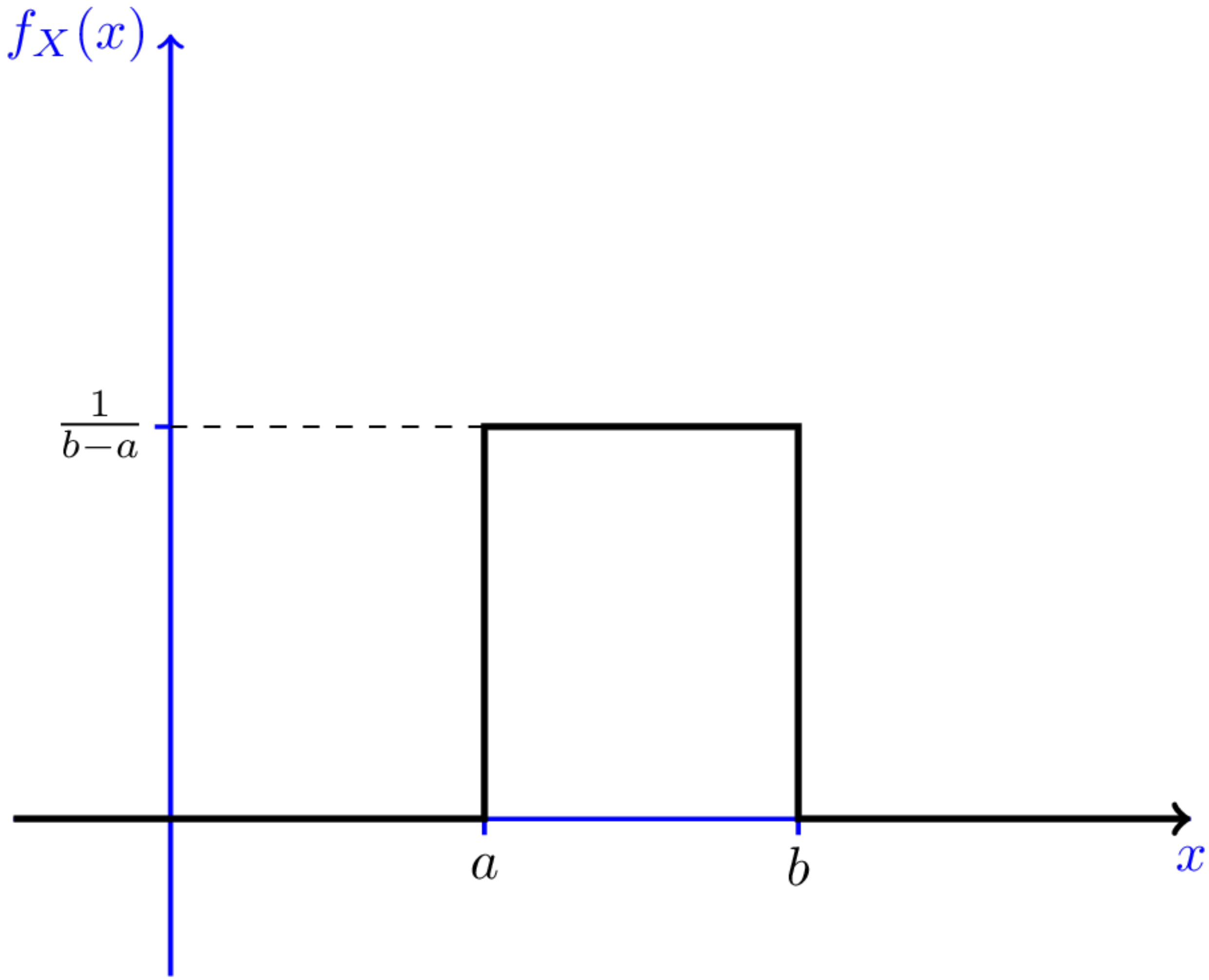
*q = 1 - p

Let's say we toss a coin 10 times, what is the probability of 5 tails?



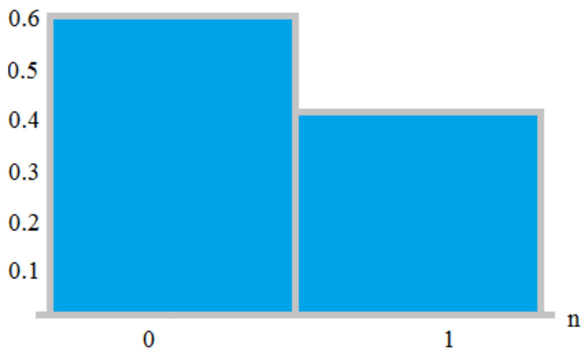
Simple distributions: Uniform distribution

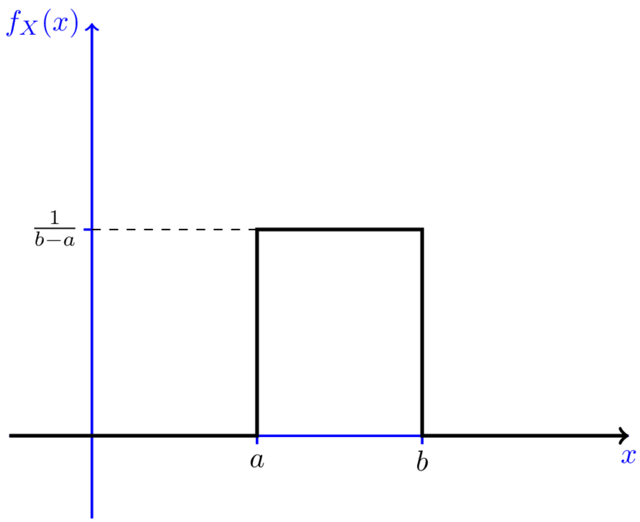
Let's say we choose a random real number between a and b

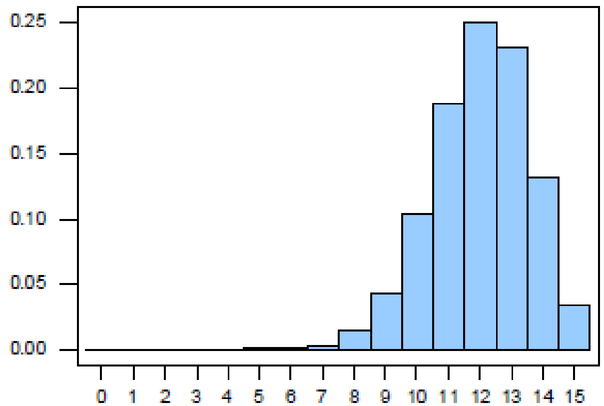


Probability density function

PDF estimates the likelihood that the value of the random variable would be equal to a specified number





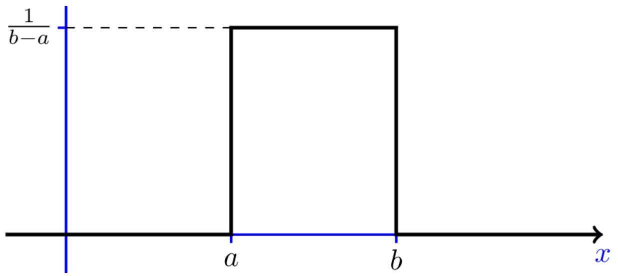


Simple distributions: Uniform distribution

Let's say we choose a random real number between 0 and 1:

- What is the probability of getting a specific number like 0.23423432432?

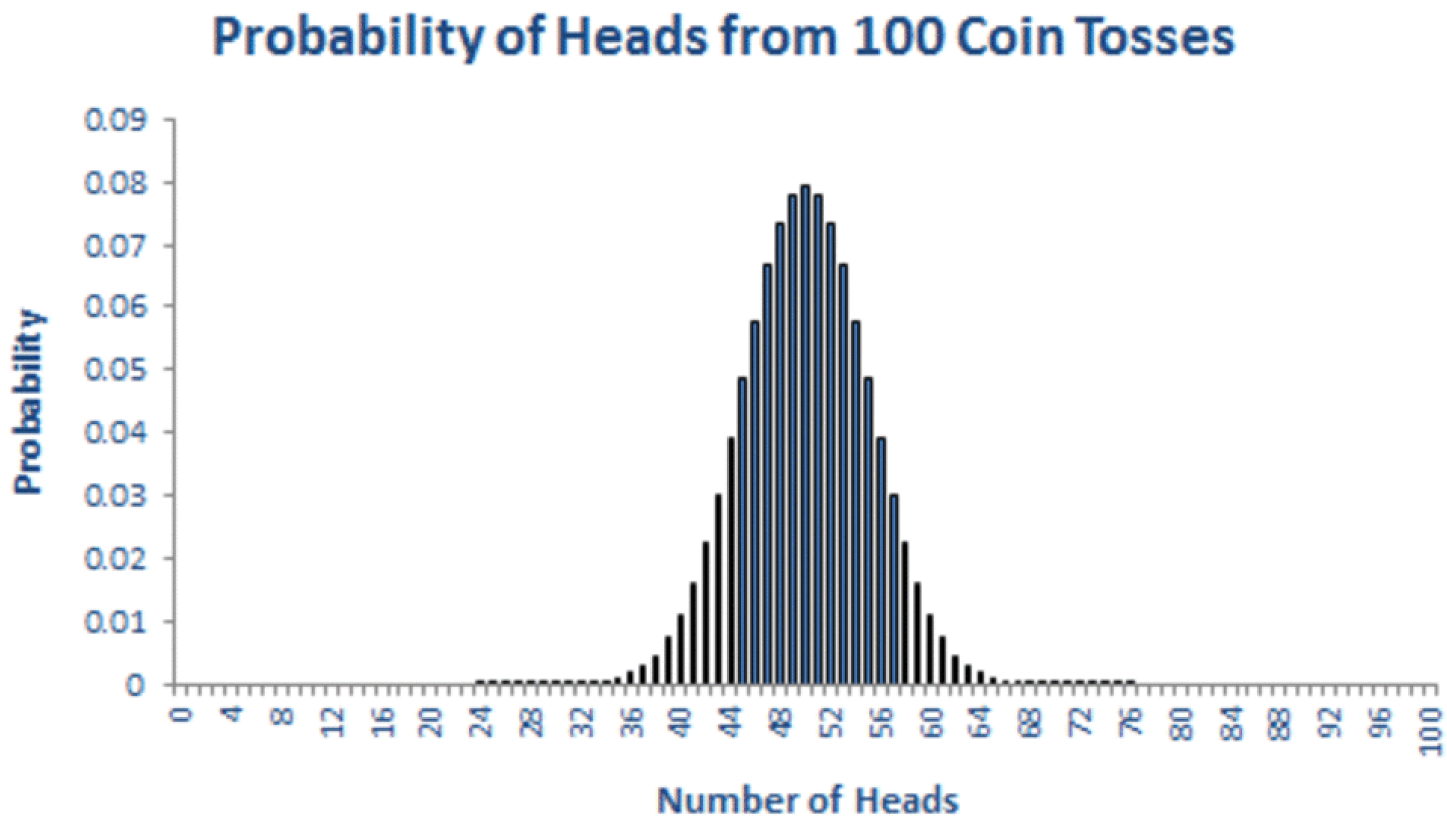
- The probability of getting a specific real number is zero.
- But we can compute the probability of getting a number in an interval from [0.2, 0.3]



$$P(0.2 \leq x \leq 0.3) = \int_{0.2}^{0.3} \frac{1}{1-0}\,dx = \frac{1}{1}(0.3-0.2) = 0.1$$

Simple distributions: Normal distribution

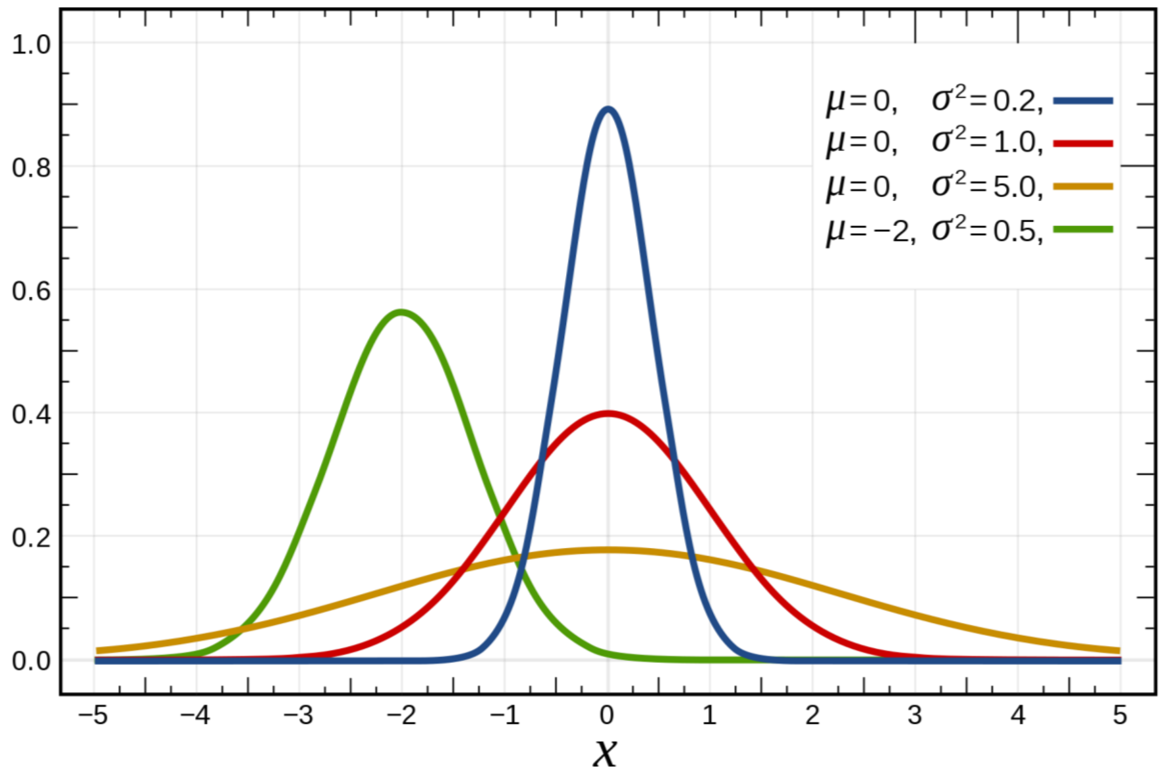
If we toss coin 100 times and count heads:



Simple distributions: Normal distribution

Normal distribution:

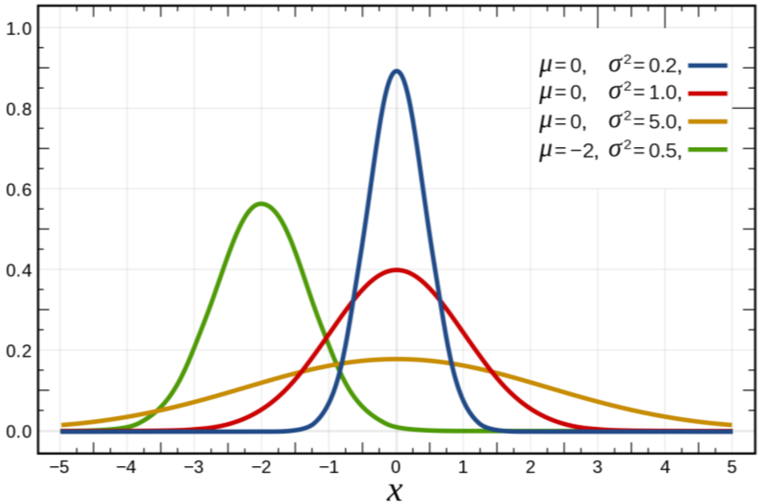
$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



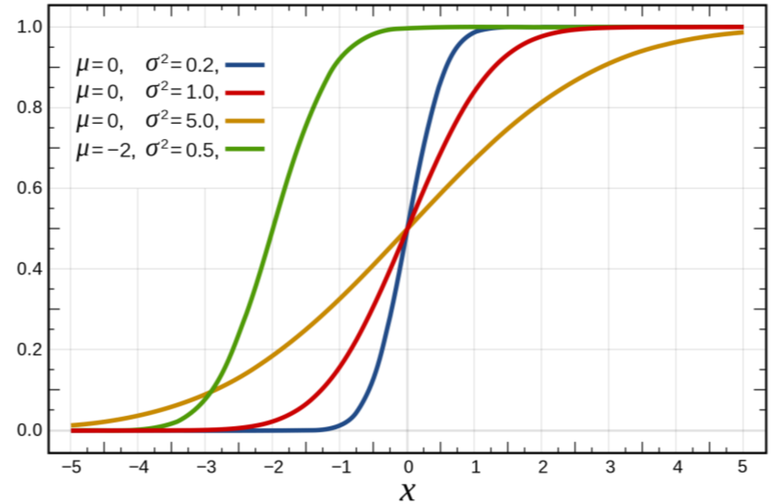
Cumulative distribution function

Plot probability of getting $x \leq t$:

$$P(x \leq t) = \int_{-\infty}^{t} f(x)\, dx$$



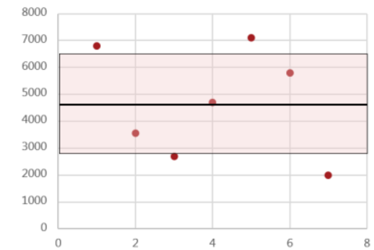
Probability density function



Cumulative distribution function

Calculating student example 2

The same formulation, but now the student lives in a big city and there are N companies in field A=computer science and N in B=physics. This time he does not have lists of salaries, but knows the means and standard deviations:



- Company in CS – $\{\mu, s\} = [4607,\ 1890]$
- Company in Ph – $\{\mu, s\} = [4700,\ \ 719]$

Which field is better to choose from if:

- N = 1. The only one position is available in CS or Ph.
- N = 10. Ten positions are available in CS (or from Ph). If he rejects a position, he cannot return to it.

What could be the student's strategy?