Lecture 4 – Basic Statistics

Bulat Ibragimov

bulat@di.ku.dk

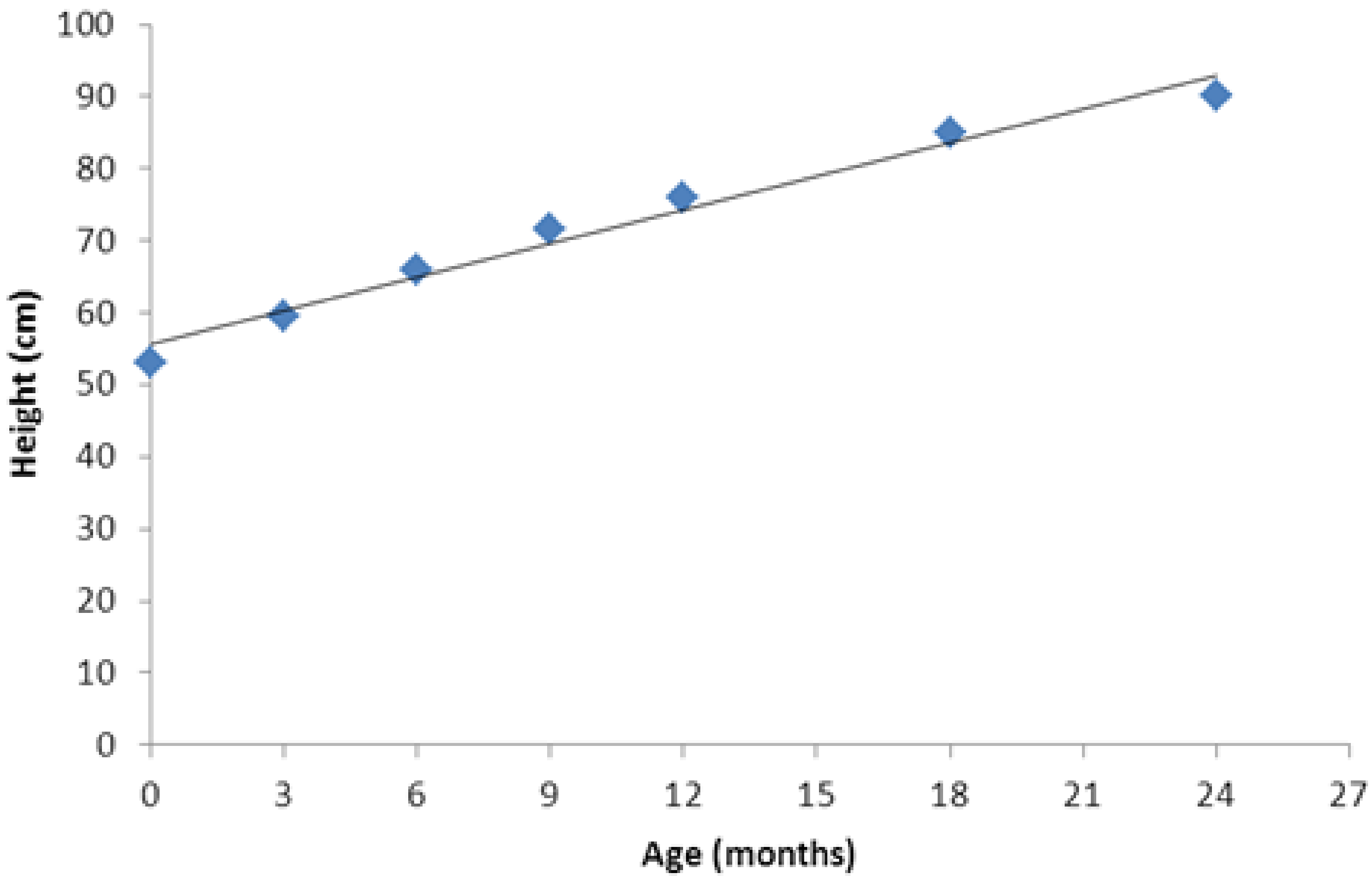
Department of Computer Science
University of Copenhagen

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:

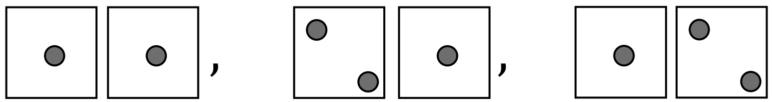
$$P(Y < 4) = P(Y = 1) + P(Y = 2) + P(Y = 3) = 1/6 + 1/6 + 1/6 = 0.5$$

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?

2 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?

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

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

$$P(+) = \underset{0.9}{P(+|D)} \cdot \underset{0.01}{P(D)} + \underset{0.05}{P(+|no\ D)} \cdot \underset{0.99}{P(no\ D)} = 0.0585$$

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

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

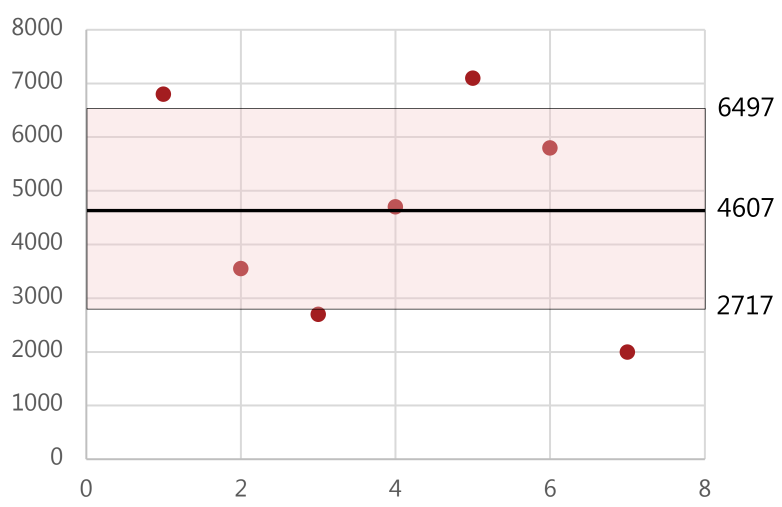
Standard deviation (SD) is the square root of variance.

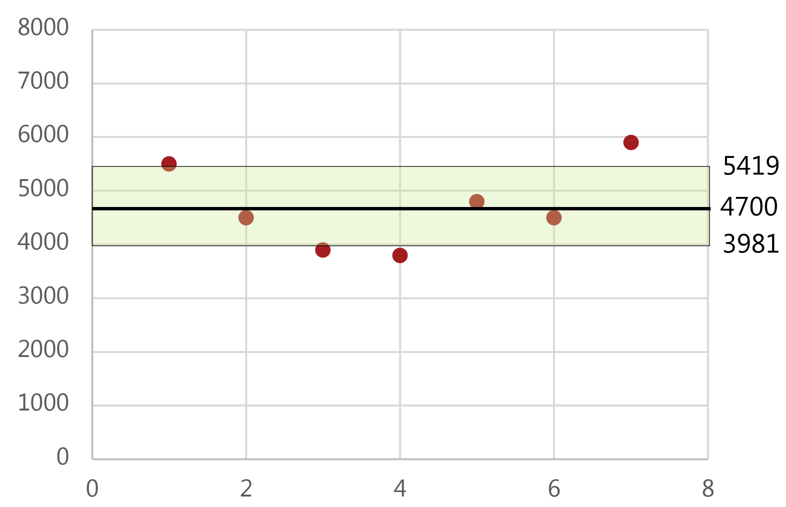
Standard deviation for company A:

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

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

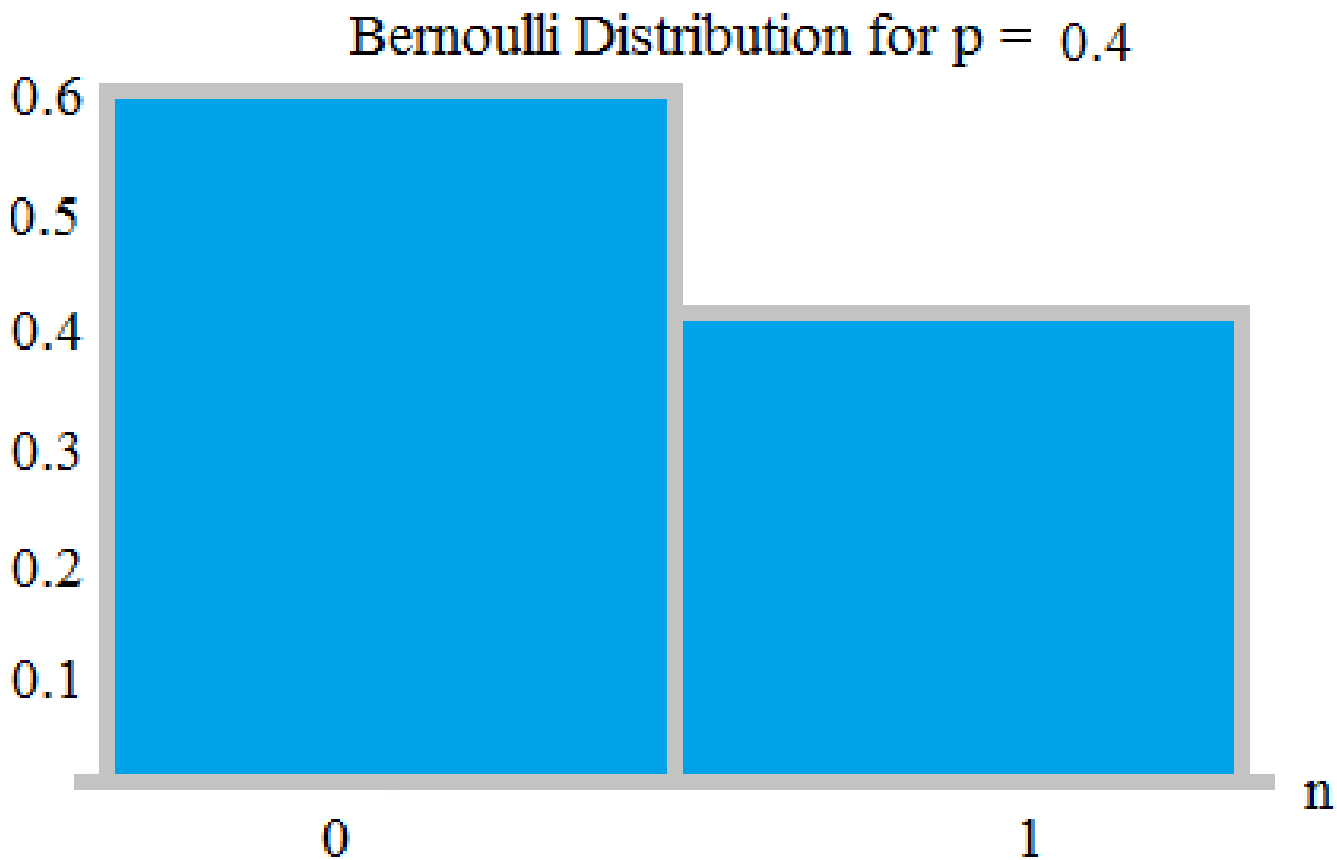
Normalized to # students

What is the meaning of the covariance matrix?

Simple distributions: Bernoulli distribution

Coin tossing is a good example

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



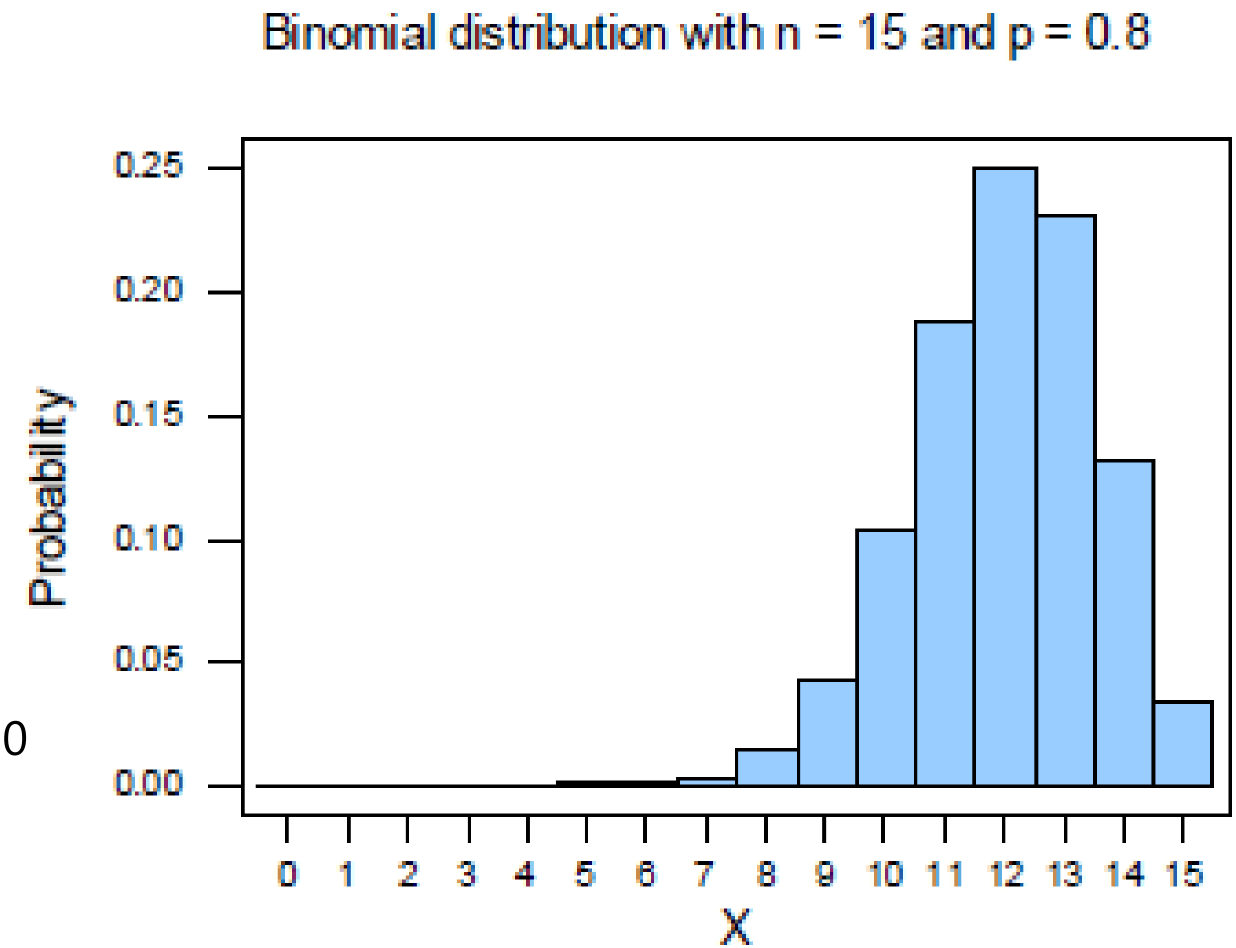
Simple distributions: Binominal distribution

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

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

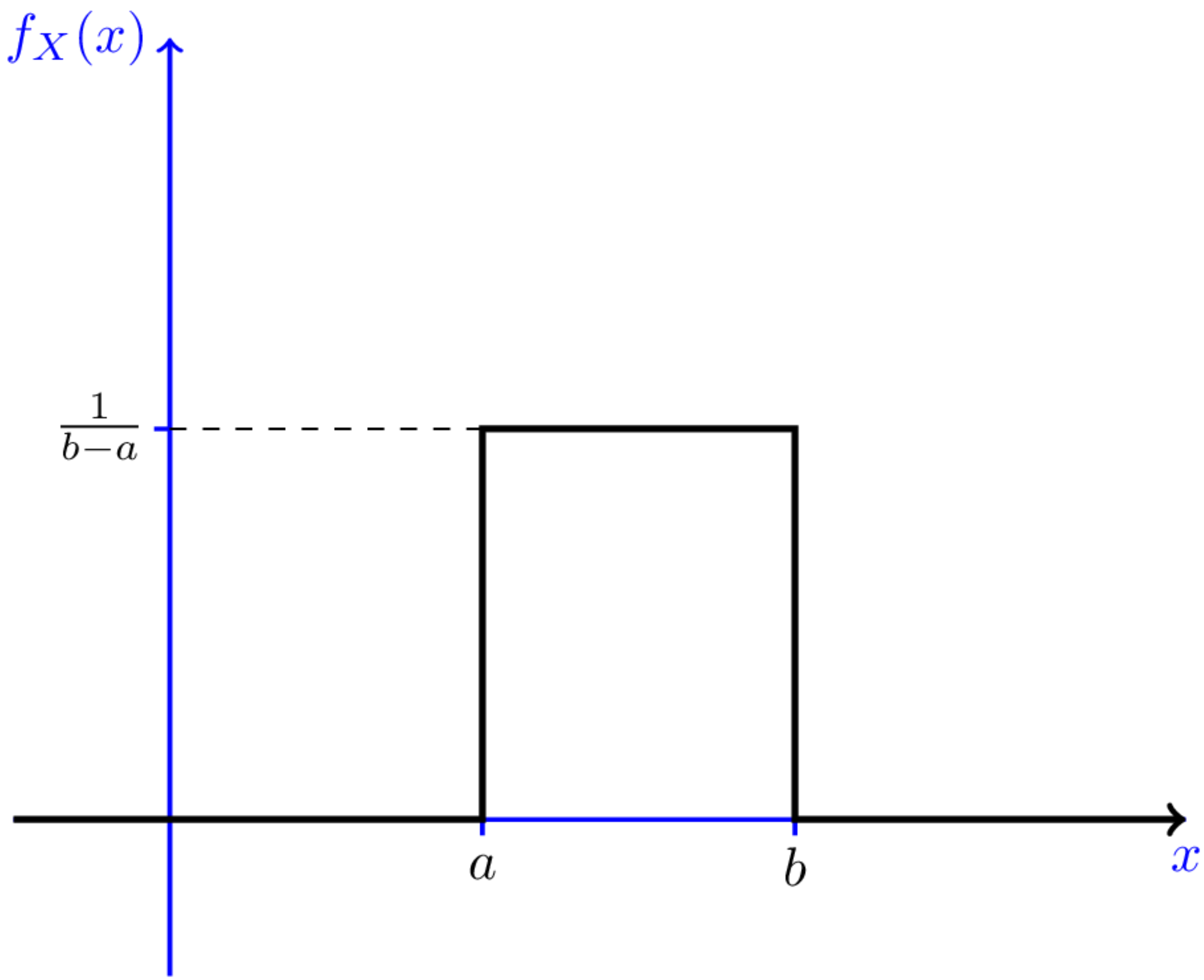
*q = 1 - p

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



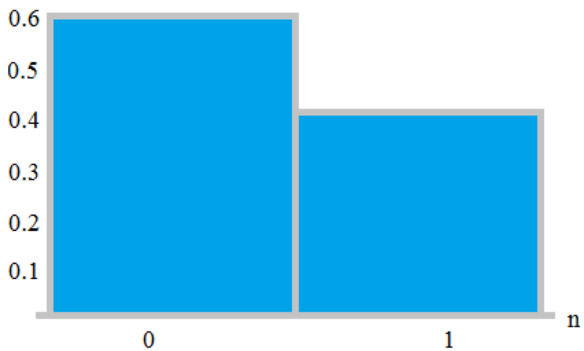
Simple distributions: Uniform distribution

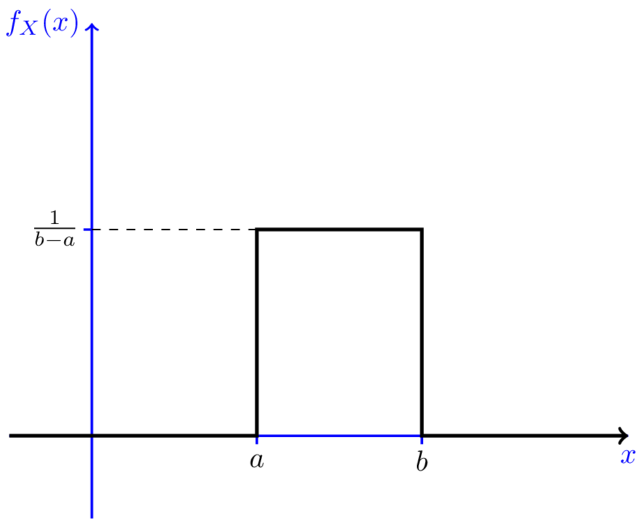
Let's say we choose a random real number between a and b

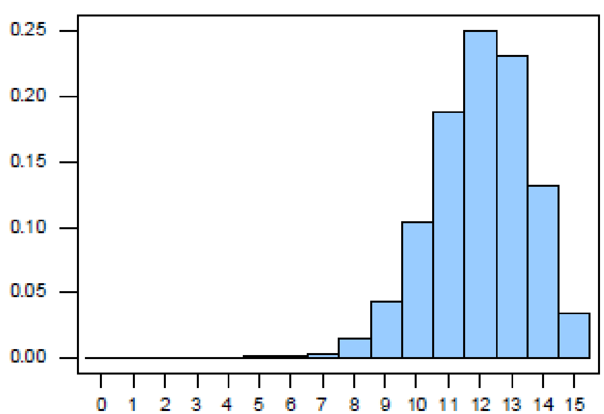


Probability density function

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





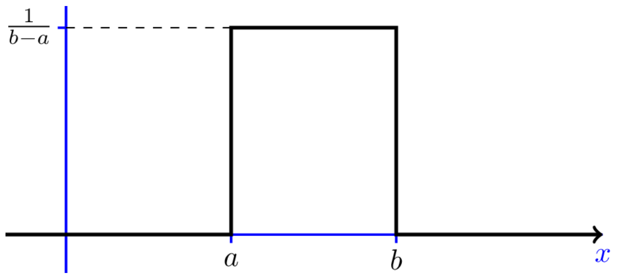


Simple distributions: Uniform distribution

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?

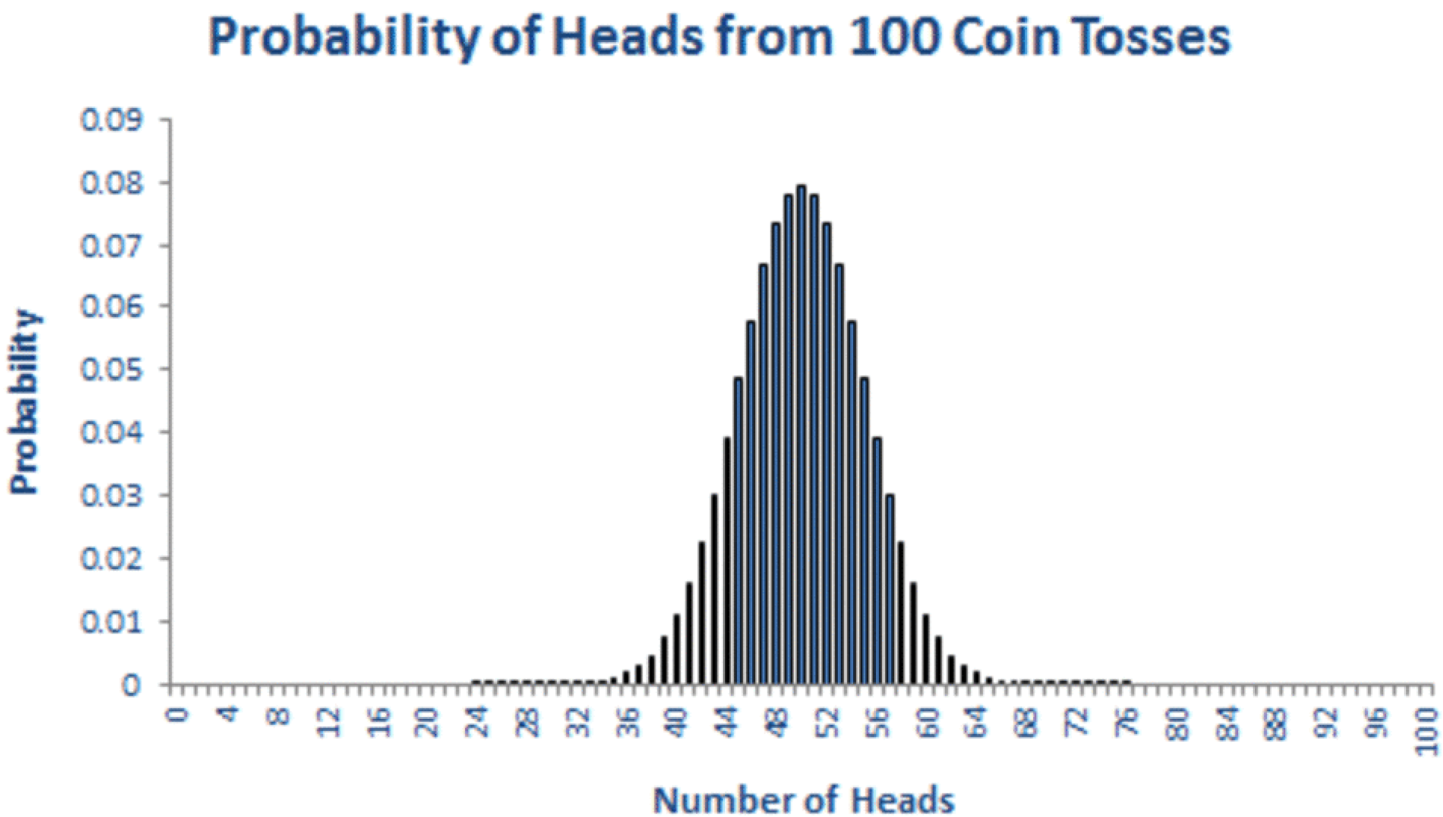
- The probability of getting a specific real number is zero.
- But we can compute the probability of getting a number in an interval from [0.2, 0.3]



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

Simple distributions: Normal distribution

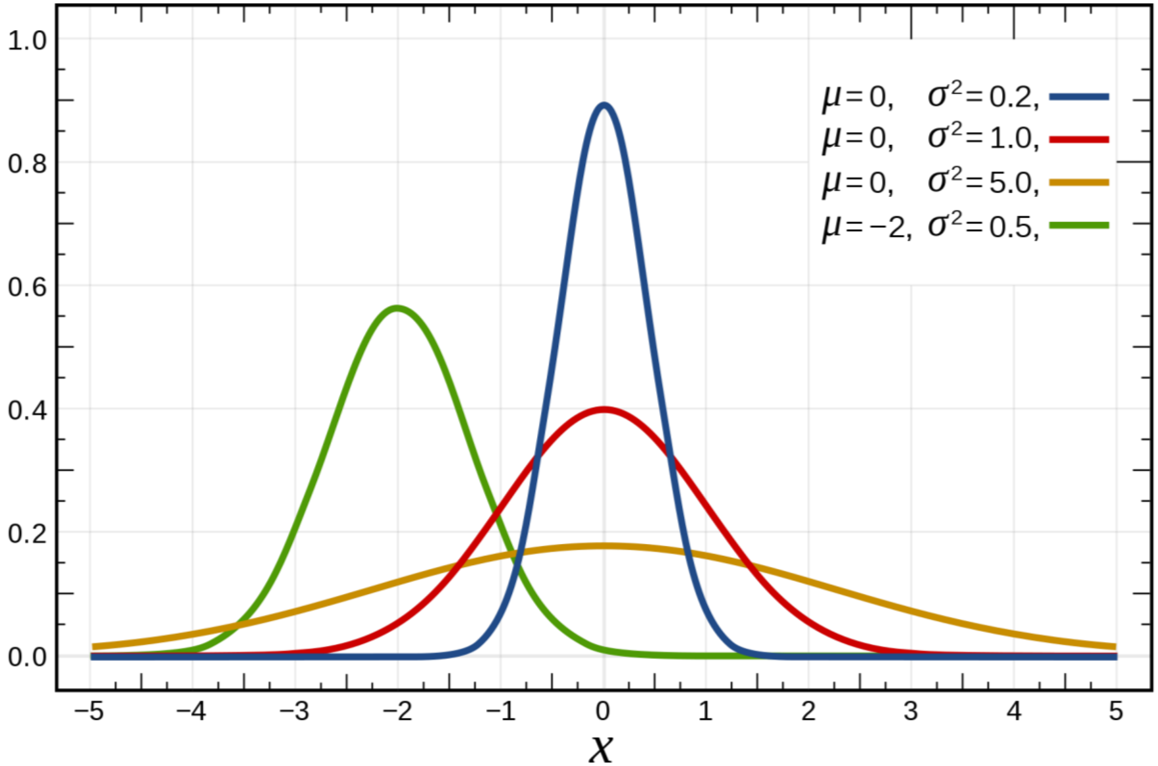
If we toss coin 100 times and count heads:



Simple distributions: Normal distribution

Normal distribution:

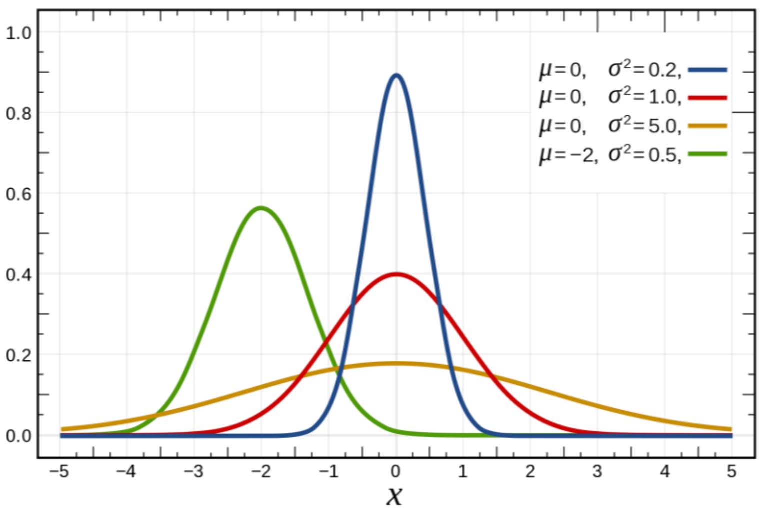
$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



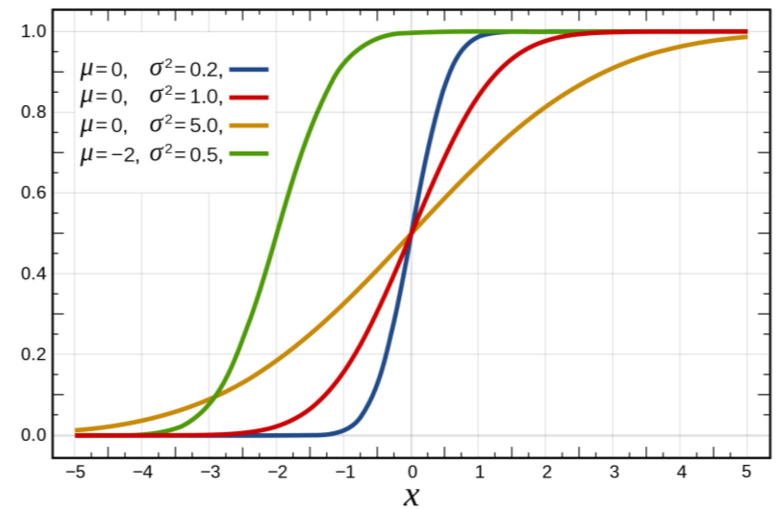
Cumulative distribution function

Plot probability of getting $x \leq t$:

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



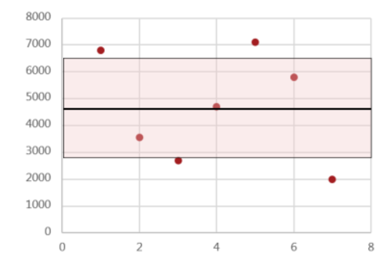
Probability density function



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:

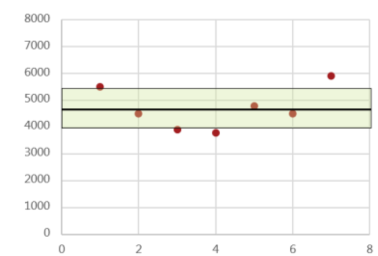


- Company in CS – $\{\mu, s\} = [4607,\ 1890]$
- Company in Ph – $\{\mu, s\} = [4700,\ \ 719]$

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?



Monte Carlo Simulation: Example

Suppose you just finished high school and ready to consider a career ever in software or finance. How to financially optimize the decision:

- Maximize the expected salary after finishing university
- Minimize the risks of being significantly underpaid
- Take your expected competitiveness into account
- Is it better to be more competitive or apply to more positions?

Monte Carlo Simulation: Example

STATISTICS DENMARK

SE

StatBank Denmark | Labour, income and wealth | Select fro

Download file as...

Excel (*.xlsx)

Codes in sep. columns

Incl. Footnotes etc.

Graphics

Line chart (interactive)

Sort data

Print

Earnings by sex, salary earners, salary, sector, tir occupation

	241 Finance professionals	25 Information and communications technology professionals
Men and women, total		
Employee group total		
Fixed salary-earners		
All sectors		
2019		
Lower quartile, standardized earnings	295.09	279.31
Median, standardized hourly earnings	346.10	345.94
Upper quartile, standardized hourly earnings	426.44	421.91
STANDARDIZED MONTHLY EARNINGS	59 540.54	57 598.37
Number of fulltime employees in the earnings statistics	33 542.00	51 541.00

The Statistics for 2013,2014 has been revised. The revision reiday allowance in DKK per hour worked is now included in thange is implemented for all sectors.

Monte Carlo Simulation: Example

Normalize to 0-1 year of experience, fresh grads:

- Coef = 354,473/542,130 = 0.654

DENMARK SALARY

SALARY SURVEYS
1,301
Search

UPDATED
August 2021
About salary data?

AVERAGE SALARY / YEAR
542,130 DKK
$88,517 USD

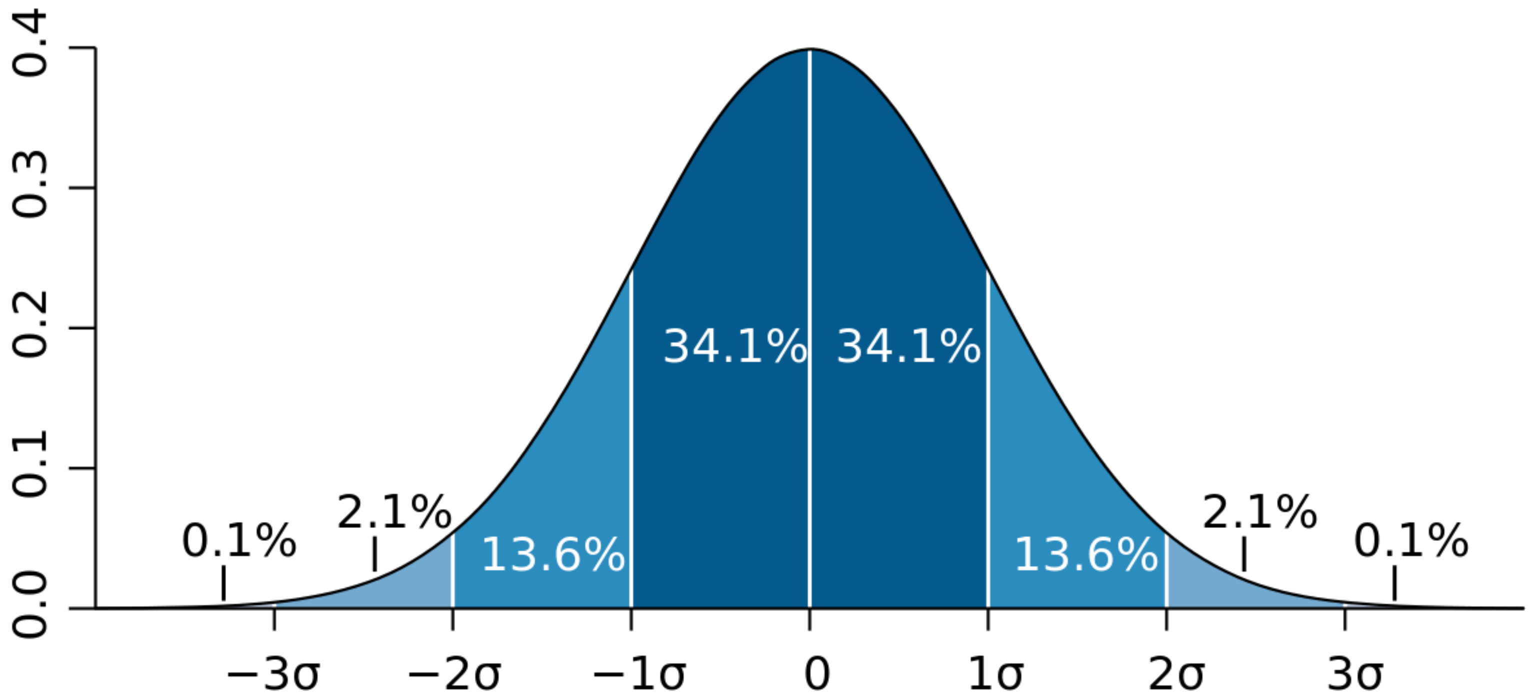
Average Salary / Denmark

Denmark (Gross DKK)

	Surveys	AVERAGE SALARY / YEAR
20+ Years	112	755,891 DKK
16-20 Years	90	728,220 DKK
12-16 Years	150	678,691 DKK
8-12 Years	280	586,105 DKK
4-8 Years	290	501,555 DKK
2-4 Years	171	416,925 DKK
0-1 Year	90	354,473 DKK
1-2 Years	118	343,657 DKK

Monte Carlo Simulation: Example

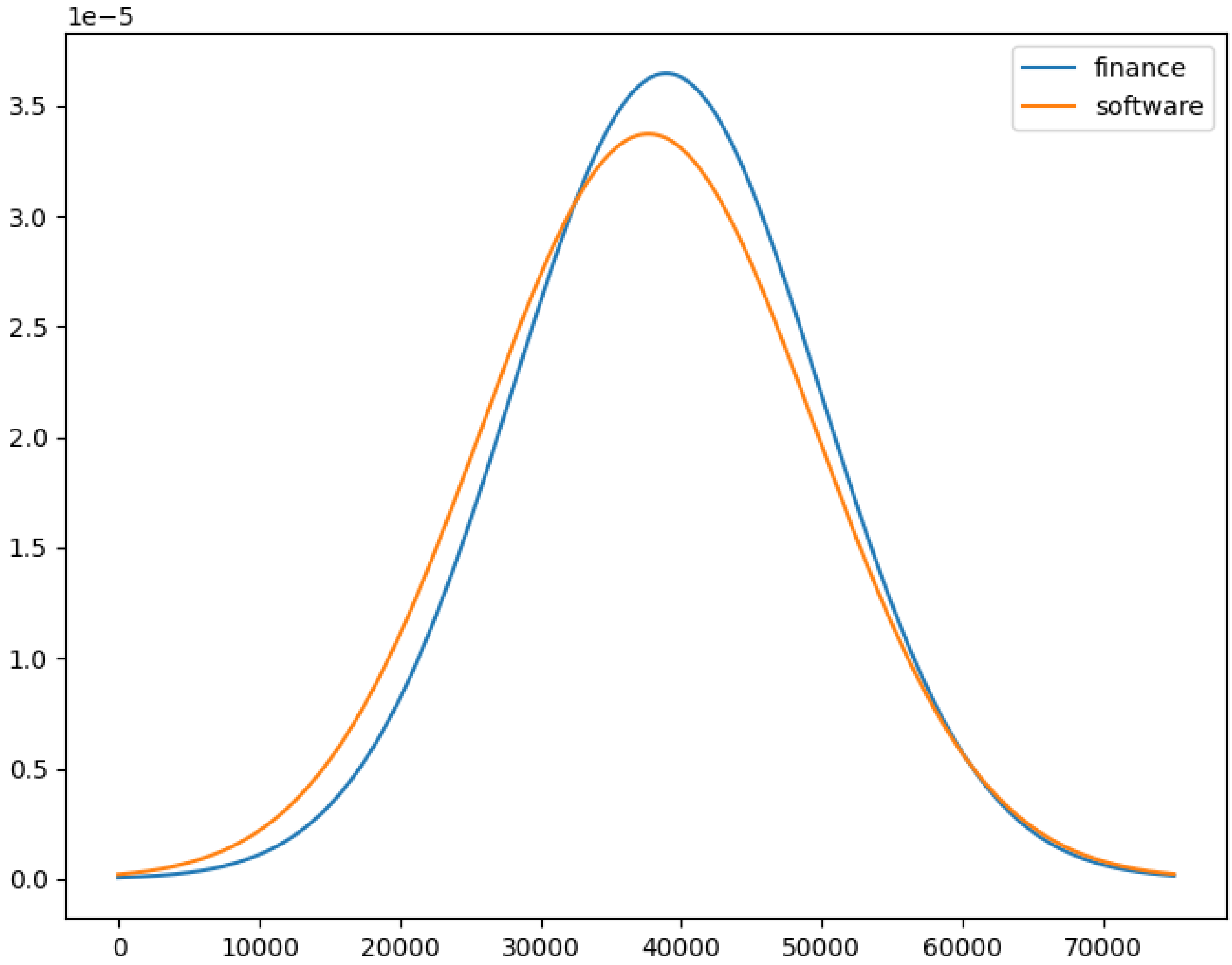
Standard deviation $\sim \frac{Q3-Q1}{1.35}$



Monte Carlo Simulation: Example

Finance:
Mean - 38927
STD ~ 10941
#jobs - 33500

Software:
Mean - 37659
STD ~ 11828
#jobs - 51500



For every 10 jobs in finance, there are 15 jobs in software development

Monte Carlo Simulation: Example

We are ready for a Monte Carlo simulation. Inputs:

- After getting a degree, student can apply for:
  - Up to 10 positions in finance, if finance education is selected
  - Up to 15 positions in software dev, if computer science education is selected
- For each application, there is a same probability of success (0,1]
- If successful, company offers a salary and student can:
  - Accept the job and stop searching for a job
  - Reject the job without potential reconsideration
- If n-1 jobs are rejected, the last application is guaranteed and cannot be rejected

What should be the student's strategy?

Monte Carlo Simulation: Example

Let's simulate thousands of random scenarios (python code):

```
min_salary = 17000

class Profession:

    def __init__(self, mean, std):
        self.mean = mean
        self.std = std

    def get_salary_sample(self):
        return max(min_salary, np.random.normal(self.mean, self.std))
```

Monte Carlo Simulation: Example

Let's simulate thousands of random scenarios (python code):

```
class JobSearchSimulator:

    def __init__(self, profession, n_positions, success_rate):
        self.profession = profession
        self.n_positions = n_positions
        self.success_rate = success_rate

    def run_constant_salary(self, requested_salary):
        for i in range(0, self.n_positions - 1):
            prob = np.random.random()
            if prob > self.success_rate:
                continue
            offer = self.profession.get_salary_sample()
            if offer >= requested_salary:
                return offer, True
        return self.profession.get_salary_sample(), False
```

Monte Carlo Simulation: Example

Let's simulate million random attempts (python code):

- A student is extremely competitive and has 100% of being offer a job.
- If the offered salary is lower than 40,000 DKK, he/she rejects it.

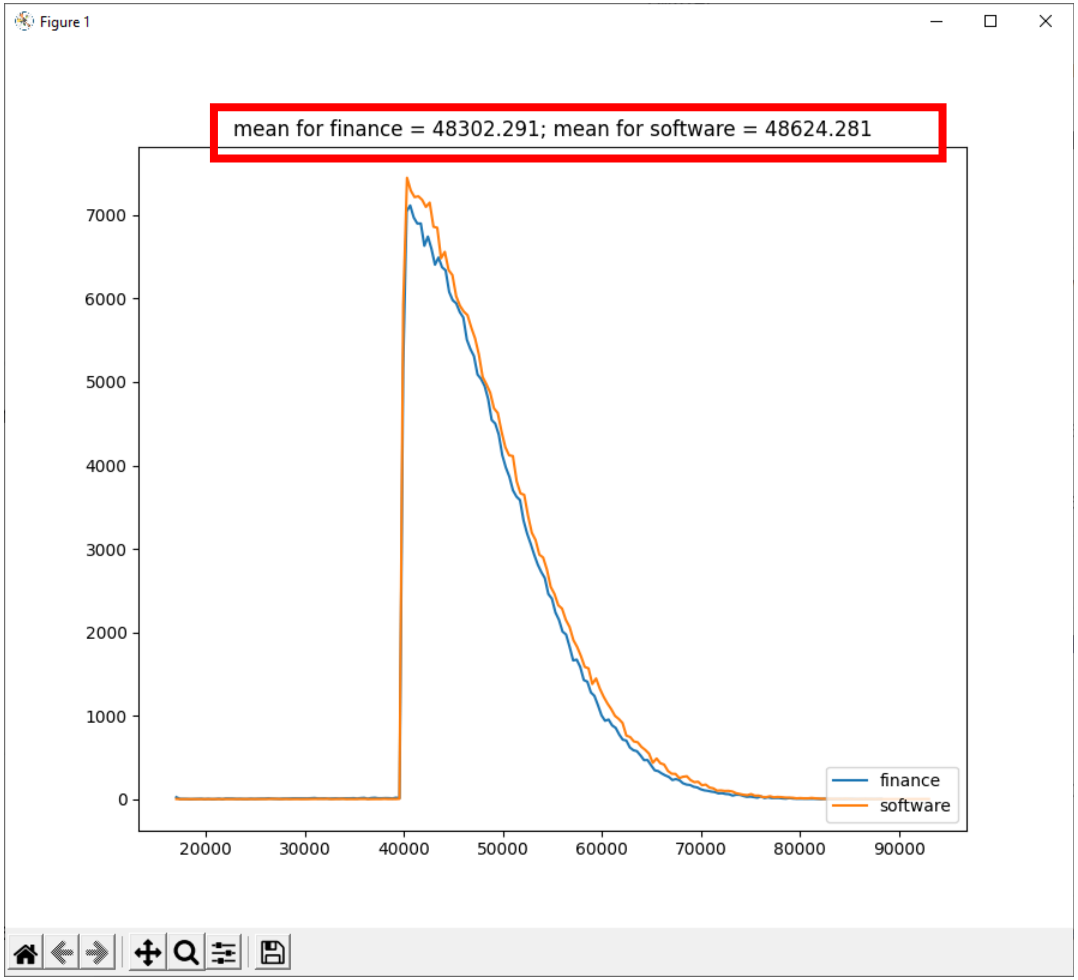
```
class MonteCarlo_Tests:

    def test_constant_salary_one():
        finance = JobSearchSimulator(Profession(38927, 10941), 10, 1)
        software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
        simulator = JobSearchStrategyTester(finance, software)
        simulator.compare_constant_salary(1000000, [40000, 40000], True)
```

Monte Carlo Simulation: Example

Let's simulate thousands of random scenarios (python code):

Finance:
- Pity offer = 0.38%
- Minimal wage risk = 0.047%

Software:
- Pity offer = 0.008%
- Minimal wage risk = 0.001%



Monte Carlo Simulation: Example

Let's simulate thousands of random scenarios (python code):

- A student got 48,302 DKK in finance and 48,624 DKK in software in previous simulation. Why not to explicitly request such salaries?

```
class MonteCarlo_Tests:

    def test_constant_salary_one():
        finance = JobSearchSimulator(Profession(38927, 10941), 10, 1)
        software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
        simulator = JobSearchStrategyTester(finance, software)
        simulator.compare_constant_salary(1000000, [48302, 48624], True)
```

Monte Carlo Simulation: Example

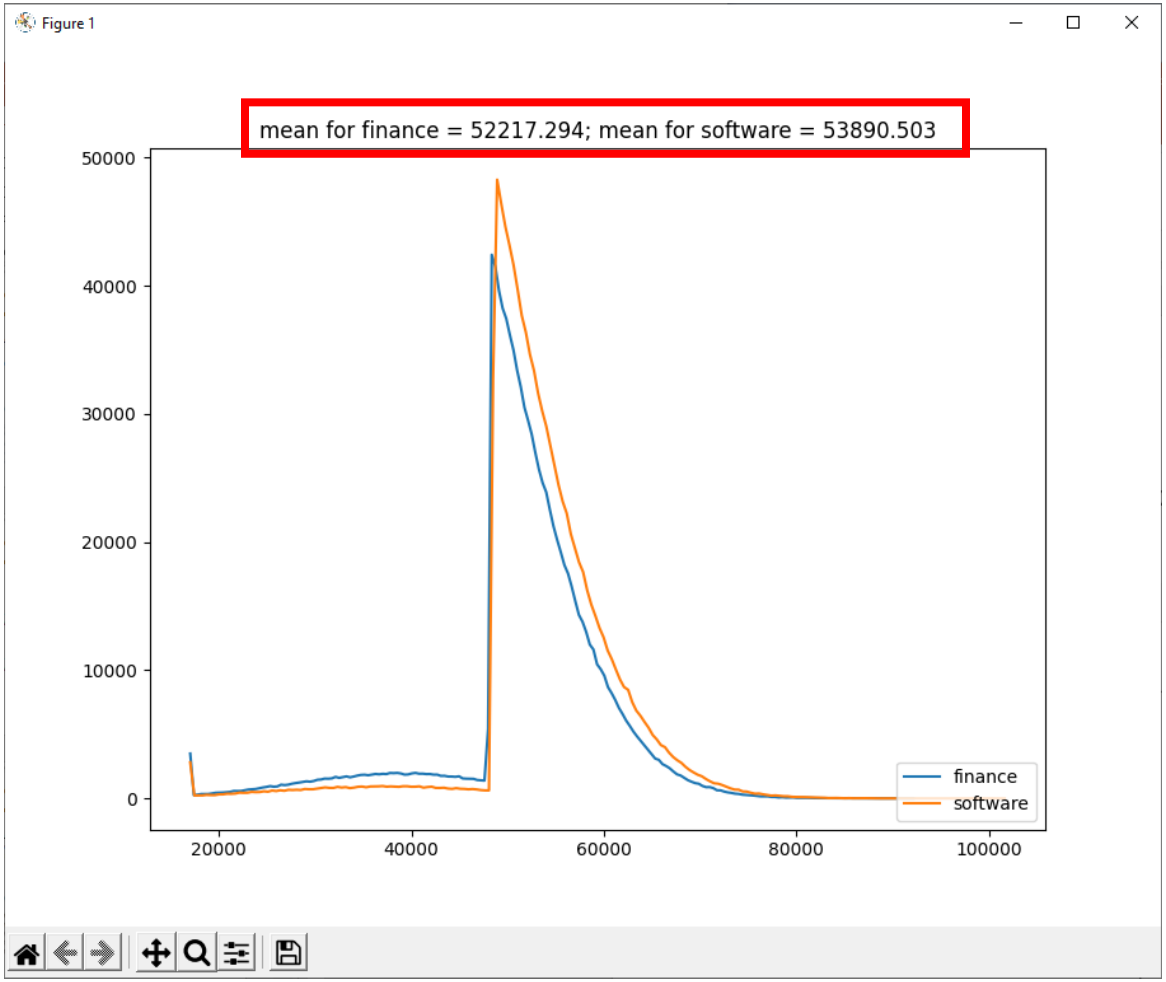
Pushing for a higher salary worked.

Finance:

- Pity offer = 14.05%
- Minimal wage risk = 0.32%

Software:

- Pity offer = 6.52%
- Minimal wage risk = 0.26%



Monte Carlo Simulation: Example

Next attempt

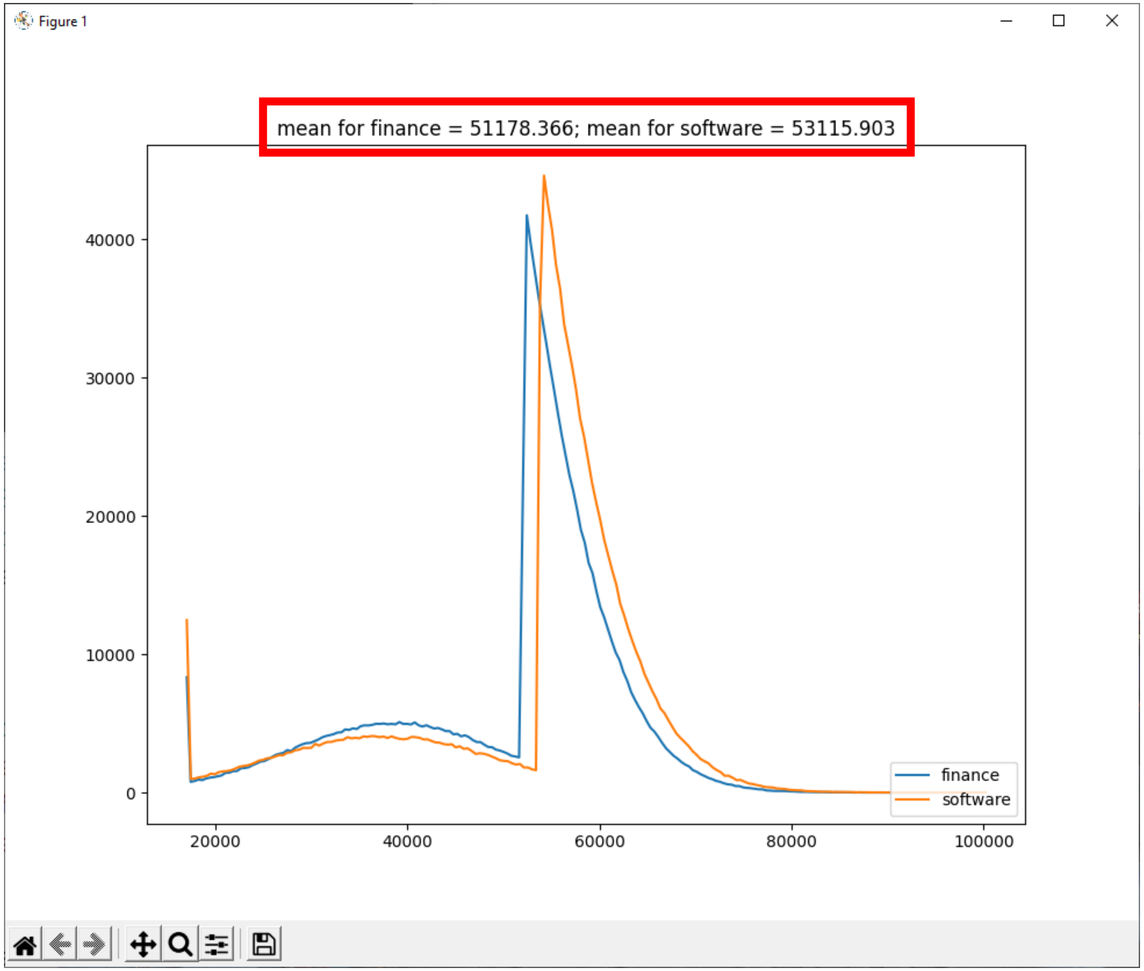
```
simulator.compare_constant_salary(1000000, [52217, 53890], True)
```

Finance:

- Pity offer = 34.25%
- Minimal wage risk = 0.76%

Software:

- Pity offer = 28.87%
- Minimal wage risk = 1.15%



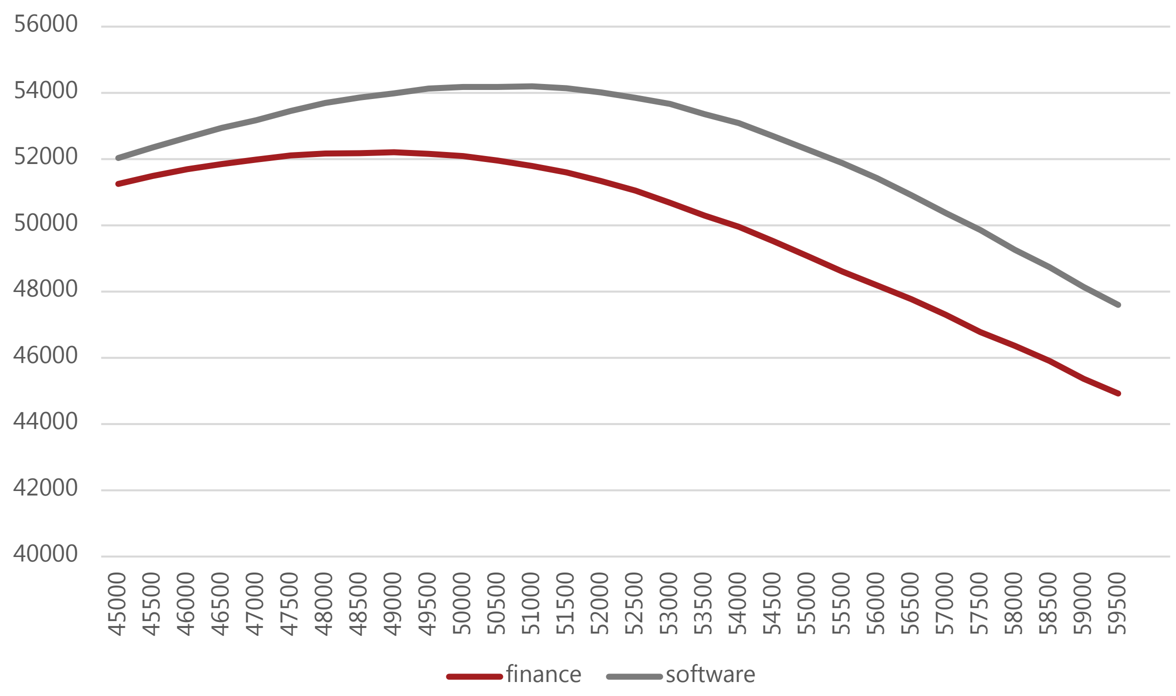
Monte Carlo Simulation: Example

Optimal strategy for finance is:

- Accept salary above 49,000 DKK. The average salary received 52,210 DKK

Optimal strategy for software is:

- Accept salary above 51,000 DKK. The average salary received 54,200 DKK



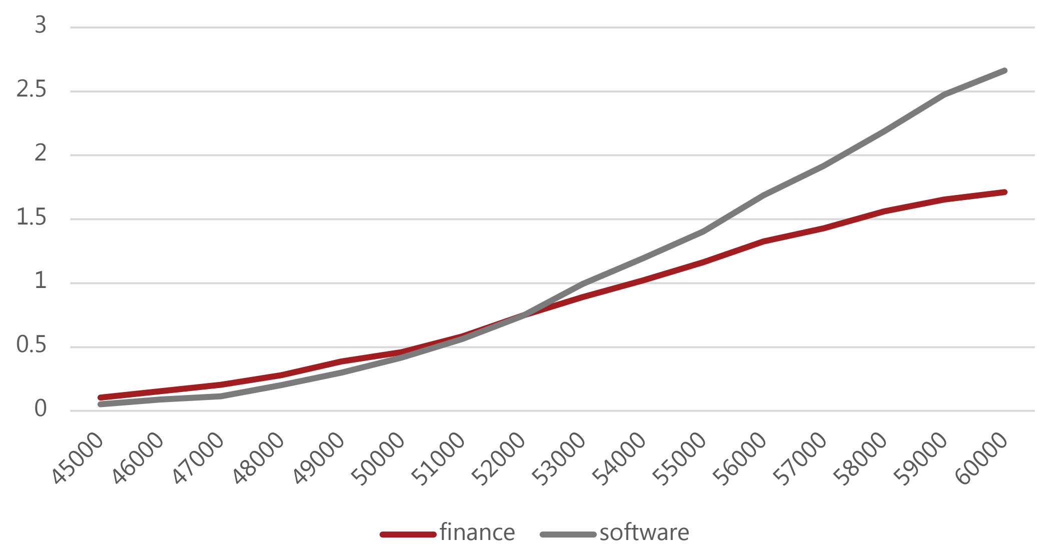
Monte Carlo Simulation: Example

Min. wage risk for finance is:

- Accept salary above 49,000 DKK. The risks of minimal wage is 0.39%.

Min. wage risk for software development is:

- Accept salary above 51,000 DKK. The risks of minimal wage is 0.56%.



Monte Carlo Simulation: Example

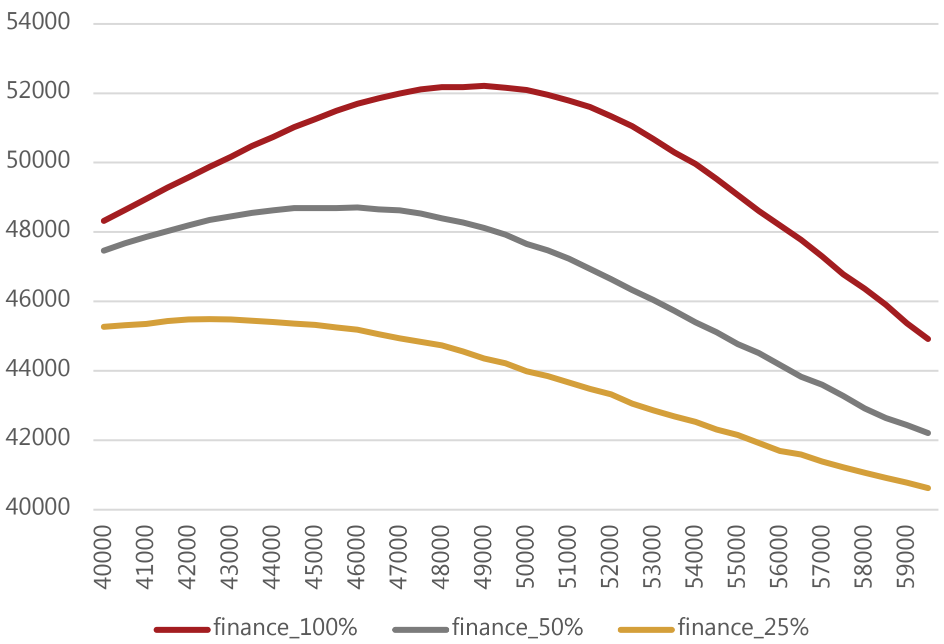
If the application success rate is only 50% or 25%:

```
finance_1   = JobSearchSimulator(Profession(38927, 10941), 10, 1)
finance_05  = JobSearchSimulator(Profession(38927, 10941), 10, 0.5)
finance_025 = JobSearchSimulator(Profession(38927, 10941), 10, 0.25)
```

Finance 100% success:
- Accept 49,000 DKK, receive 52,210 DKK

Finance 50% success:
- Accept 46,000 DKK, receive 48,712 DKK

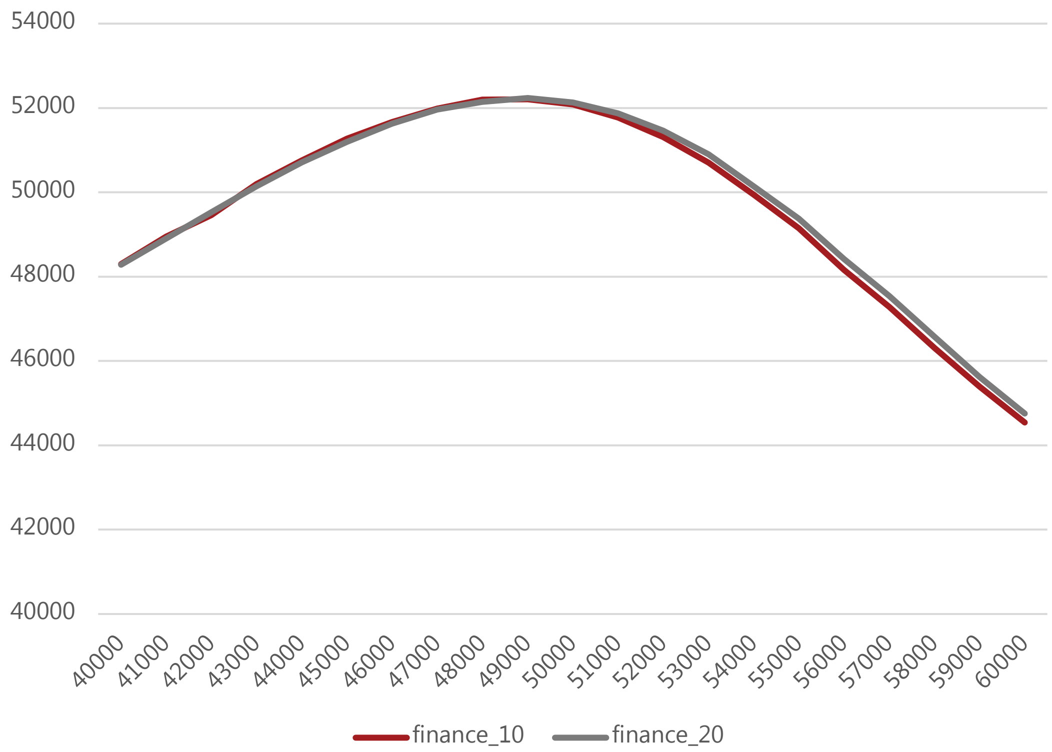
Finance 25% success:
- Accept 42,500 DKK, receive 45,491DKK



Monte Carlo Simulation: Example

Is it more beneficial to apply to more positions or become more competitive?

```
finance_1 = JobSearchSimulator(Profession(38927, 10941), 10, 1)
finance_2 = JobSearchSimulator(Profession(38927, 10941), 20, 0.5)
```



Monte Carlo Simulation: Example

The student predefines an optimal salary value and is not willing to accept any offer with a lower salary. Can we improve this strategy?

```
software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
```

- Let's say he was rejected n-2 times and only two jobs left. Shall he insist on the originally optimal 51,000 DKK salary on the n-1th attempt?

- He may consider gradually reducing the minimal acceptable salary with every reject

Monte Carlo Simulation: Example

We can find the salary reduction schedule with the Monte Carlo simulation by trying different combinations of salaries.

Very slow

```
software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
```

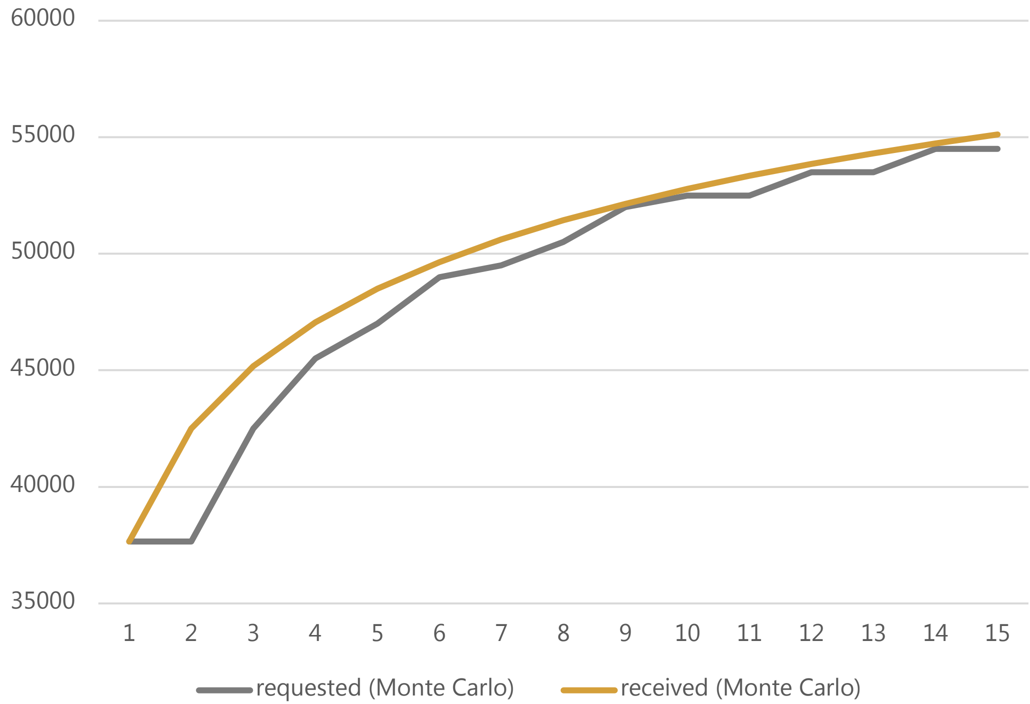
We can use Monte Carlo to find the optimal salary reduction schedule for 2 jobs, then gradually add jobs

Monte Carlo Simulation: Example

We can use Monte Carlo to find the optimal salary reduction schedule for 2 jobs, then gradually add jobs

Software:

- Start from 54,500 DKK, and gradually reduce the requested salary
- Receive 55,122 DKK



Monte Carlo Simulation: Example

Let's find the optimal schedule:

- Suppose only two jobs are available.
- We know if a first offer is rejected, we will get in average 37,659 DKK offer in the next attempt and will have to accept it.
- We should therefore accept any salary > 37,659 DKK in the first attempt.
- The average received salary will be:

$$\mathbf{S_2} = p_1 S_1 + p_2 S_2$$

p_1 - probability of success of the first attempt
S_1 - average salary expected for a successful first attempt
$p_2 = 1 - p_1$ - probability of having the second attempt
S_2 - average salary expected in the second attempt

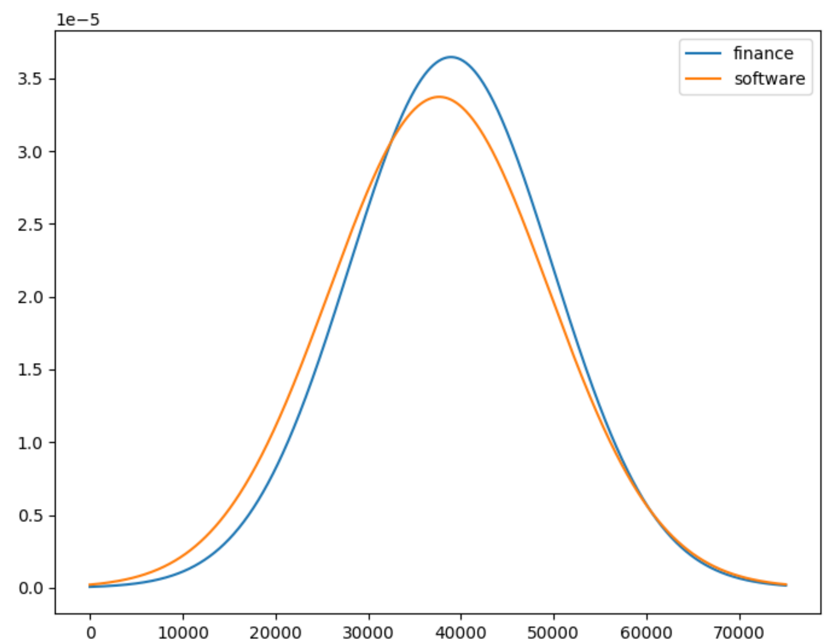
Monte Carlo Simulation: Example

```
software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
```

If we are lucky at the first attempt, the expected salary will be:

$$\int_{37659}^{\infty} 2 \cdot \frac{x}{11828\sqrt{2\pi}} e^{-0.5\left(\frac{x-37659}{11828}\right)^2} dx = 47096$$

$$\mathbf{S_2} = p_1 S_1 + p_2 S_2$$
$$= 0.5 \cdot 47096 + 0.5 \cdot 37659 = 42337$$



Monte Carlo Simulation: Example

```
software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
```

For three positions:

- We know if a first offer is rejected, we will get in average 42,337 DKK in the next two attempts.
- We should therefore accept any salary > 42,337 DKK in the first attempt.
- The probability of such an offer:

$$\int_{42337}^{\infty} \frac{1}{11828\sqrt{2\pi}} \mathrm{e}^{-0.5\left(\frac{x-37659}{11828}\right)^2} \mathrm{d}x = 0.346$$

Monte Carlo Simulation: Example

```
software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
```

For three positions:

- The probability of such an offer:

$$\int_{42337}^{\infty} \frac{1}{11828\sqrt{2\pi}} e^{-0.5\left(\frac{x-37659}{11828}\right)^2} \mathrm{d}x = 0.346$$

- The salary, if we are lucky at the first attempt:

$$\int_{42337}^{\infty} \frac{1}{0.346} \cdot \frac{x}{11828\sqrt{2\pi}} e^{-0.5\left(\frac{x-37659}{11828}\right)^2} \mathrm{d}x = 50296$$

- The total expected salary:

$$\begin{aligned} \mathbf{S_3} &= p_3 S_3 + (1 - p_3)\mathbf{S_2} \\ &= 0.346 \cdot 50296 + (1 - 0.346) \cdot 42337 = 45091 \end{aligned}$$

Monte Carlo Simulation: Example

```
software = JobSearchSimulator(Profession(37659, 11828), 15, 1)
```

For four positions:

- The probability of such an offer:

$$\int_{45091}^{\infty} \frac{1}{11828\sqrt{2\pi}} e^{-0.5\left(\frac{x-37659}{11828}\right)^2} \mathrm{d}x = 0.265$$

- The salary, if we are lucky at the first attempt:

$$\int_{45091}^{\infty} \frac{1}{0.265} \cdot \frac{x}{11828\sqrt{2\pi}} e^{-0.5\left(\frac{x-37659}{11828}\right)^2} \mathrm{d}x = 52260$$

- The total expected salary:

$$\mathbf{S_4} = p_4 S_4 + (1 - p_4)\mathbf{S_3}$$
$$= 0.265 \cdot 52260 + (1 - 0.265) \cdot 45091 = 46991$$

Monte Carlo Simulation: Example

```
software = JobSearchSimulator(Profession(57600, 18091), 15, 1)
```

Comparison between Monte Carlo and explicitly computed value:

Monte Carlo:

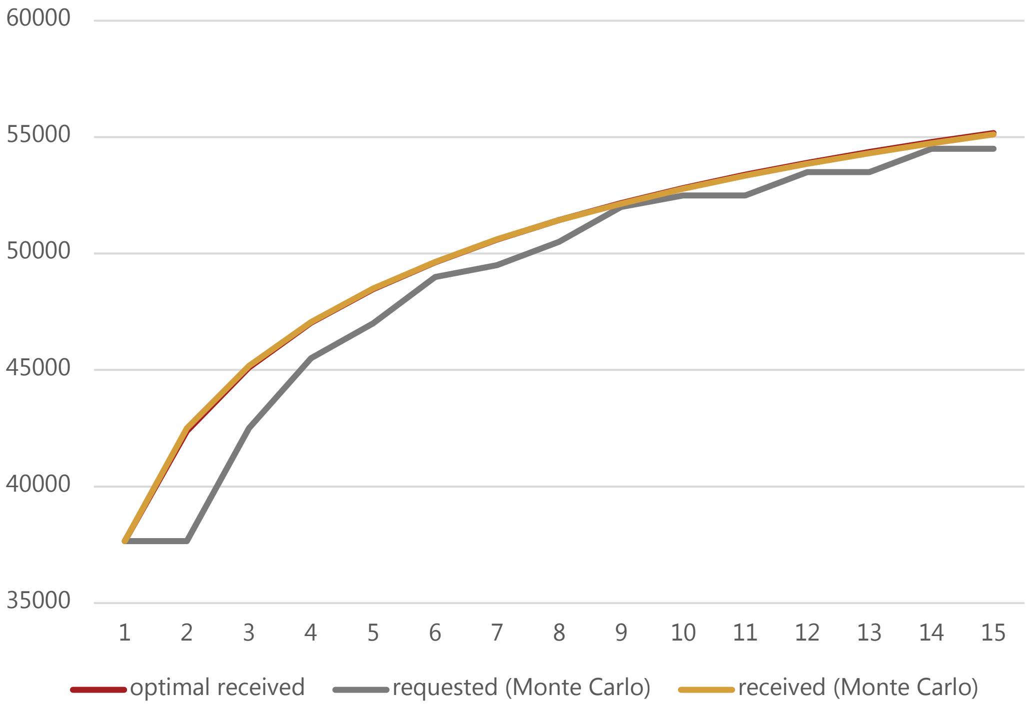
- Start from 54,500 DKK, and gradually reduce the requested salary

Receive 55,122 DKK

Optimal:

- Start from 54,787 DKK, and gradually reduce the requested salary

Receive 55,172 DKK



Monte Carlo Simulation: Example

Limitations:

- The salaries we have are for all types of jobs in the field, including managerial positions. The salary mean and standard deviation for fresh graduates will be different.
- It is very difficult to correctly estimate how many openings will be available for you.

Conclusions:

- Apply as much as you can. The probability of success is probably way lower than 50%.
- High salary variability is a positive thing.