PROBABILITY

Goal: estimate/model uncertainty

Random experiment/phenomena

1) flip a coin: outcome H(eads) or T(ails)

2) flip a coin 3 times: outcomes
$HHH, HHT, HTH, THH, HTT, THT, TTH, TTT$

3) throw a dice: outcomes 1,2,3,4,5,6

4) throw 2 dice: outcomes (36)
(1,1),(1,2),. . . ,(6,6)

5) 100 items, 15 defective, 85 non-defective. Check 10. Outcomes are all 10 item collections from 100.

If all N outcomes are equally likely , then the probability of each is $1/N$.

An event A is a collection of outcomes and it's probability

$$P(A) = \frac{\text{number of outcomes in } A}{N}$$

1) $P(H) = P(T) = 1/2$

2) $P(HHH) = \cdots = P(TTT) = 1/8$,
$P(2\ H)$=$P(2\ T)$=$3/8$

3) $P(1) = \cdots = P(6) = 1/6$,
$P(5 \text{ or } 6)$= $2/6$

4) $P(1,1) = \cdots = P(6,6) = 1/36$,
$P(\text{sum=7})$= $6/36$

5) $P(\text{find at least one defective})$=$\frac{14181146783804}{17310309456440} \approx 0.82$

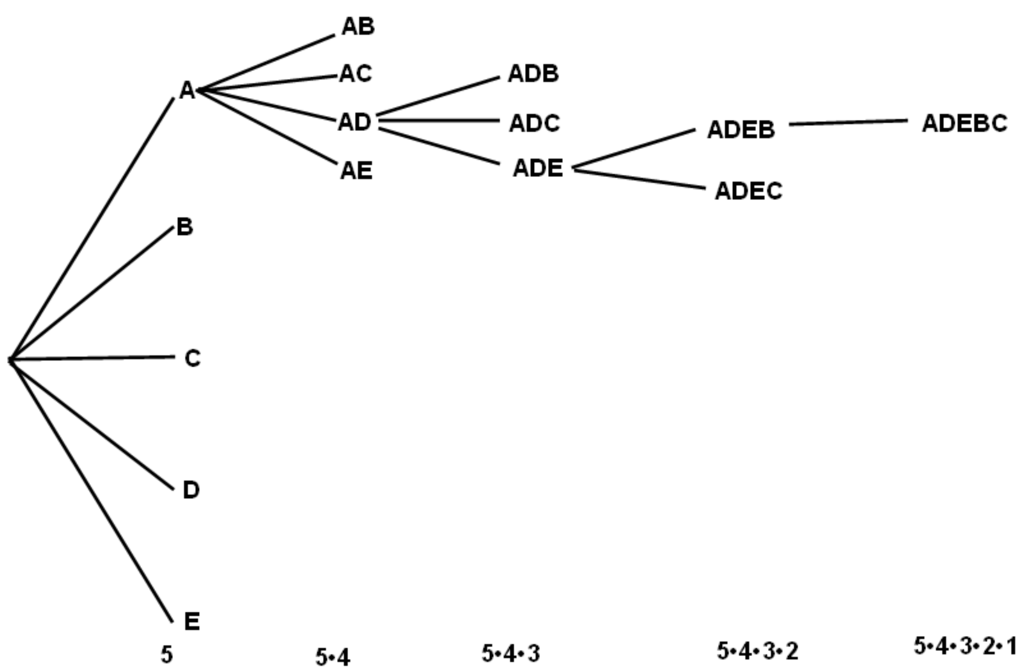
Combinatorics

= art of counting

Ex. How many different arrangements can be formed from five things, for example A,B,C,D,E, i.e ABCDE, CAEBD, ECBDA,...

Basic principle:

First can be any of the 5, second any of the 4 remaining etc



i.e the number of different arrangements (**permutations**) is

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

Similarly, there are

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 1$$

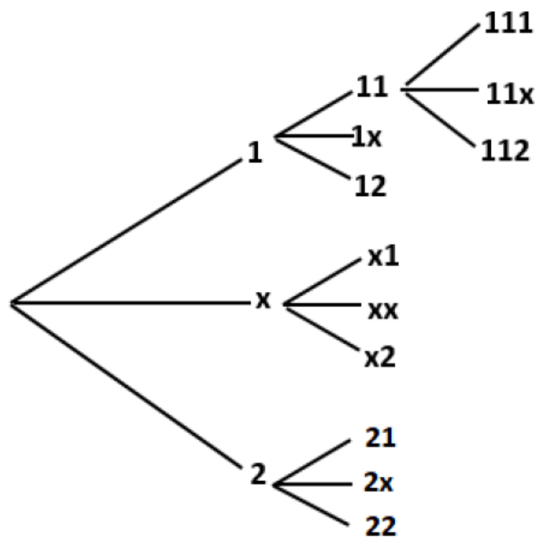
different permutations of n things

For example

$$10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 3628800$$

Ex: In vakioveikkaus (Finnish football pool) there are 13 games, each with 3 different choices (1 = home win, x = draw, 2 = visitor win). Number of possibilities to fill the coupon is

$$\underbrace{3 \cdot 3 \cdots 3}_{13 \text{ kpl}} = 3^{13} = 1594323$$



Ex: Passwords of length 8 consists of small or capital letters ($2 \cdot 29$), number (10) or special character (15). Number of different passwords is
$83^8 = 2\,252\,292\,232\,139\,041$

Ex: Throw dice 6 times. Calculate probabilities
P(6 different numbers)
P(no result 6)

Outcomes are sequences $abcdef$, and there are 6^6 of them

There are $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ outcomes with 6 different numbers i.e

$$P(6 \text{ different numbers}) = \frac{6!}{6^6} \approx 0.015$$

There are 5^6 outcomes without result 6 i.e

$$P(\text{no } 6) = \frac{5^6}{6^6} \approx 0.33$$

Simulation: generate outcomes of 6 throws from 1,2,...,6, repeat for example 100000 rounds, and calculate the fraction of rounds that produce the outcome in question

Ex. (Birthday problem) Calculate probability that n random persons have all different birthday

Birthdays form a vector of length n from numbers 1,2,...,365, and there are 365^n outcomes

There are $365 \cdot 364 \cdots (365-(n-1))$ outcomes with different days, i,e

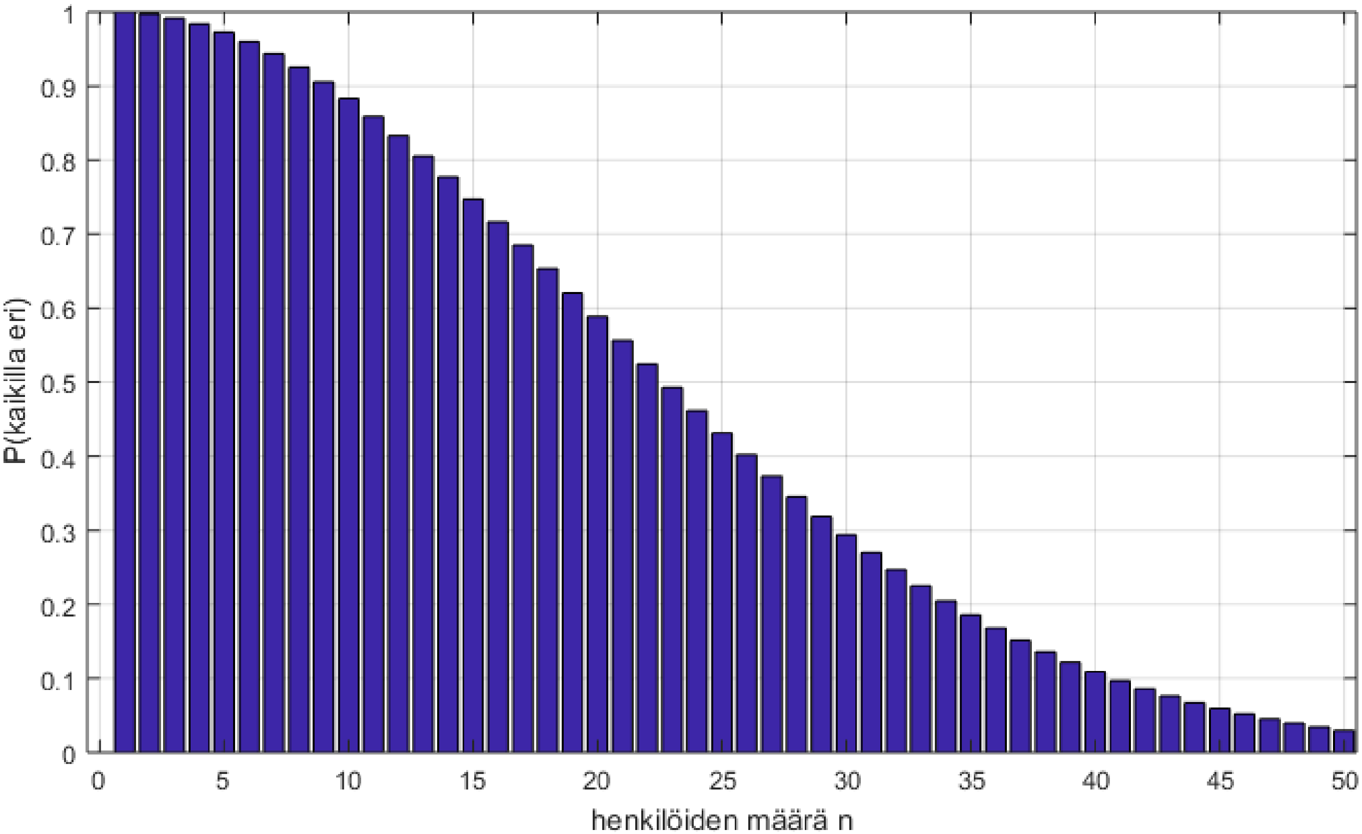
P(all birthdays different) =

$$= \frac{365 \cdot 364 \cdots (365-(n-1))}{365^n}$$

In particular, if $n \geq 23$, then

P(all different) < 0.5 i.e

P(at least 2 have same)> 0.5



P(kaikilla eri)
henkilöiden määrä n
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15
20
25
30
35
40
45
50

Esim: Derangement is a permutation that has no fixed points

Counting derangements of a set amounts to the hat-check problem, in which one considers the number of ways in which n hats (call them h_1 through h_n) can be returned to n people (P_1 through P_n) such that no hat makes it back to its owner.
Person P_1 may receive any of the $n-1$ hats that is not his own. Call the hat which the person P_1 receives h_i and P_i receives either P_1's hat, h_1, or some other. Accordingly, the problem splits into two possible cases:

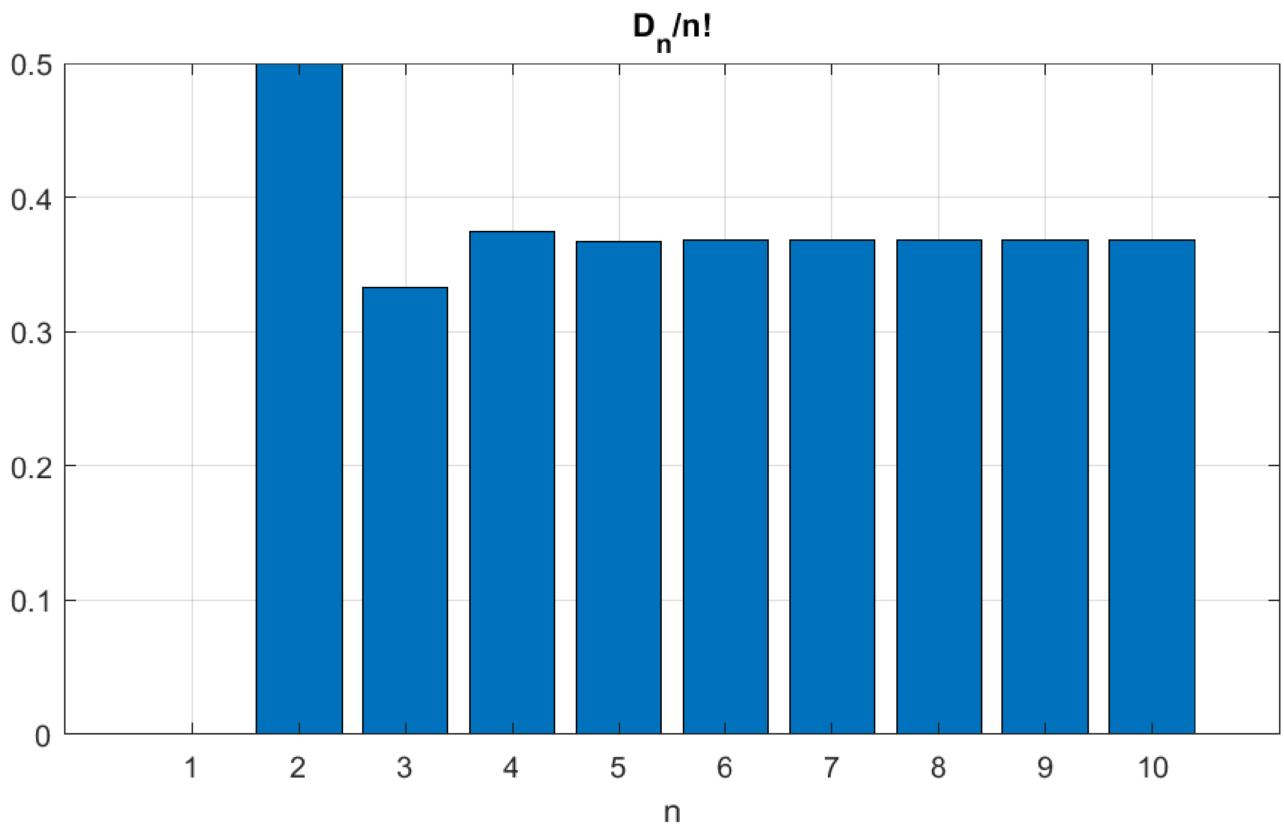
1) P_i receives a hat other than h_1. This case is equivalent to solving the problem with $n-1$ people and $n-1$ hats because for each of the $n-1$ people besides P_1 there is exactly one hat from among the remaining $n-1$ hats that they may not receive (for any P_j besides P_i, the unreceivable hat is h_j, while for P_i it is h_1).

2) P_i receives h_1. In this case the problem reduces to $n-2$ people and $n-2$ hats, because P_1 received P_i's hat and P_i received P_1's hat, effectively putting both out of further consideration.

For each of the $n-1$ hats that P_1 may receive, the number of ways that $P_2, \ldots, P_n$ may all receive hats is the sum of the counts for the two cases, i.e if D_n is the number of derangements, then

$$D_1 = 0,\ D_2 = 1$$

$$D_n = (n-1)(D_{n-1} + D_{n-2}),\ n = 3, 4, \ldots$$



$$P(\text{all wrong}) = \frac{D_n}{n!} \approx \frac{1}{e} \approx 0.37,\ \text{ if } n \geq 5$$

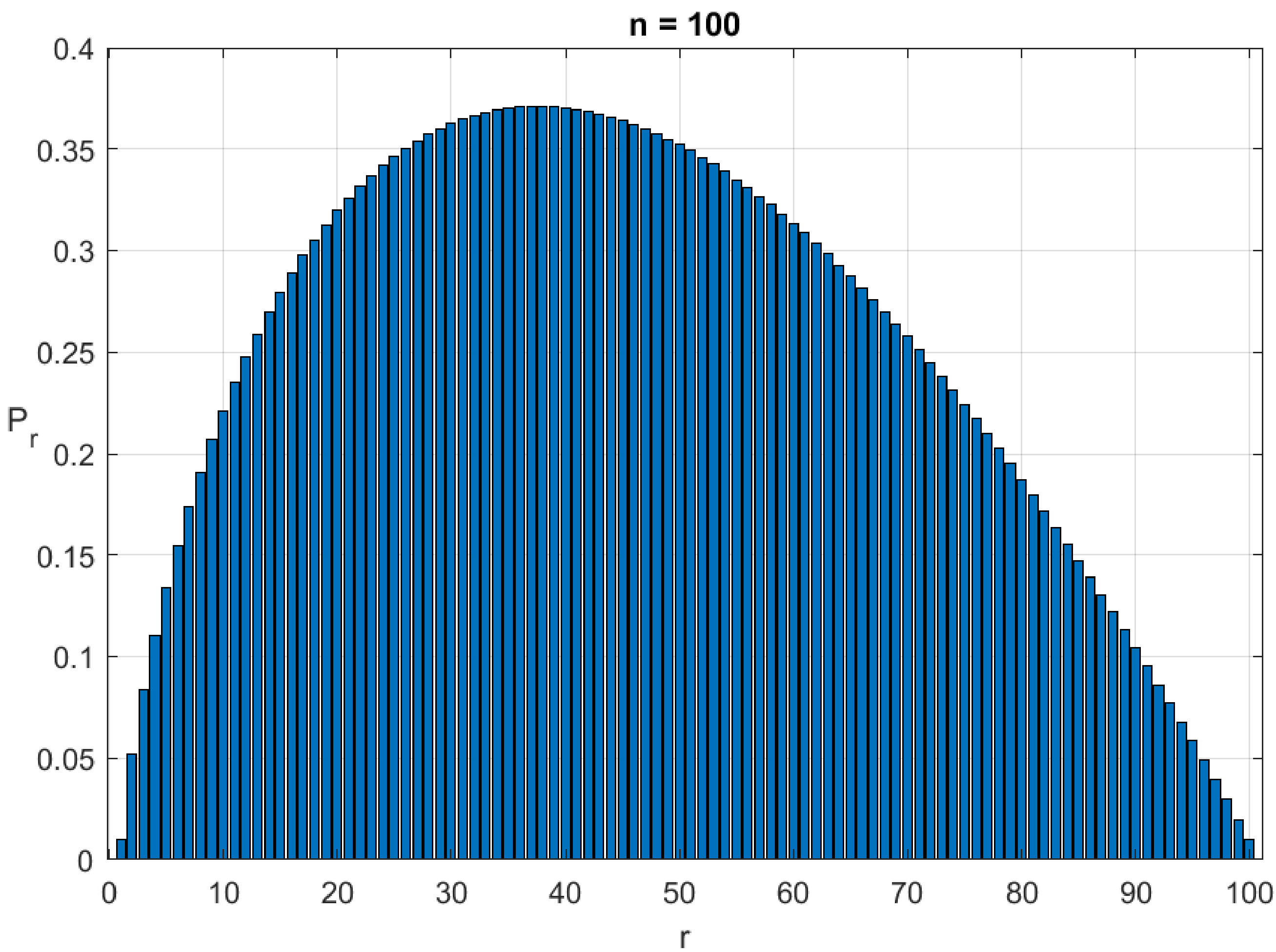
Ex: Secretary problem.

The basic form of the problem is the following: imagine an administrator who wants to hire the best secretary out of n rankable applicants for a position. The applicants are interviewed one by one in random order. A decision about each particular applicant is to be made immediately after the interview. Once rejected, an applicant cannot be recalled. During the interview, the administrator gains information sufficient to rank the applicant among all applicants interviewed so far, but is unaware of the quality of yet unseen applicants. The question is about the optimal strategy to maximize the probability of selecting the best applicant.

The optimal policy for the problem is a **stopping rule**: the interviewer rejects the first $r-1$ applicants, $r = 2, 3, \ldots, n$, and then selects the first subsequent applicant that is better than the best applicant among first $r-1$ applicants.
If there is no such applicant, the last applicant is selected.

$$P_r = P(\text{the best applicant is selected})$$

$$= \sum_{i=r}^{n} \frac{1}{n} \cdot \frac{r-1}{i-1},\ r = 2, 3, ..., n$$



P_r is largest, when $r \approx n/e \approx 0.37n$, and largest value is $\approx 1/e \approx 0.37$

Ex. How many ways there is to choose 3 from 5 items, for example A,B,C,D,E ?

i.e the order of items doesn't matter (**combinations**)

$$\text{A,B,C,} \quad \text{A,C,E,} \quad \text{B,C,D,} \quad \ldots$$

Number of ordered triples (ABC, CAB, BDC, ...) is $5 \cdot 4 \cdot 3 = 60$

Each triple appears in $3! = 3 \cdot 2 \cdot 1 = 6$ different order i.e

$$\text{number of triples} \cdot 3! = 5 \cdot 4 \cdot 3$$

i.e

$$\text{number of triples} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

This is called 5 over 3 and denoted by

$$\binom{5}{3}$$

In the same way, there are

$$\binom{10}{4} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210$$

ways to choose 4 from 10 items

Binomial coefficient (n over k)

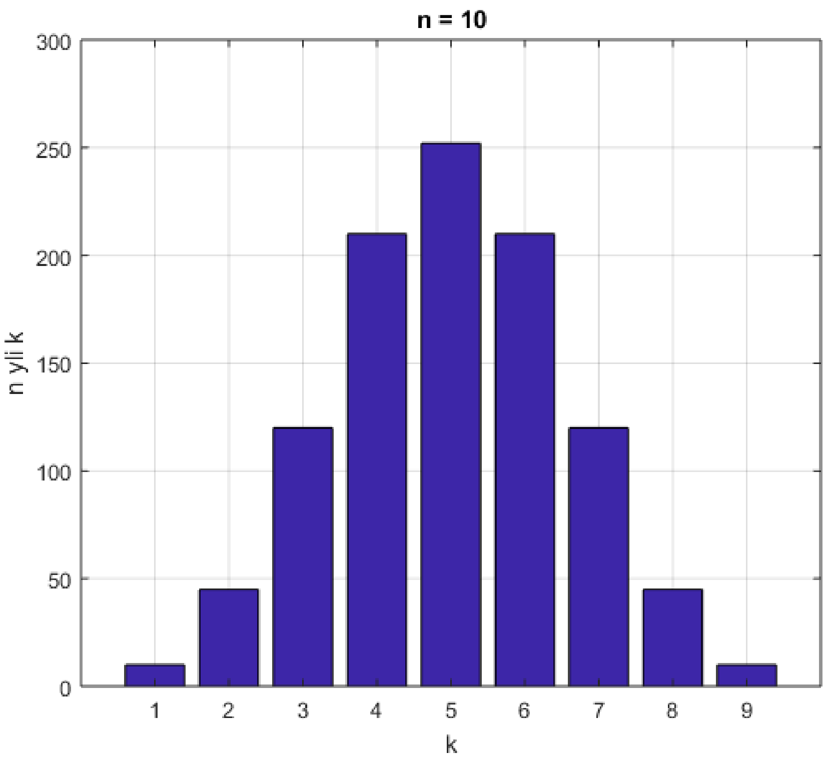
$$\binom{n}{k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$$

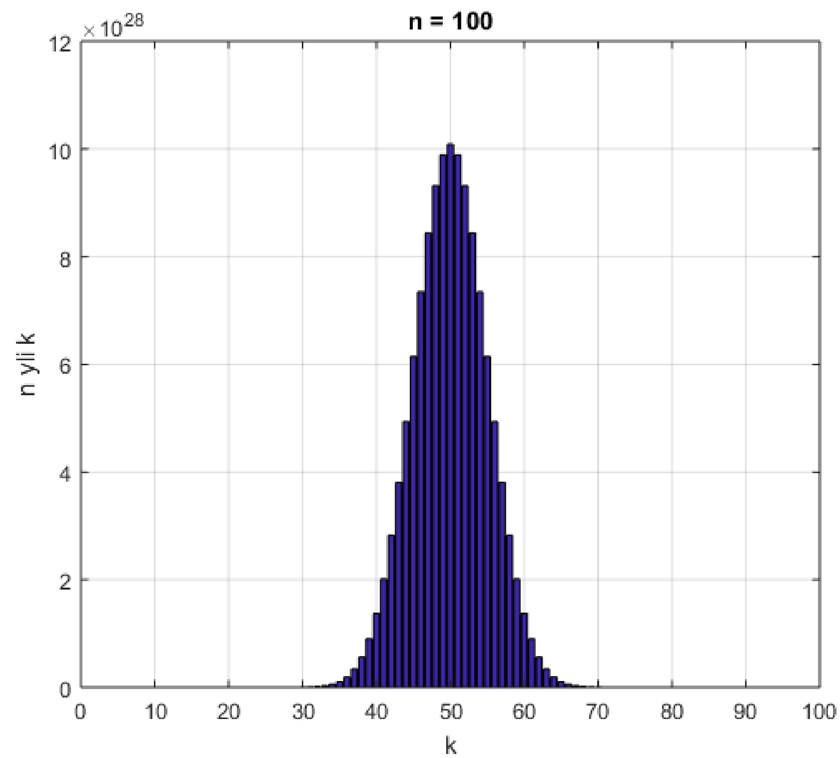
$$= \frac{n!}{(n-k)! \cdot k!}, \quad k = 0, 1, 2, \ldots, n$$

is the number of ways to choose k from n items

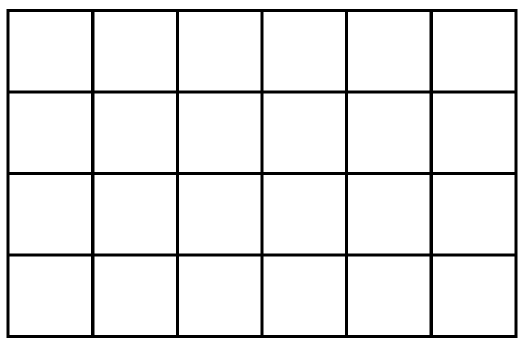
Note: 0!=1 and

$$\binom{n}{0} = \binom{n}{n} = 1$$


n = 10
300
250
200
150
100
50
0
n yli k
1
2
3
4
5
6
7
8
9
k


n = 100
×10^28
12
10
8
6
4
2
0
n yli k
0
10
20
30
40
50
60
70
80
90
100
k

Ex. How many routes from lower left corner to upper right corner, when one can move either one step to the right or up ?
How many routes go through the center point of the grid ?



Two routes (R=right, U=up)

URRUURURRR

RURRURRURU

Common to all routes: total 10 steps, 6 right and 4 up
Number of different routes is

$$\binom{10}{4} = \binom{10}{6} = 210$$

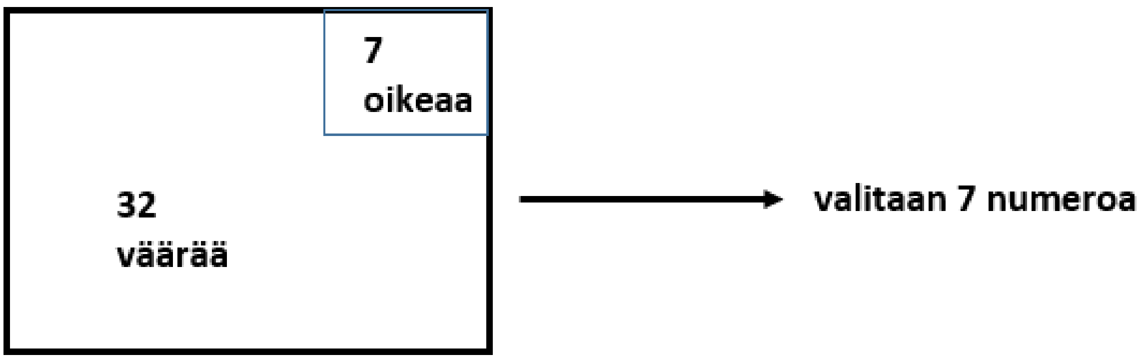
(choose 4 U:s or 6 R:s from 10).

Route URRUU|RURRR doesn't go therough the center point, route RURRU | RRURU does

In order to go through the center point, first 5 and last 5 steps steps should contain 2 U:s and 3 R:s
Number of different routes through the center point is

$$\binom{5}{2} \cdot \binom{5}{2} = 100$$

Ex. How many ways to fill a lotto coupon i.e choose 7 numbers from 39 ? Probability to have 0,1,2,3,4,5,6,7 winning numbers ?



Number of different coupons

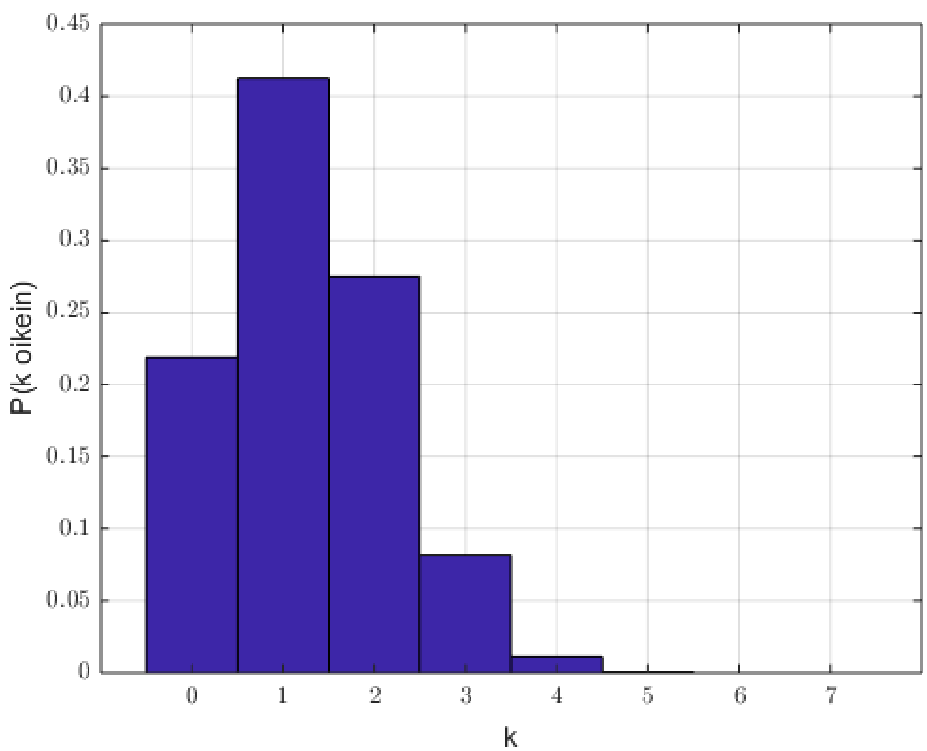
$$\binom{39}{7} = \frac{39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 15380937$$

Coupons with k winning and $7 - k$ non-winning numbers

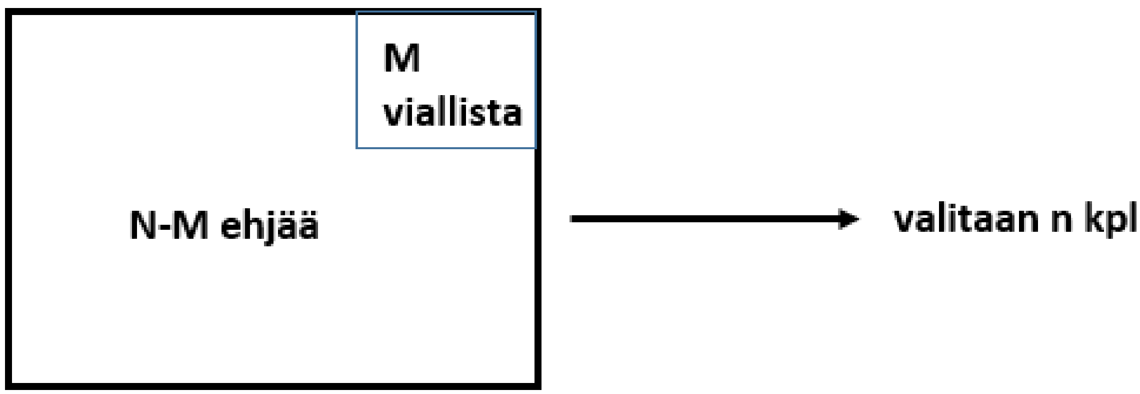
$$\underbrace{\binom{7}{k}}_{\text{winning}} \cdot \underbrace{\binom{32}{7-k}}_{\text{non-winning}}, \quad k = 0, 1, 2, \ldots, 7$$

$$P(k \text{ winning}) = \frac{\binom{7}{k} \cdot \binom{32}{7-k}}{\binom{39}{7}}, \quad k = 0, 1, 2, \ldots, 7$$

```
0.2188    0.4124    0.2749    0.0818    0.0113    0.0007    0.0000    0.0000
```



Ex. In quality control n items from total of N items are checked.

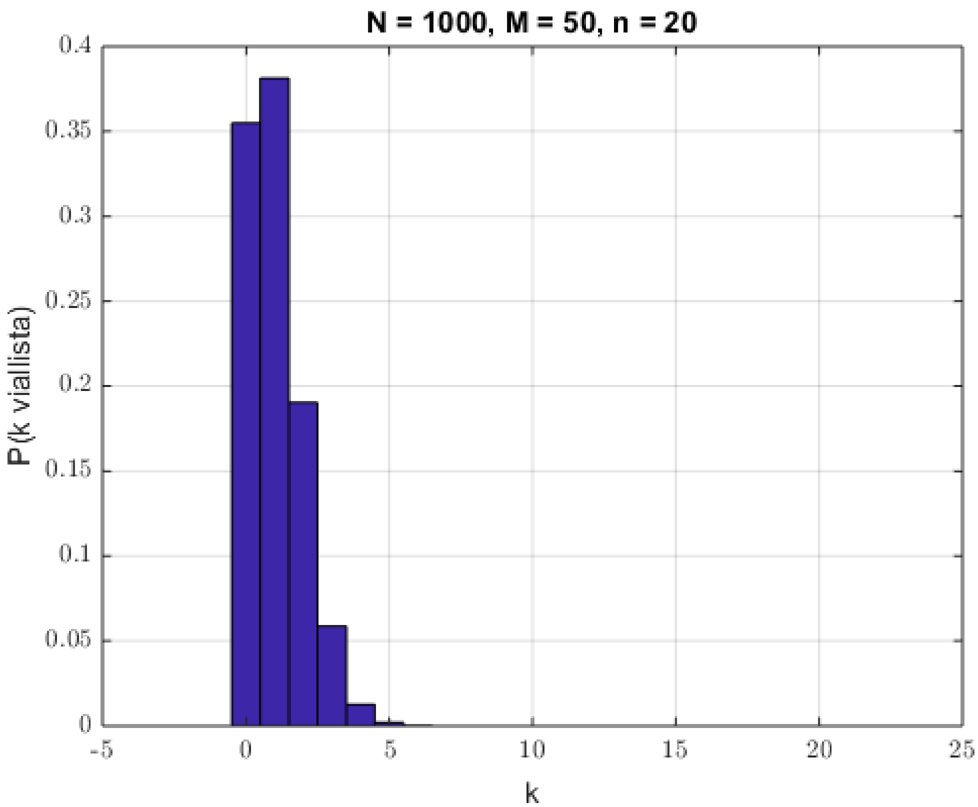


If there are M defective items, then

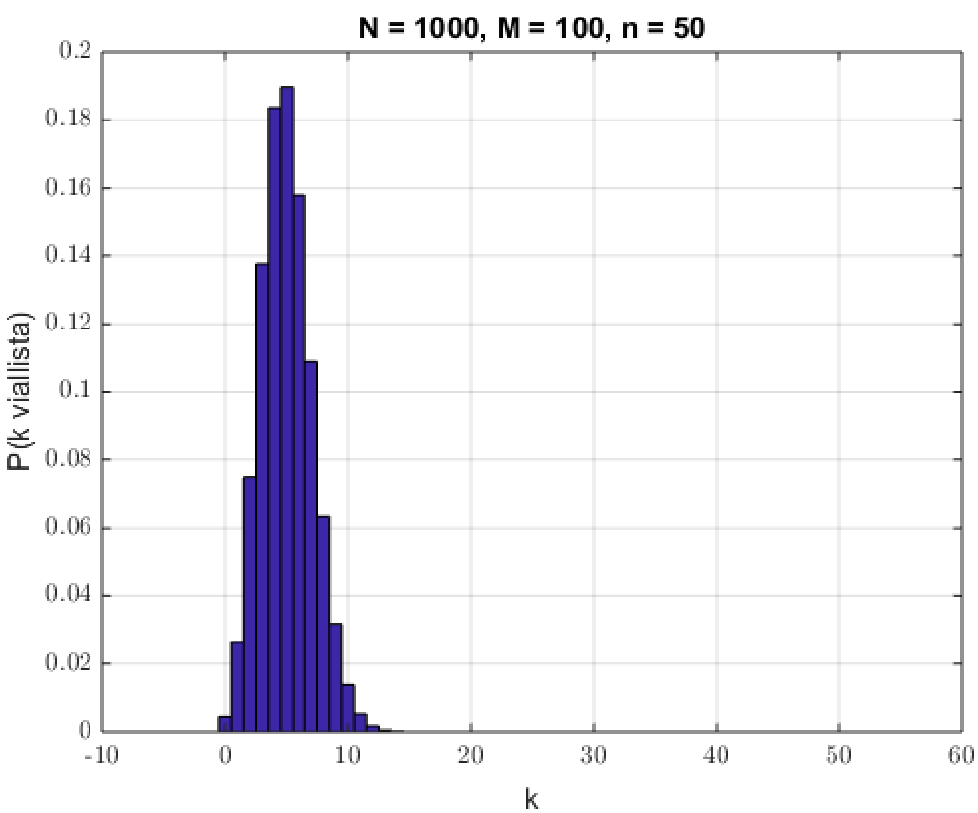
$$P(k \text{ defectives are found})$$

$$= \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}, \ k = 0, 1, \ldots, n$$

(Hypergeometric distribution)



N = 1000, M = 50, n = 20
P(k viallista)
k
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-5
0
5
10
15
20
25



N = 1000, M = 100, n = 50
P(k viallista)
k
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-10
0
10
20
30
40
50
60

Ex. Deck of 52 cards, 4 suits, each with numbers $1-13$. Number of different 5 card poker hands

$$\binom{52}{5} = 2598960$$

flush (5 cards of same suit):

$$4 \cdot \binom{13}{5} = 5148$$

full house (3 cards with one number and 2 cards with another):

$$13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} = 3744$$

straight (5 consecutive numbers, 1-5,2-6,...,9-13):

$$9 \cdot 4^5 = 9216$$

straight flush:

$$4 \cdot 9 = 36$$

4 same numbers :

$$13 \cdot 48 = 624$$

3 same numbers:

$$13 \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot 4^2 = 54912$$

2 pairs (2 cards with one numbers, 2 with another):

$$\binom{13}{2} \cdot \binom{4}{2}^2 \cdot 44 = 123552$$

pair (2 cards with one number):

$$13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3 = 1098240$$

rest:

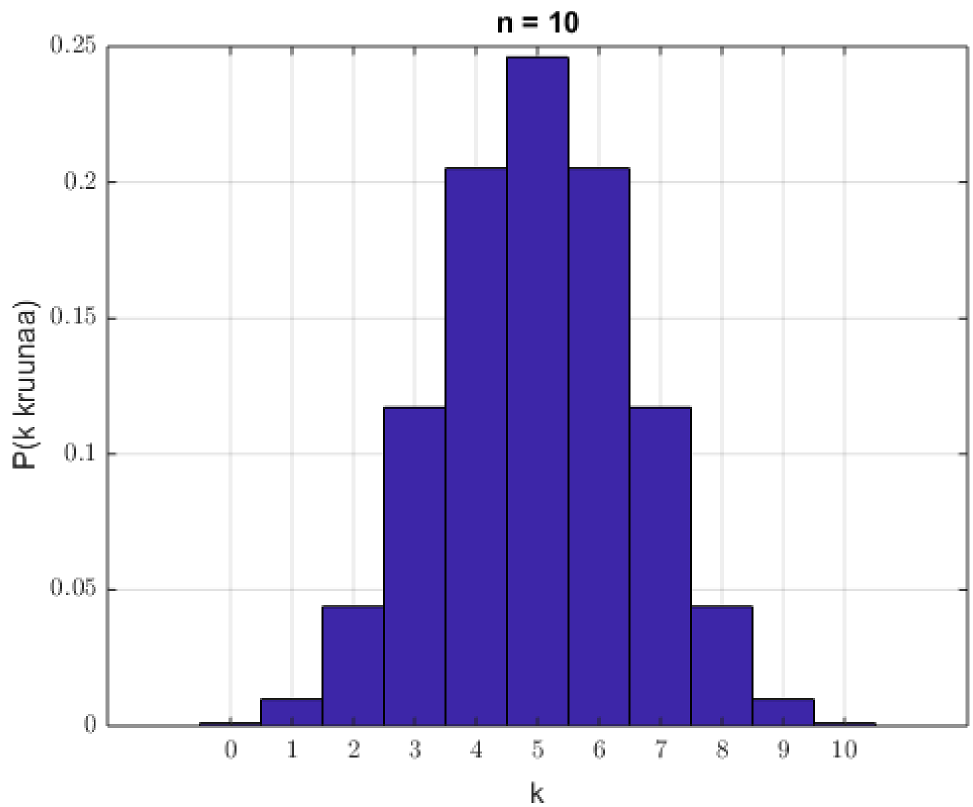
$$1303560$$

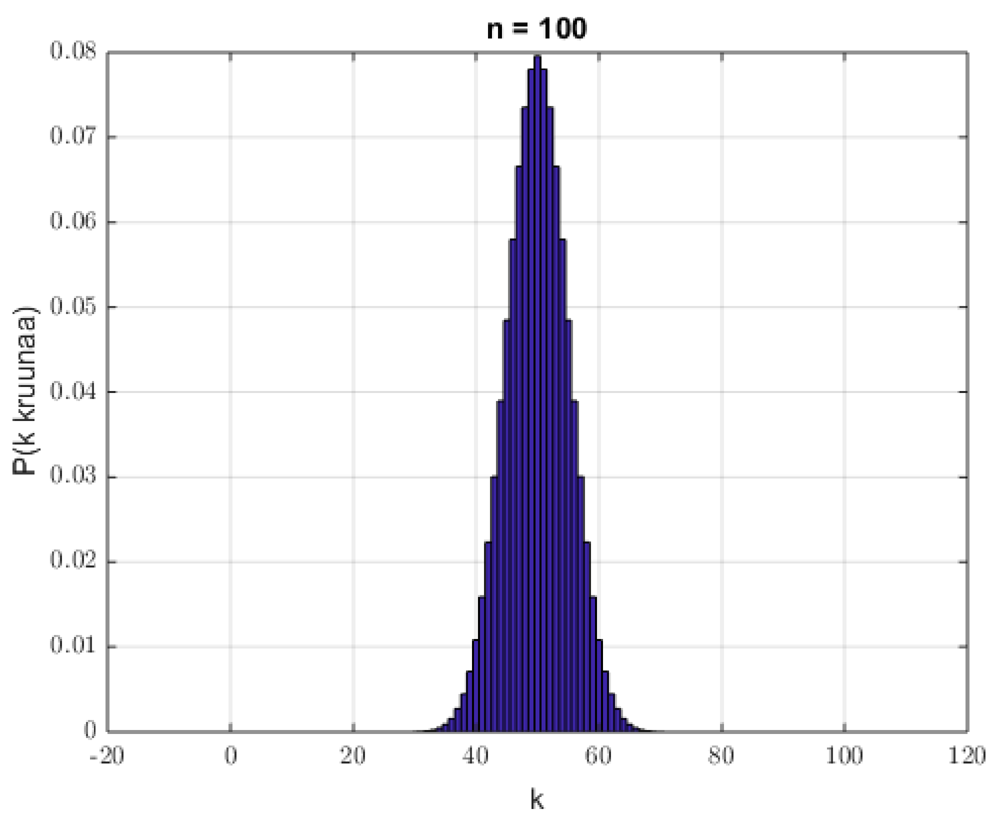
Ex. Flip a coin n times.
2^n different outcomes, number of outcomes with k heads and $n-k$ tails is

$$\binom{n}{k} = \binom{n}{n-k}$$

$$P(k \text{ heads}) = \frac{\binom{n}{k}}{2^n} = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}$$

$$k = 0, 1, 2, \ldots, n$$


n = 10
P(k kruunaa)
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6 7 8 9 10
k


n = 100
P(k kruunaa)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-20 0 20 40 60 80 100 120
k

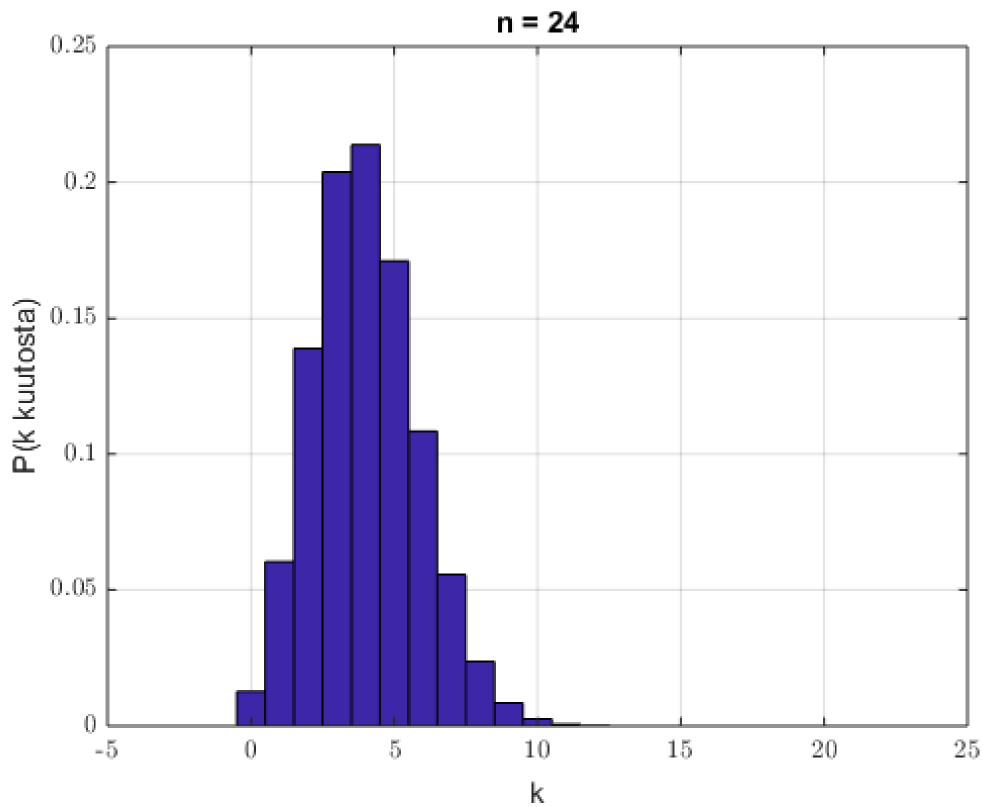
Ex. Throw dice n times.
6^n different outcomes, number of outcomes with k times 6 is

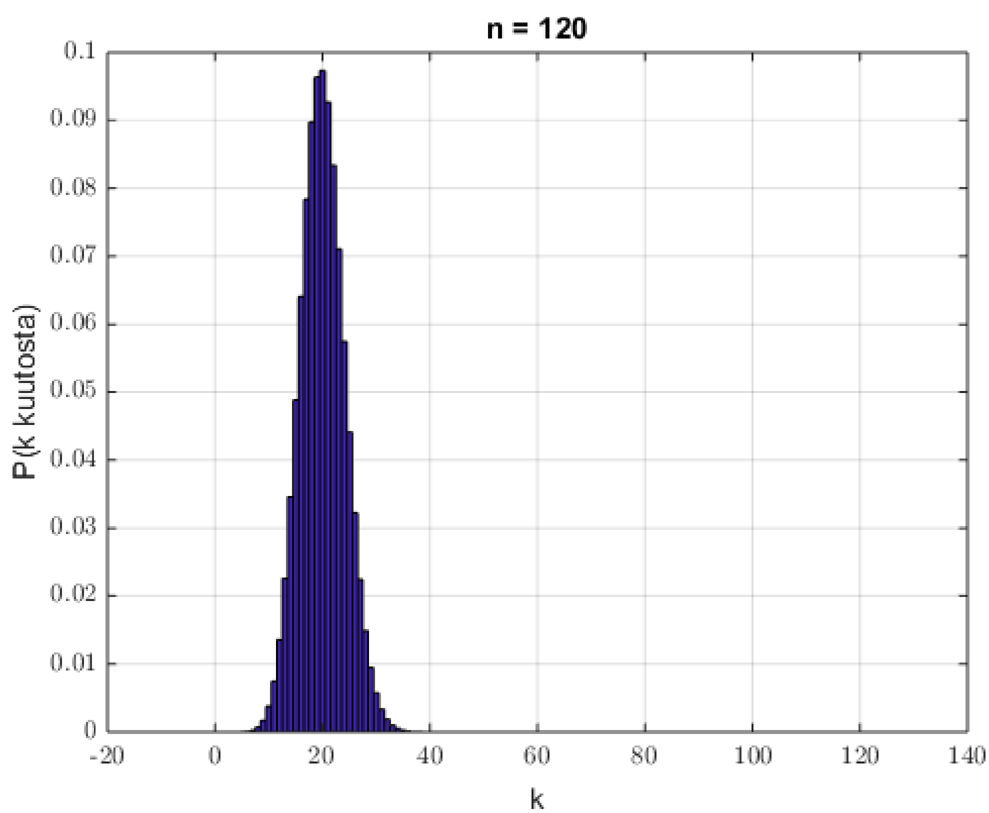
$$\binom{n}{k} \cdot 5^{n-k}$$

(places of 6:s $\cdot$ results of other $n-k$ throws)

$$P(k \text{ times } 6) = \frac{\binom{n}{k} \cdot 5^{n-k}}{6^n} = \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

$$k = 0, 1, 2, \ldots, n$$


n = 24
P(k kuutosta)
0.25
0.2
0.15
0.1
0.05
0
-5
0
5
10
15
20
25
k


n = 120
P(k kuutosta)
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-20
0
20
40
60
80
100
120
140
k

Binomial distribution

Random experiment with two possible outcomes, 1 and 0.
If

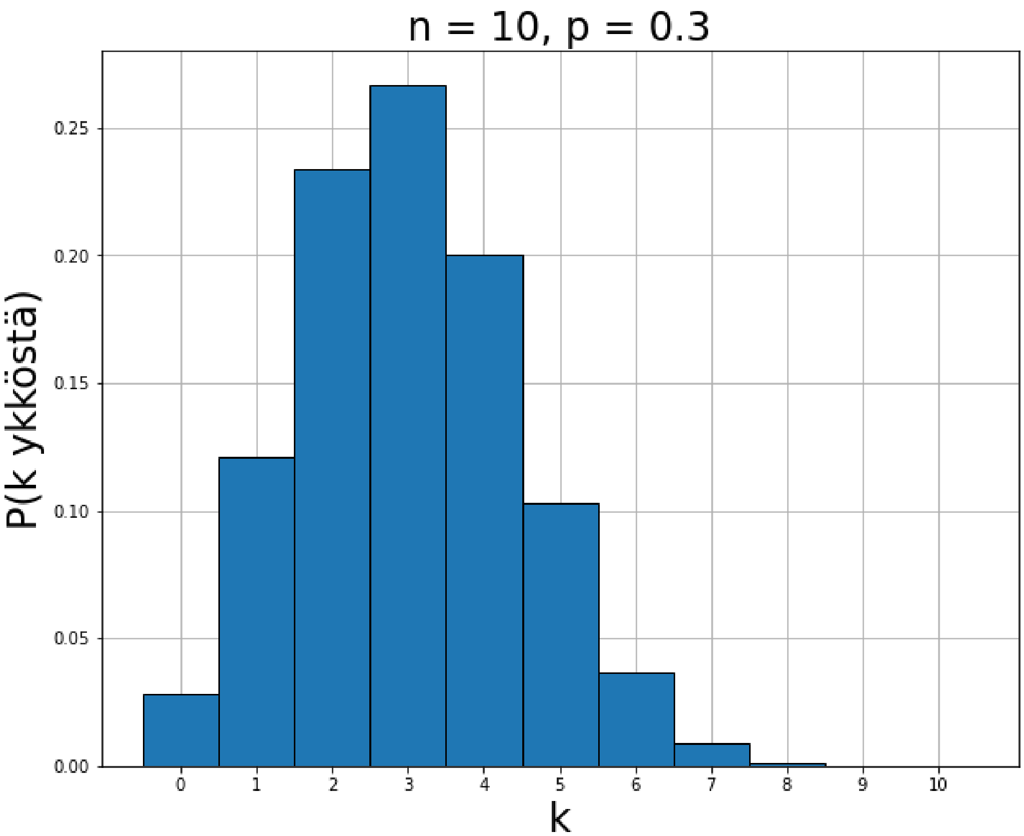
$$P(1) = p \quad \text{and} \quad P(0) = 1 - p$$

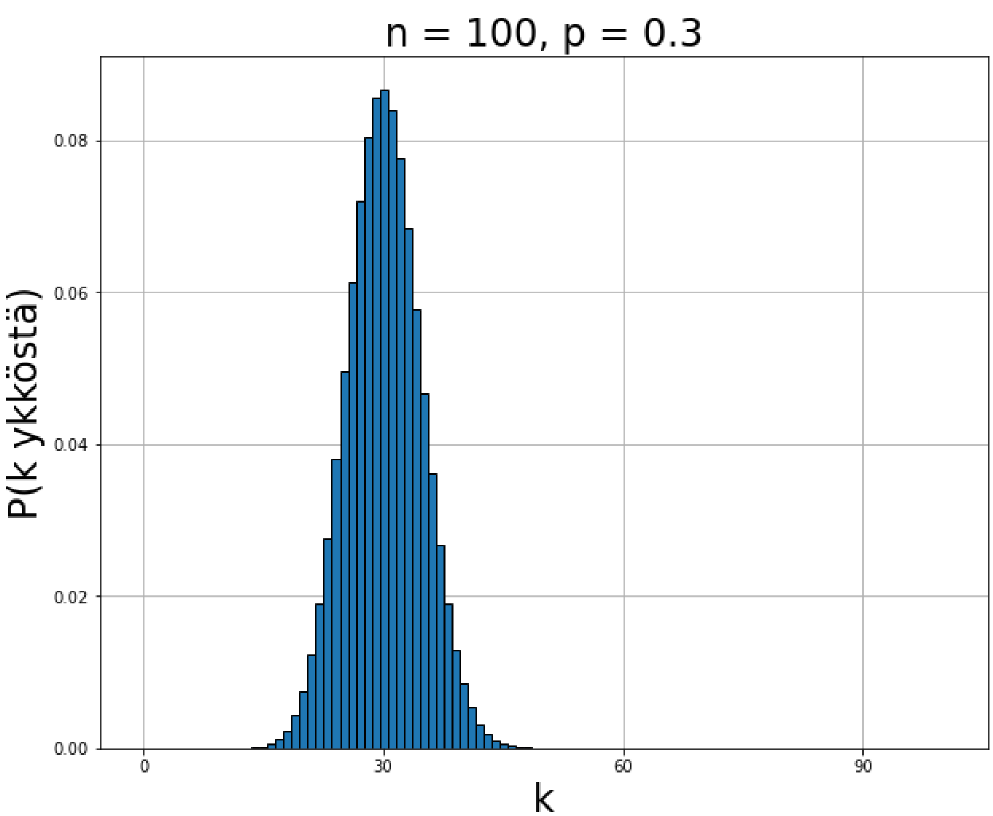
and if the experiment is repeated n times, then

$$P(k \text{ times } 1) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$k = 0, 1, 2 \ldots, n$$

Idea: outcome is a vector of length n of zeros and ones. Probability of an outcome with k ones and $n-k$ zeros is $p^k(1-p)^{n-k}$, and there are $\binom{n}{k}$ such outcomes


n = 10, p = 0.3
P(k ykköstä)
k
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7
8
9
10


n = 100, p = 0.3
P(k ykköstä)
k
0.00
0.02
0.04
0.06
0.08
0
30
60
90

Ex: Vakioveikkaus (Finnish football pool, 13 matches, 3 alternatives for each)

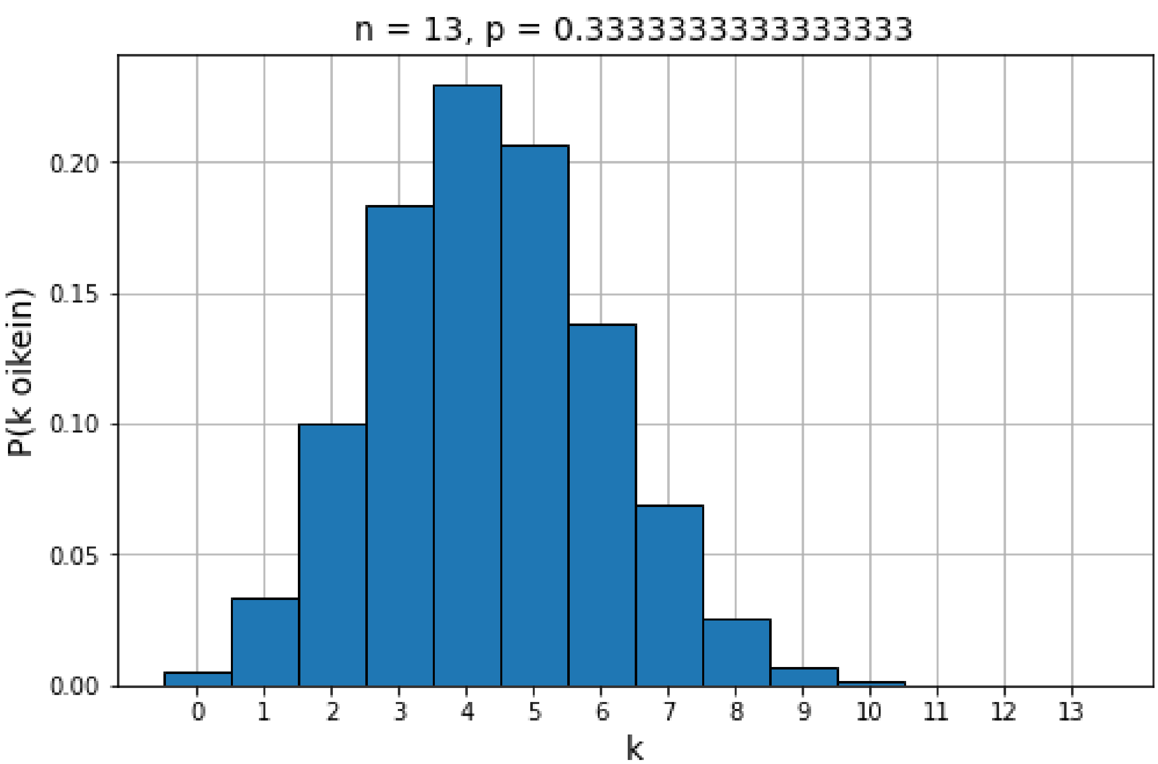
1.	1 x 2
2.	1 x 2
3.	1 x 2
4.	1 x 2
5.	1 x 2
6.	1 x 2
7.	1 x 2
8.	1 x 2
9.	1 x 2
10.	1 x 2
11.	1 x 2
12.	1 x 2
13.	1 x 2

$$n = 13,\ p = P(\text{right choise}) = \frac{1}{3}$$

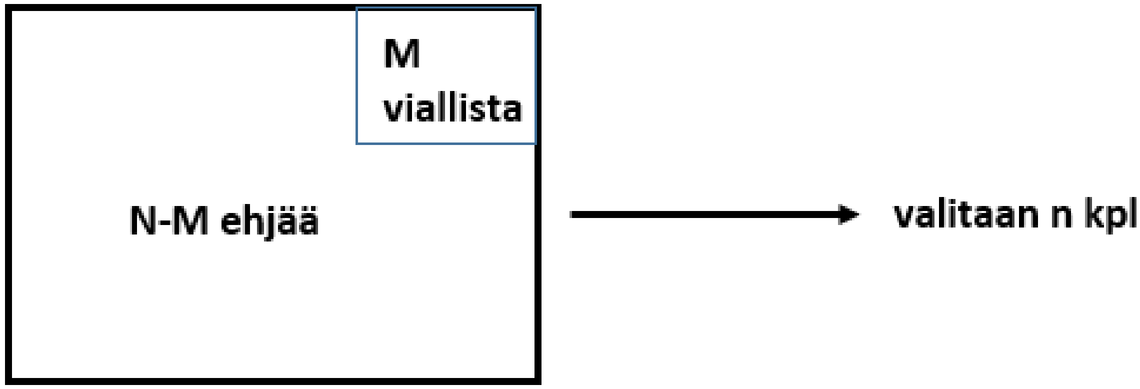
$$P(k \text{ right choices}) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \binom{13}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{13-k}$$

$$k = 0, 1, 2, ..., 13$$



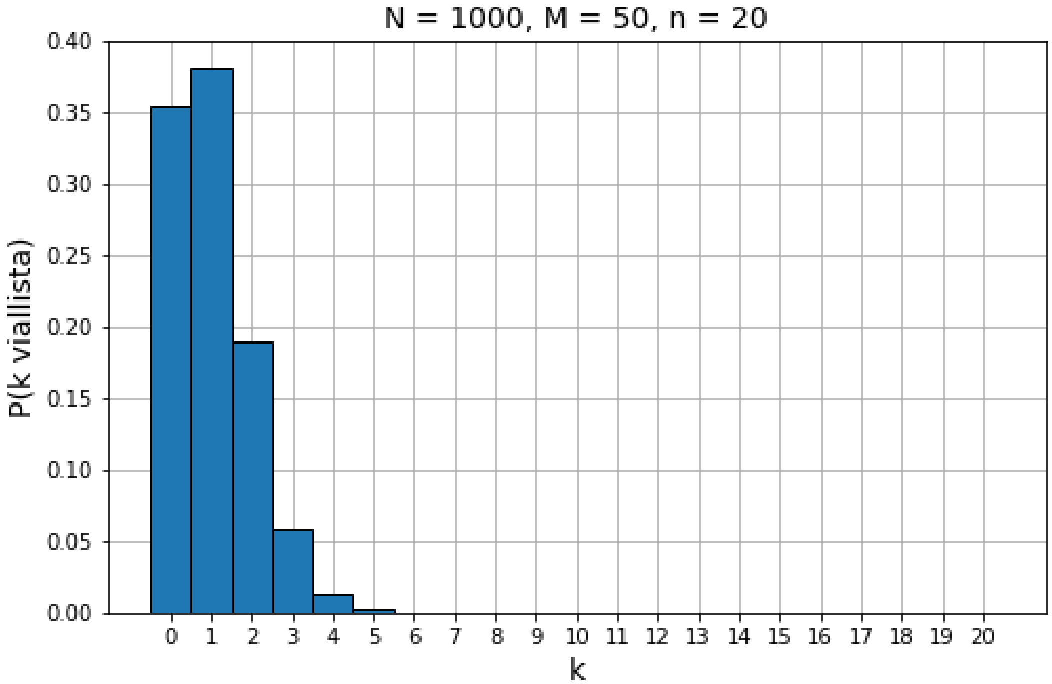
Ex. Quality control, N items, M defective, check n

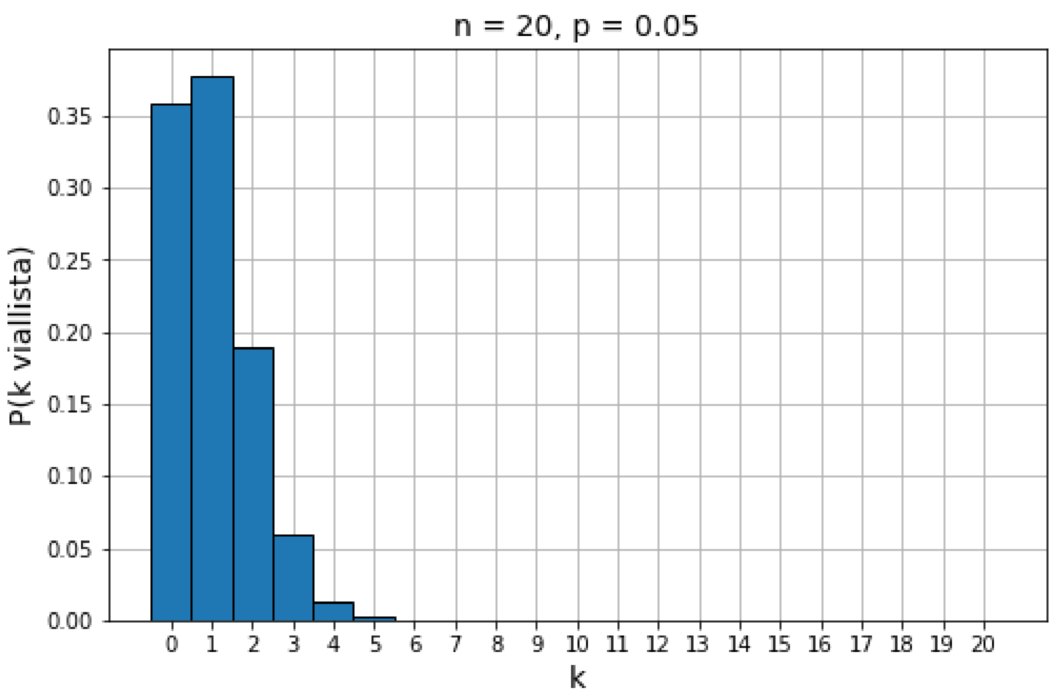


$$P(k \text{ defectives}) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}, \; k = 0, 1, \ldots, n$$

If N is large, then $p = P(\text{defective}) \approx \dfrac{M}{N}$
and

$$P(k \text{ defectives}) \approx \binom{n}{k} p^k (1-p)^{n-k}$$

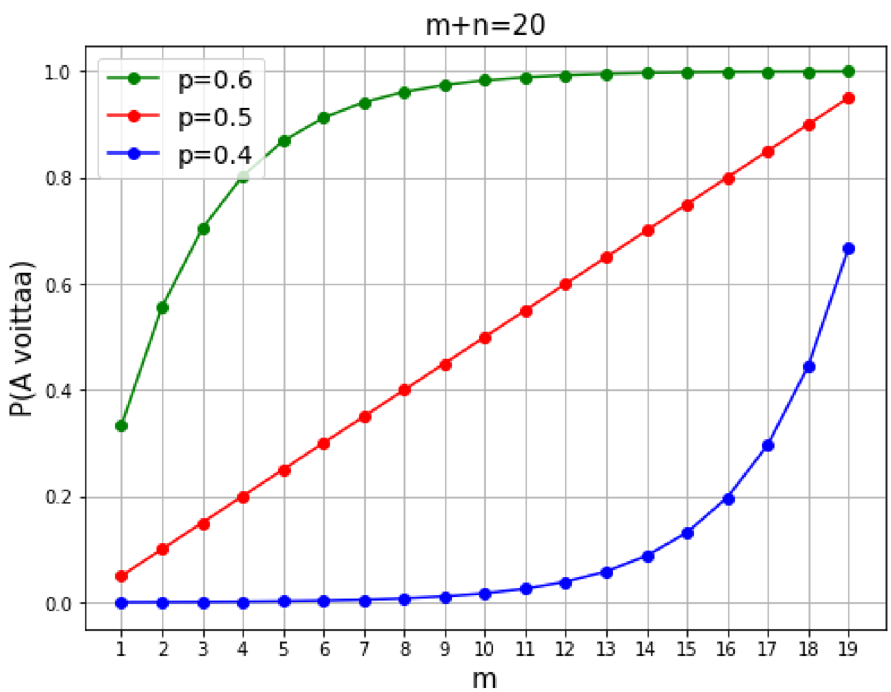

N = 1000, M = 50, n = 20
P(k viallista)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
k

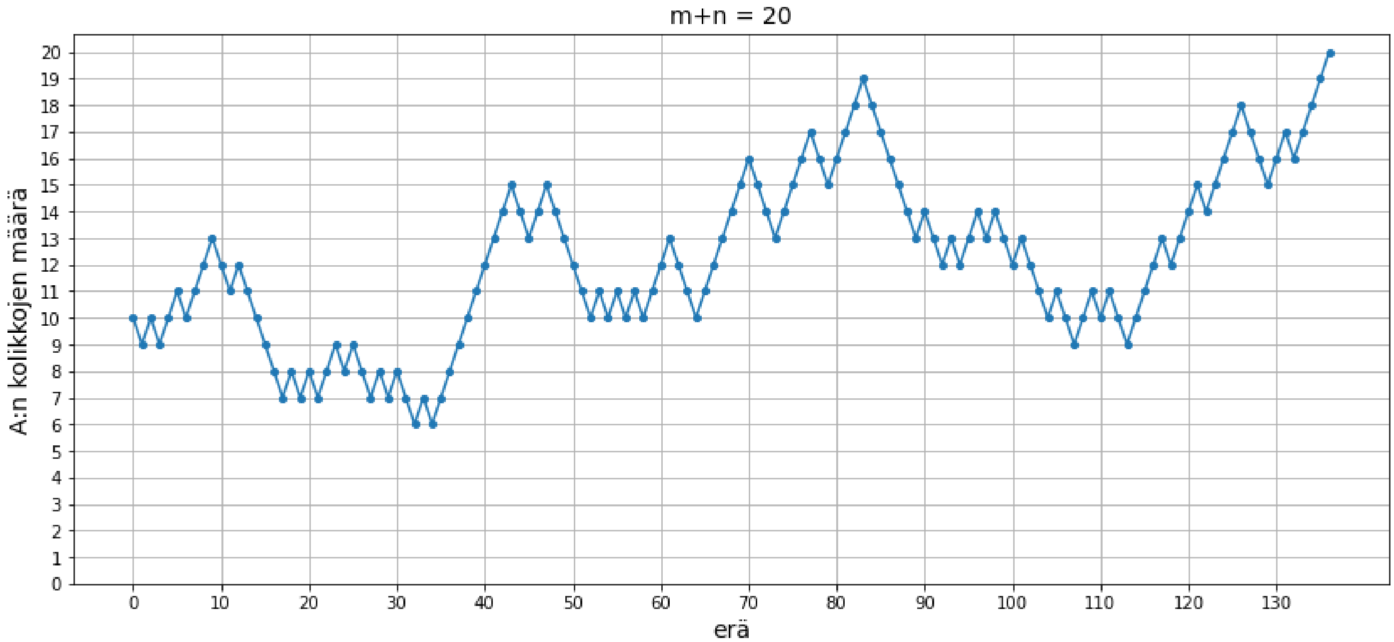

n = 20, p = 0.05
P(k viallista)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
k

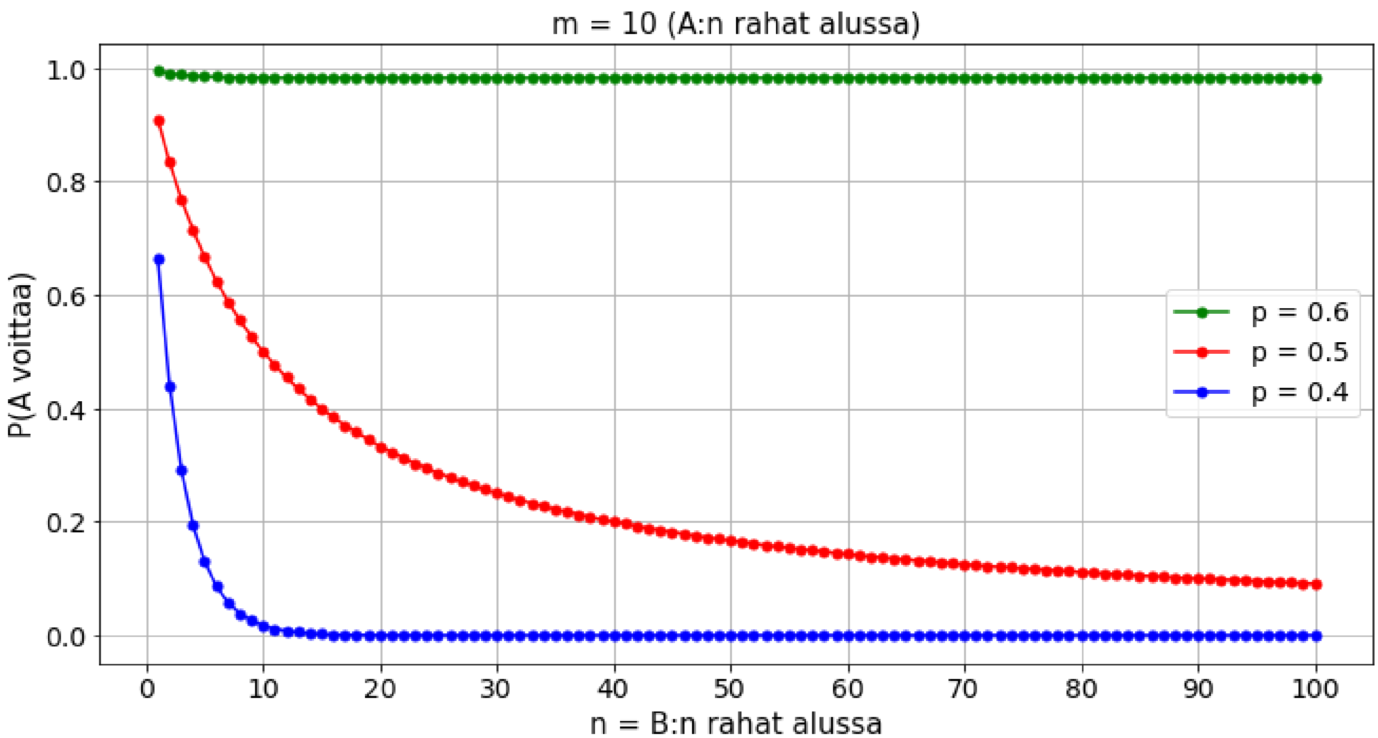
Ex: Gambler's ruin. A and B play a game which A wins with probability p and B wins with probability $1-p$. Loser gives one coin to the winner, and games are continued until one player doesn't have any coins left. Calculate the probability that A wins, if he has initially n coins and B has $N-n$ coins.

If $\alpha = (1-p)/p$, then A wins with probability

$$P_A = \begin{cases} \dfrac{1-\alpha^m}{1-\alpha^{m+n}}, & \text{jos } \alpha \neq 1 \text{ eli } p \neq 0.5 \\[2ex] \dfrac{m}{m+n}, & \text{jos } \alpha = 1 \text{ eli } p = 0.5 \end{cases}$$







Game against the bank:

if $p \leq 0.5$, then $P_A \to 0$, as n increases

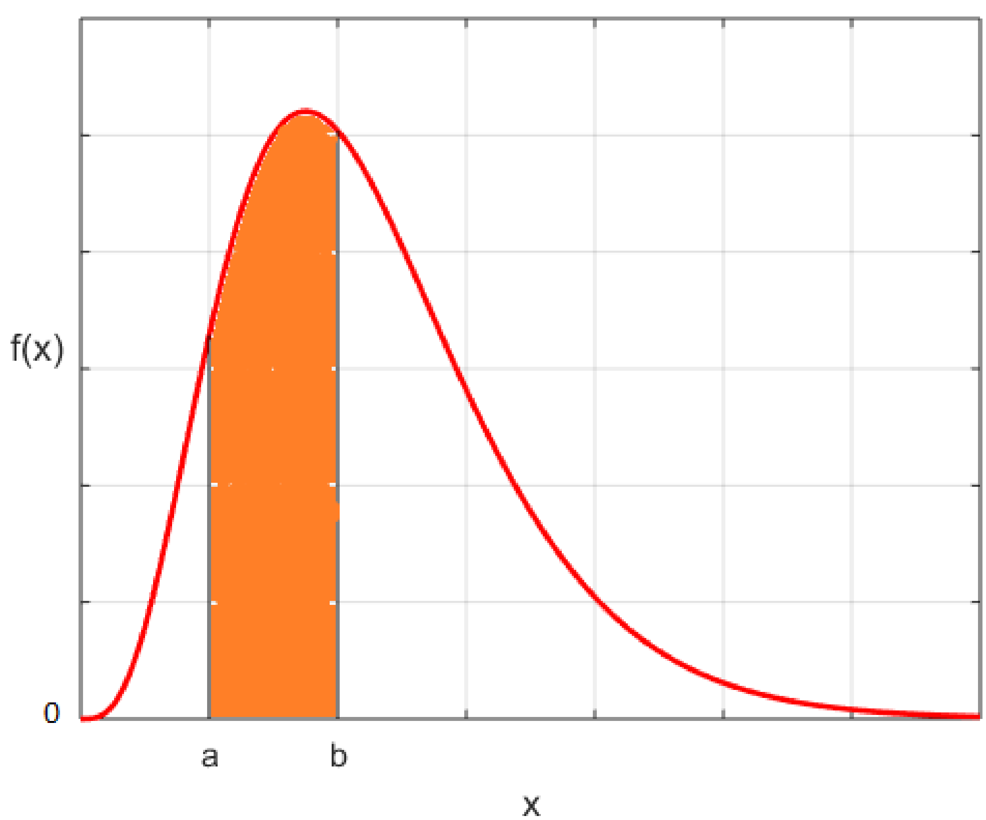
if $p > 0.5$, then $P_A \to 1 - \alpha^m$, as n increases

Probability density

Function $f(x)$ is the probability density function, pdf, for values of X, if the probability

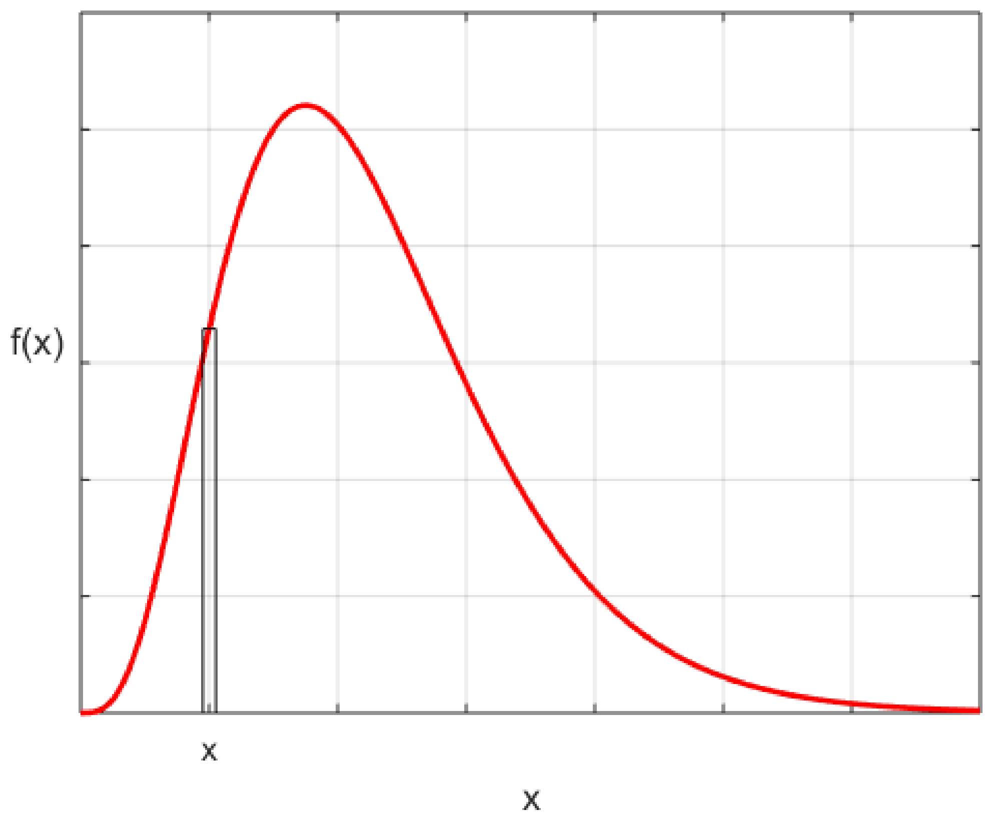
$$P(\text{values of } X \text{ are on the interval } a \ldots b) = \int_a^b f(x)\,dx$$

=the area between the graph of $f(x)$ and x-axis on the interval $a \ldots b$



Note 1. The total area between graph of $f(x)$ and x-axis $= 1$

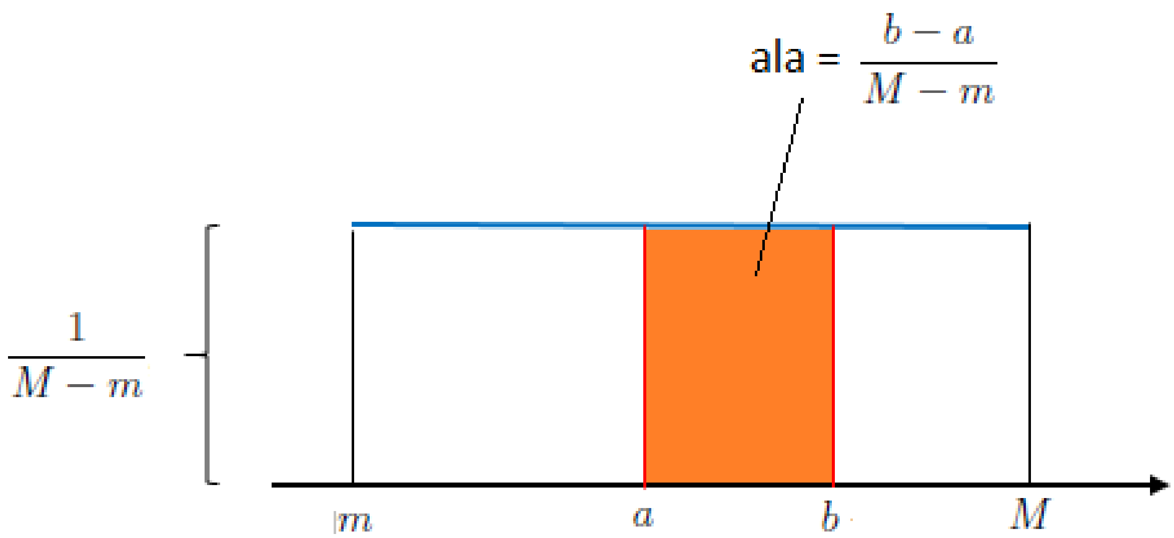
Note 2: $f(x)$ is probability density i.e the values of X are on the small interval $x \pm dx/2$ of length dx with probablity $f(x)dx$



UNIFORM DISTRIBUTION

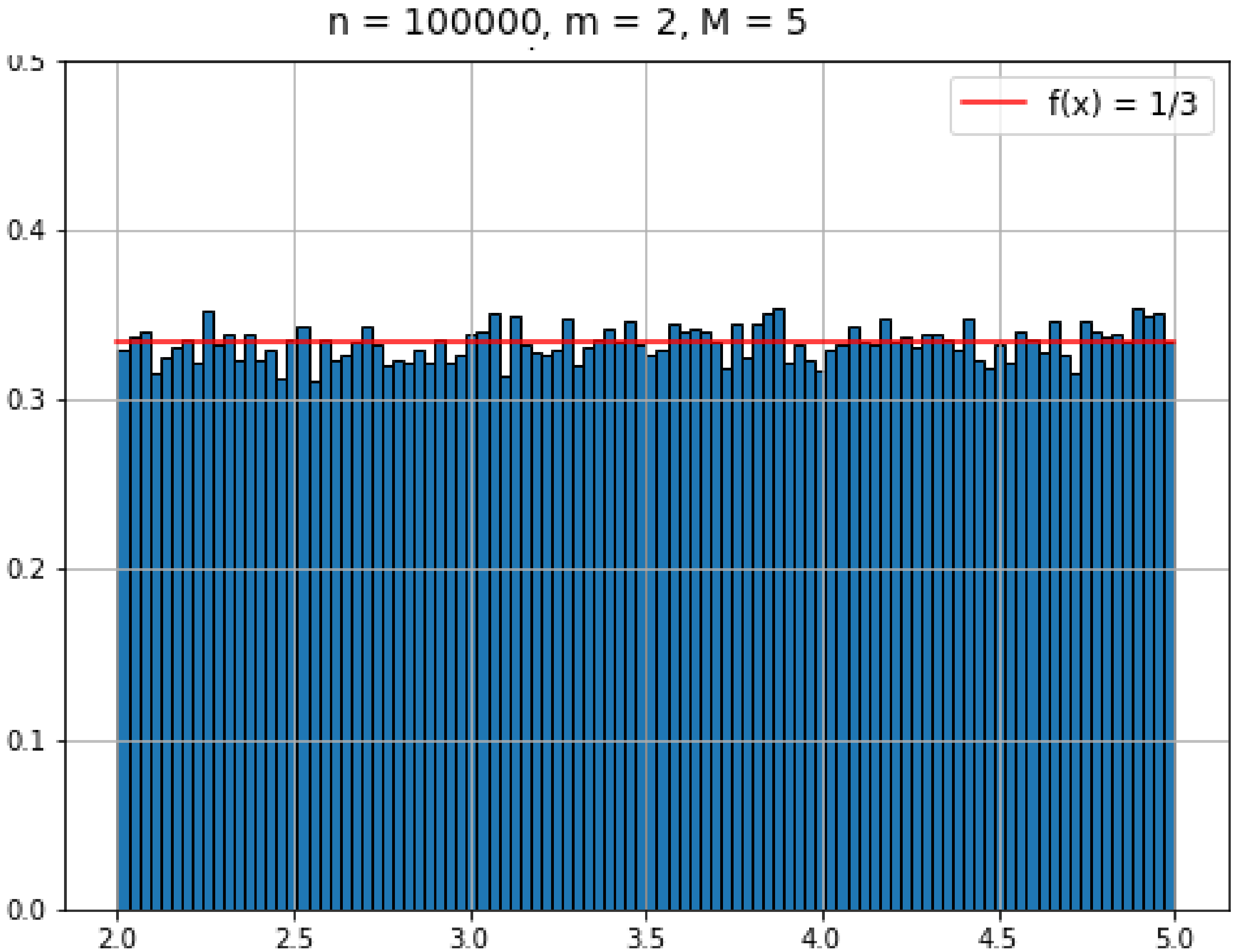
The values of X are uniformly distributed on the interval $m \ldots M$, if the probability density function

$$f(x) = \frac{1}{M-m}, \quad x = m \ldots M$$



i.e the value of X is on the interval $a \ldots b$ with probability

$$\int_a^b f(x)\,dx = \frac{b-a}{M-m}$$


n = 100000, m = 2, M = 5
f(x) = 1/3
0.0
0.1
0.2
0.3
0.4
0.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

NORMAL DISTRIBUTION

Normal distributions are often used in the natural sciences to represent random variables whose distributions are not known. Their importance is partly due to the central limit theorem. It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a random variablewhose distribution converges to a normal distribution as the number of samples increases. Therefore, physical quantities that are expected to be the sum of many independent processes, such as measurement errors, often have distributions that are nearly normal.

Values of X are normally distributed with mean μ and standard deviation σ (variance $= \sigma^2$)

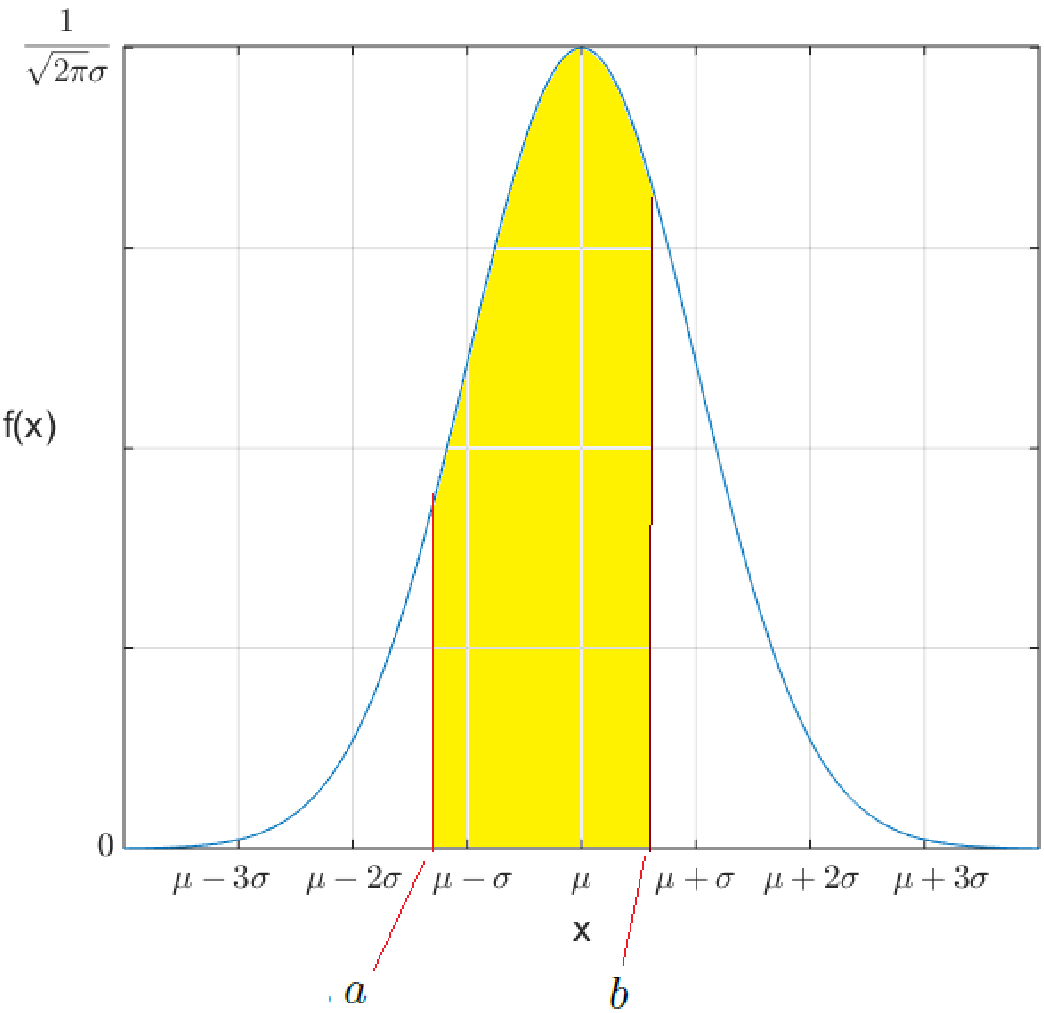
$$X \sim N(\mu, \sigma^2)$$

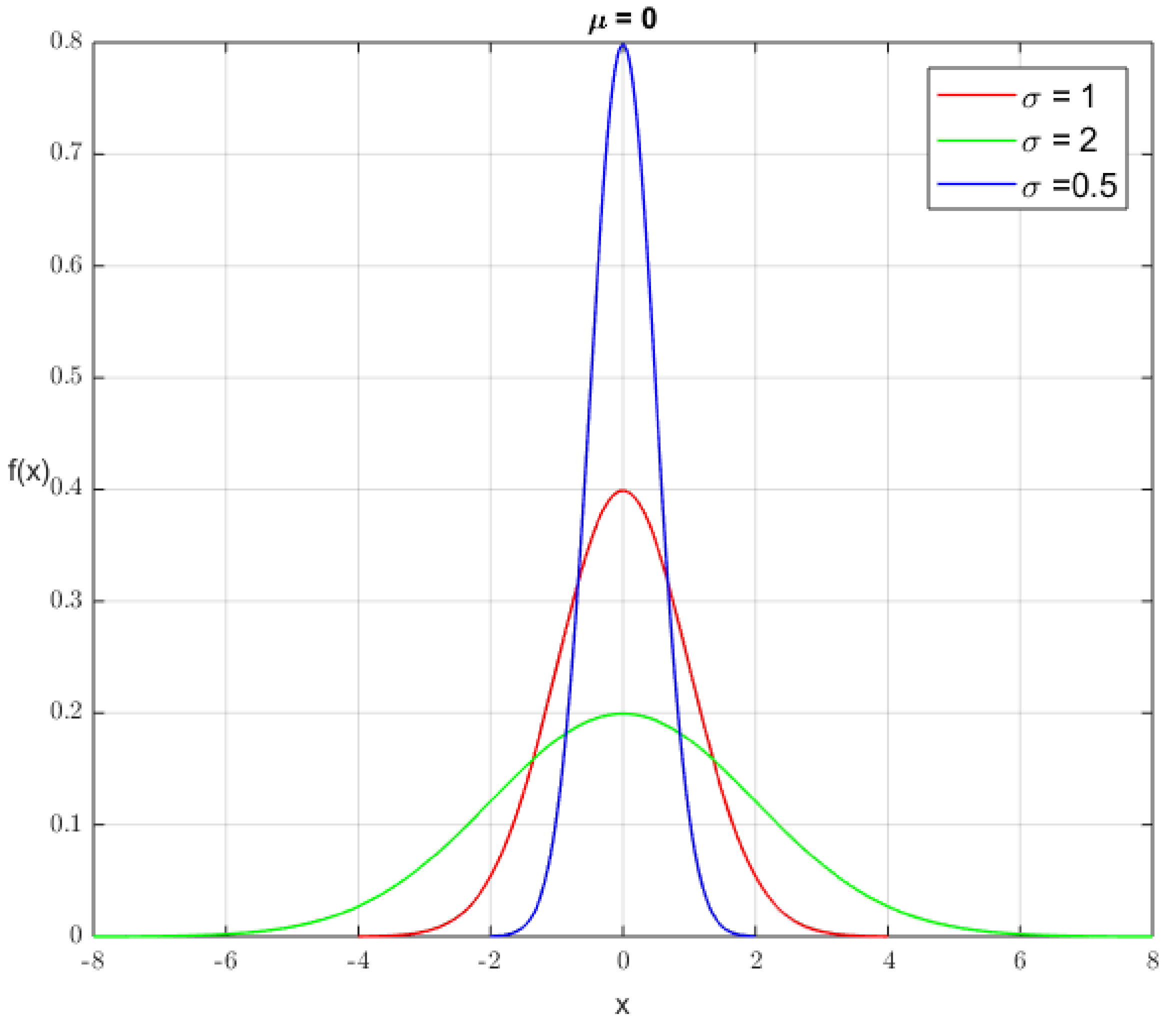
if the probability density function is Gaussian

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

where $\exp(...) = e^{\cdots}$ is the exponential function i.e the values of X are on the interval $a \ldots b$ with probability

$$\int_a^b f(x)\,dx$$





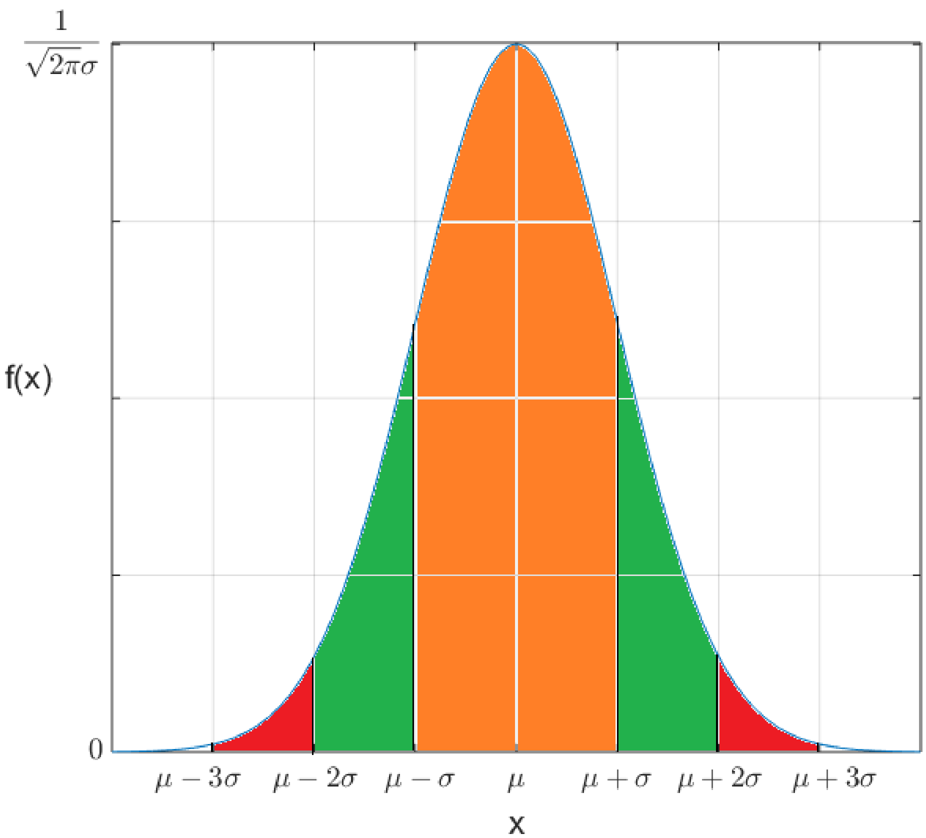
For example, results of measurements are (usually) approximately normally distributed

The value of X is with probability

68 % on the interval $\mu \pm \sigma$

95 % on the interval v $\mu \pm 2\sigma$

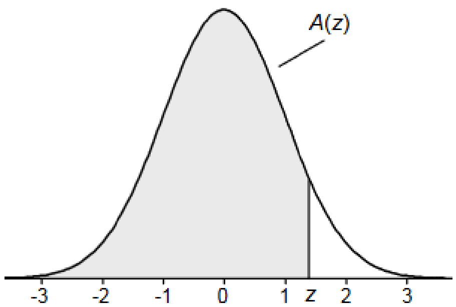
99.7 % on the interval $\mu \pm 3\sigma$

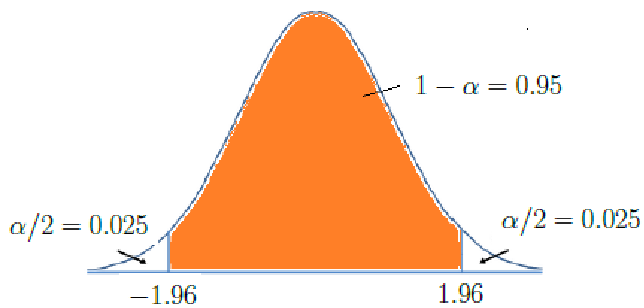


Critical value z_α:

$$P(X \leq \mu + z_\alpha \cdot \sigma) = 1 - \alpha$$

$$P(X \text{ between } \mu \pm z_{\alpha/2} \cdot \sigma) = 1 - \alpha$$





z	0.00	0.01	0.02	0.03	0.04	0.05	$\boxed{0.06}$	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
$\boxed{1.9}$	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	$\boxed{0.9750}$	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

$$z = 1.96 : P(X \leq \mu + 1.96\sigma) = 0.975 \rightarrow$$

$$P(X > \mu + 1.96\sigma) = P(X < \mu - 1.96\sigma) = 0.025$$

$$\rightarrow P(X \text{ between } \mu \pm 1.96\sigma) = 0.95$$

$$\rightarrow z_{\alpha/2} = 1.96, \text{ when } \alpha = 0.05$$

$$z = 1.00 : P(X \leq \mu + 1\sigma) = 0.8413 \rightarrow$$

$$P(X > \mu + 1\sigma) = P(X < \mu - 1\sigma) = 0.1587$$

$$\rightarrow P(X \text{ between } \mu \pm 1\sigma) = 0.6826$$

$$z = 3.00 : P(X \leq \mu + 3\sigma) = 0.9987 \rightarrow$$

$$P(X > \mu + 3\sigma) = P(X < \mu - 3\sigma) = 0.0013$$

$$\rightarrow P(X \text{ between } \mu \pm 3\sigma) = 0.9974$$

Note: numbers $x_1, x_2, \ldots, x_n$

mean

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{n} x_k,$$

standard deviation

$$s = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (x_k - \overline{x})^2}$$

variance

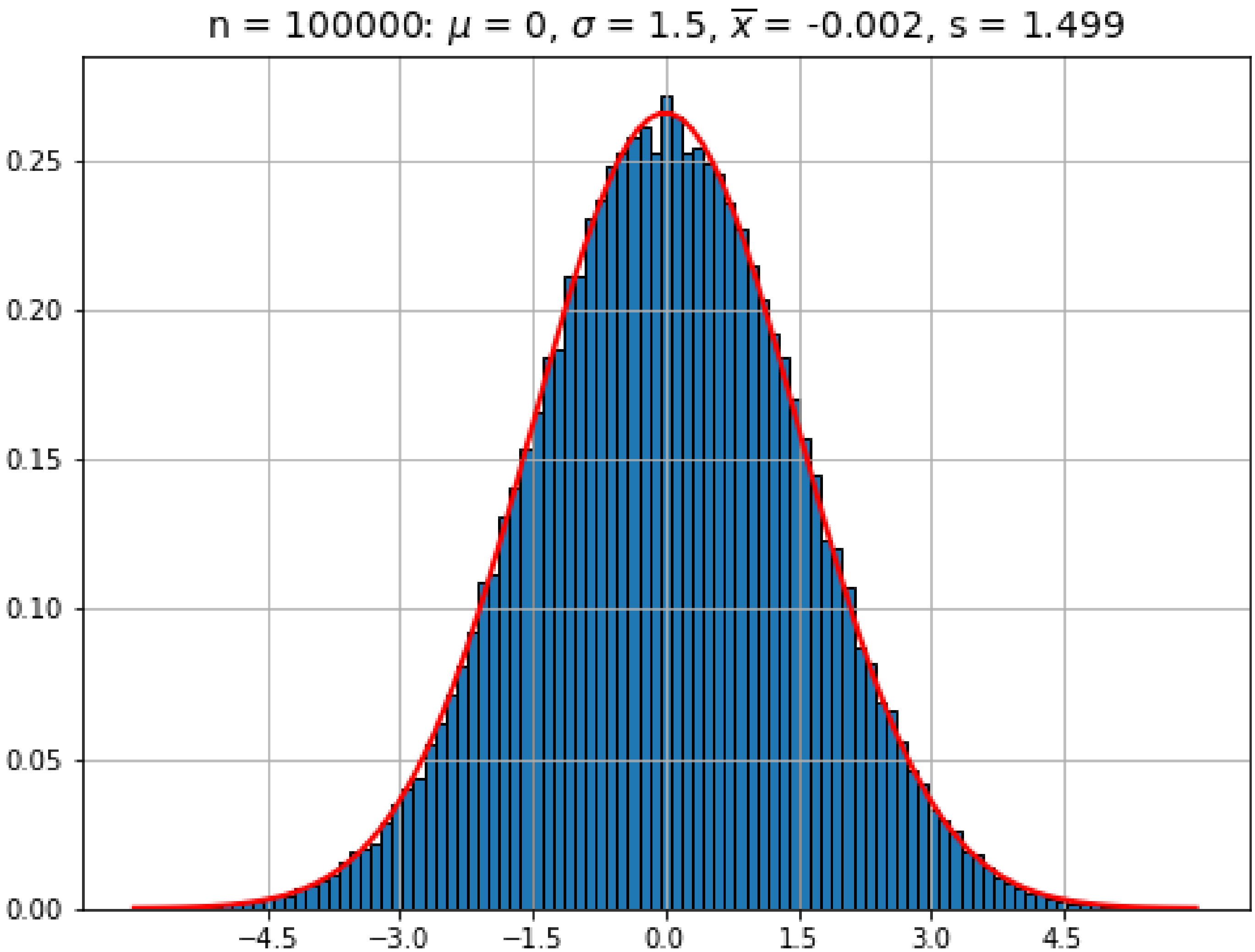
$$s^2 = \sum_{k=1}^{n} (x_k - \overline{x})^2$$

If numbers $x_1, x_2, \ldots, x_n$ are approximately $N(\mu, \sigma^2)$-distributed, then

$$\overline{x} \approx \mu \text{ and } s \approx \sigma$$

and

n. 68 % of the numbers are between $\overline{x} \pm s$
n. 95 % of the numbers are between $\overline{x} \pm 2s$
n. 99.7 % of the numbers are between $\overline{x} \pm 3s$



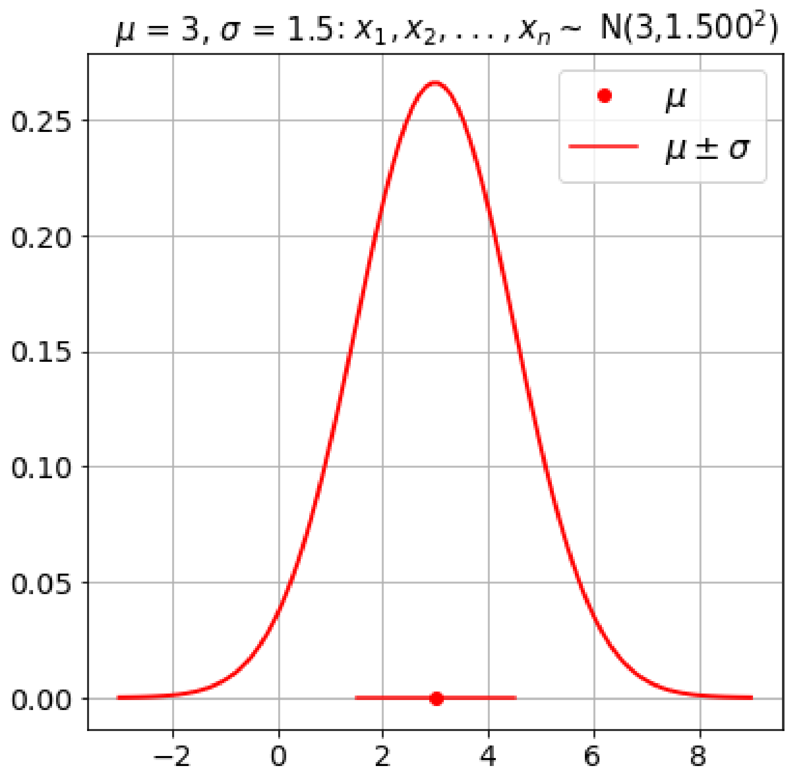
Ex: Sum and average of normally distributed numbers are also normally distributed
If

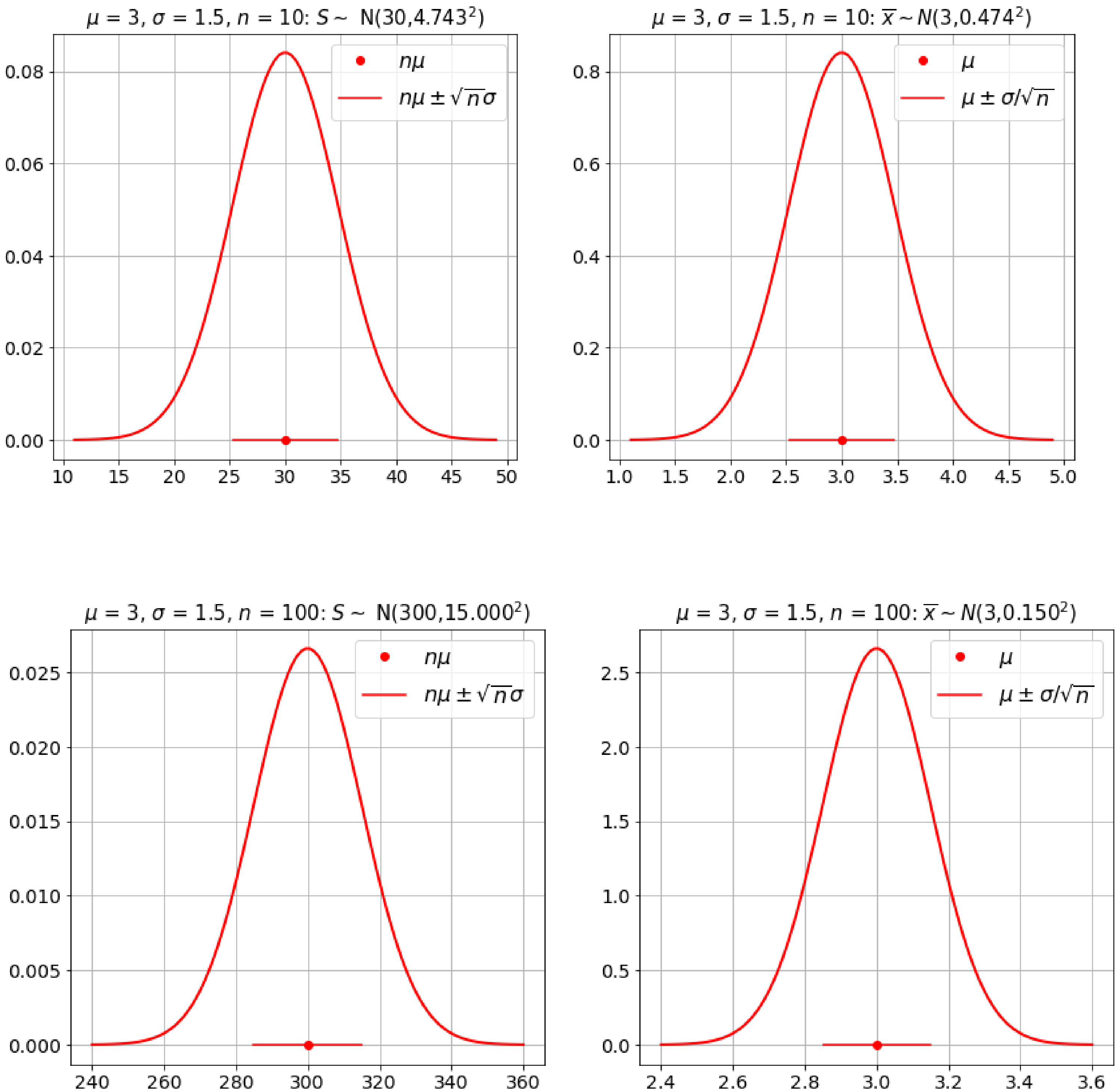
$$x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$$

then

$$S = \sum_{k=1}^{n} x_k \sim N(n\mu, (\sigma \cdot \sqrt{n})^2)$$

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{n} x_k \sim N(\mu, (\sigma/\sqrt{n})^2)$$




$\mu = 3$, $\sigma = 1.5$, $n = 10$: $S \sim$ N(30,4.743^2)
$n\mu$
$n\mu \pm \sqrt{n}\sigma$
$\mu = 3$, $\sigma = 1.5$, $n = 10$: $\overline{x} \sim N$(3,0.474^2)
μ
$\mu \pm \sigma/\sqrt{n}$
$\mu = 3$, $\sigma = 1.5$, $n = 100$: $S \sim$ N(300,15.000^2)
$n\mu$
$n\mu \pm \sqrt{n}\sigma$
$\mu = 3$, $\sigma = 1.5$, $n = 100$: $\overline{x} \sim N$(3,0.150^2)
μ
$\mu \pm \sigma/\sqrt{n}$

Ex: If true value is μ and measurement $x \sim N(\mu, \sigma^2)$, then with probability

68 % x is between $\mu \pm \sigma$

95 % x is between $\mu \pm 2\sigma$

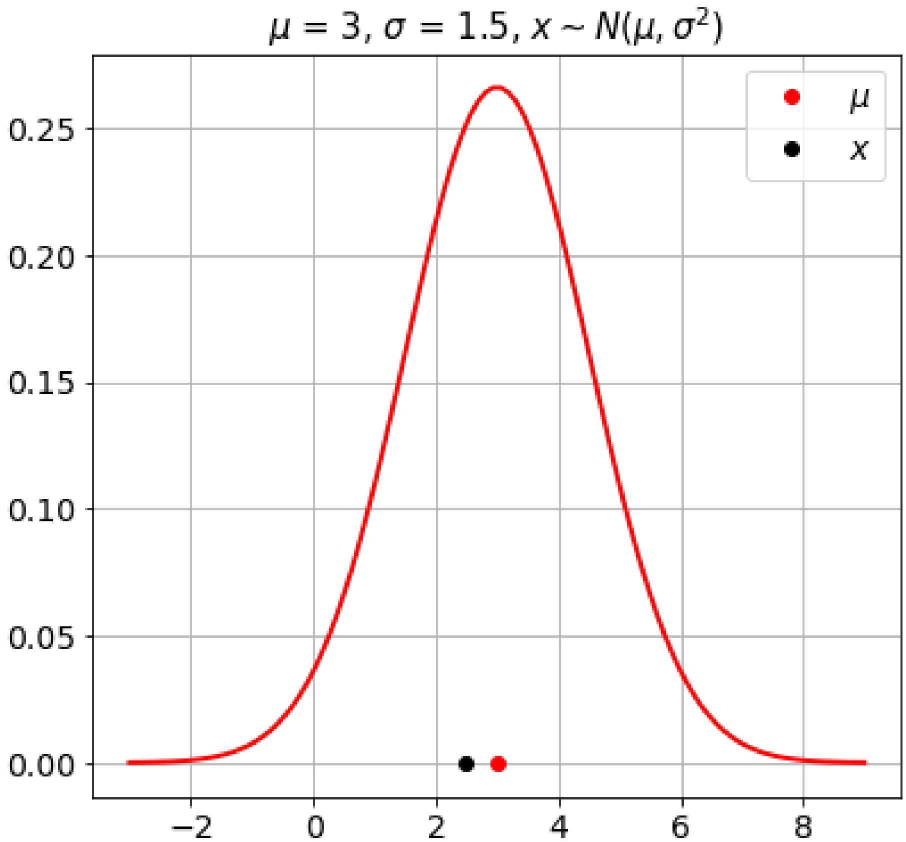
99.7 % x is between $\mu \pm 3\sigma$

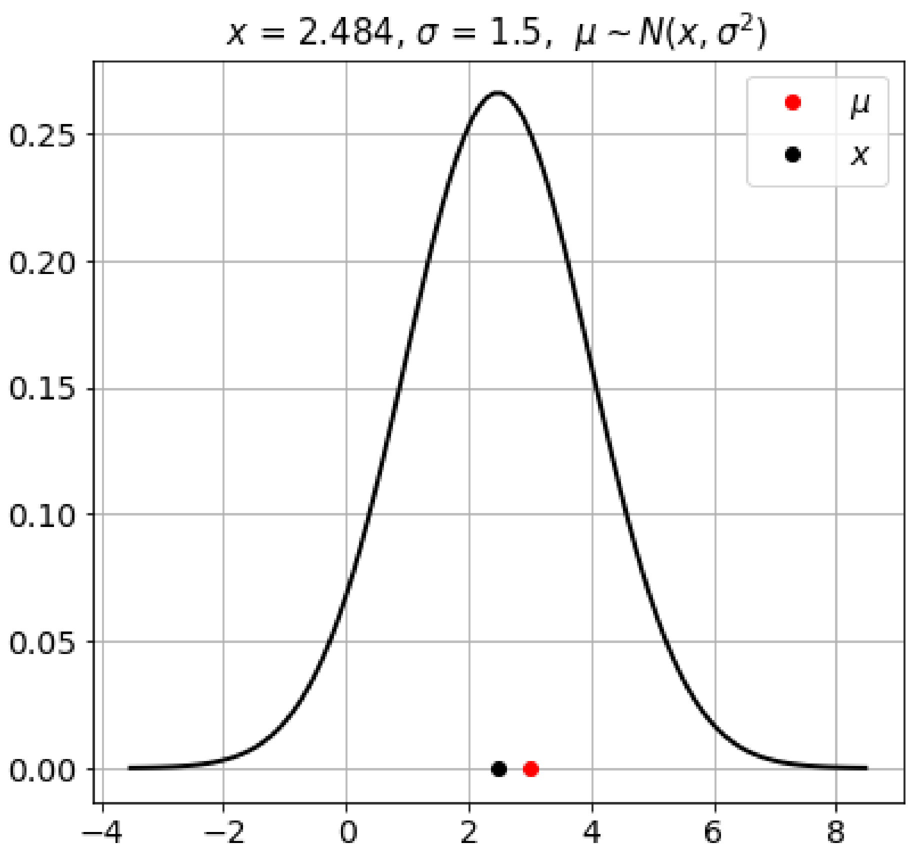
Conversely, measurement gives an estimate for the true value: $\mu \sim N(x, \sigma^2)$ i.e with probability

68 % μ is between $x \pm \sigma$

95 % μ is between $x \pm 2\sigma$

99.7 % μ is between $x \pm 3\sigma$


$\mu = 3, \sigma = 1.5, x \sim N(\mu, \sigma^2)$
μ
x
0.25
0.20
0.15
0.10
0.05
0.00
−2
0
2
4
6
8


$x = 2.484, \sigma = 1.5, \mu \sim N(x, \sigma^2)$
μ
x
0.25
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4
6
8

Ex: If true value is μ and measurements $x_1, x_2, ..., x_n \sim N(\mu, \sigma^2)$, then the average

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{n} x_k \sim N(\mu, (\sigma/\sqrt{n})^2)$$

i.e with probability

68 % $\overline{x}$ is between $\mu \pm \sigma/\sqrt{n}$

95 % $\overline{x}$ is between $\mu \pm 2\sigma/\sqrt{n}$

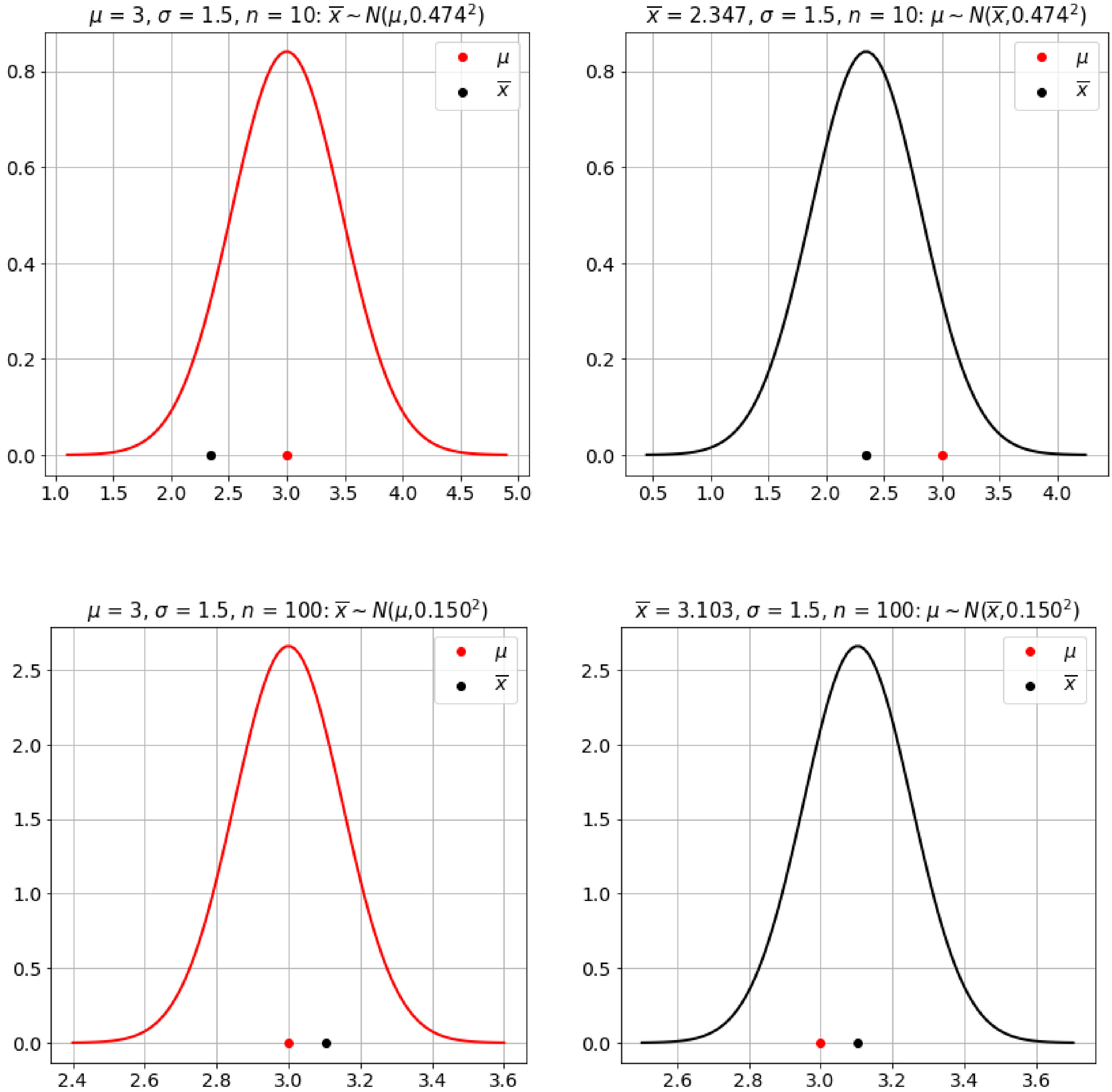
99.7 % $\overline{x}$ is between $\mu \pm 3\sigma/\sqrt{n}$

Conversely, $\mu \sim N(\overline{x}, (\sigma/\sqrt{n})^2$ i.e with probability

68 % μ is between $\overline{x} \pm \sigma/\sqrt{n}$

95 % μ is between $\overline{x} \pm 2\sigma/\sqrt{n}$

99.7 % μ is between $\overline{x} \pm 3\sigma/\sqrt{n}$


μ = 3, σ = 1.5, n = 10: $\overline{x} \sim N(\mu, 0.474^2)$
$\overline{x}$ = 2.347, σ = 1.5, n = 10: $\mu \sim N(\overline{x}, 0.474^2)$
μ = 3, σ = 1.5, n = 100: $\overline{x} \sim N(\mu, 0.150^2)$
$\overline{x}$ = 3.103, σ = 1.5, n = 100: $\mu \sim N(\overline{x}, 0.150^2)$
μ
$\overline{x}$

Ex: True value μ, measurements

$$x_1 \sim N(\mu, \sigma_1^2), \quad x_2 \sim N(\mu, \sigma_2^2)$$

i.e $\mu \sim N(x_1, \sigma_1^2), \quad \mu \sim N(x_2, \sigma_2^2)$

Weighted average

$$x = a_1 \, x_1 + a_2 \, x_2, \quad a_1 + a_2 = 1$$

is also normally distributed, $x \sim N(\mu, \sigma^2)$, where

$$\sigma^2 = a_1^2 \, \sigma_1^2 + a_2^2 \, \sigma_2^2$$

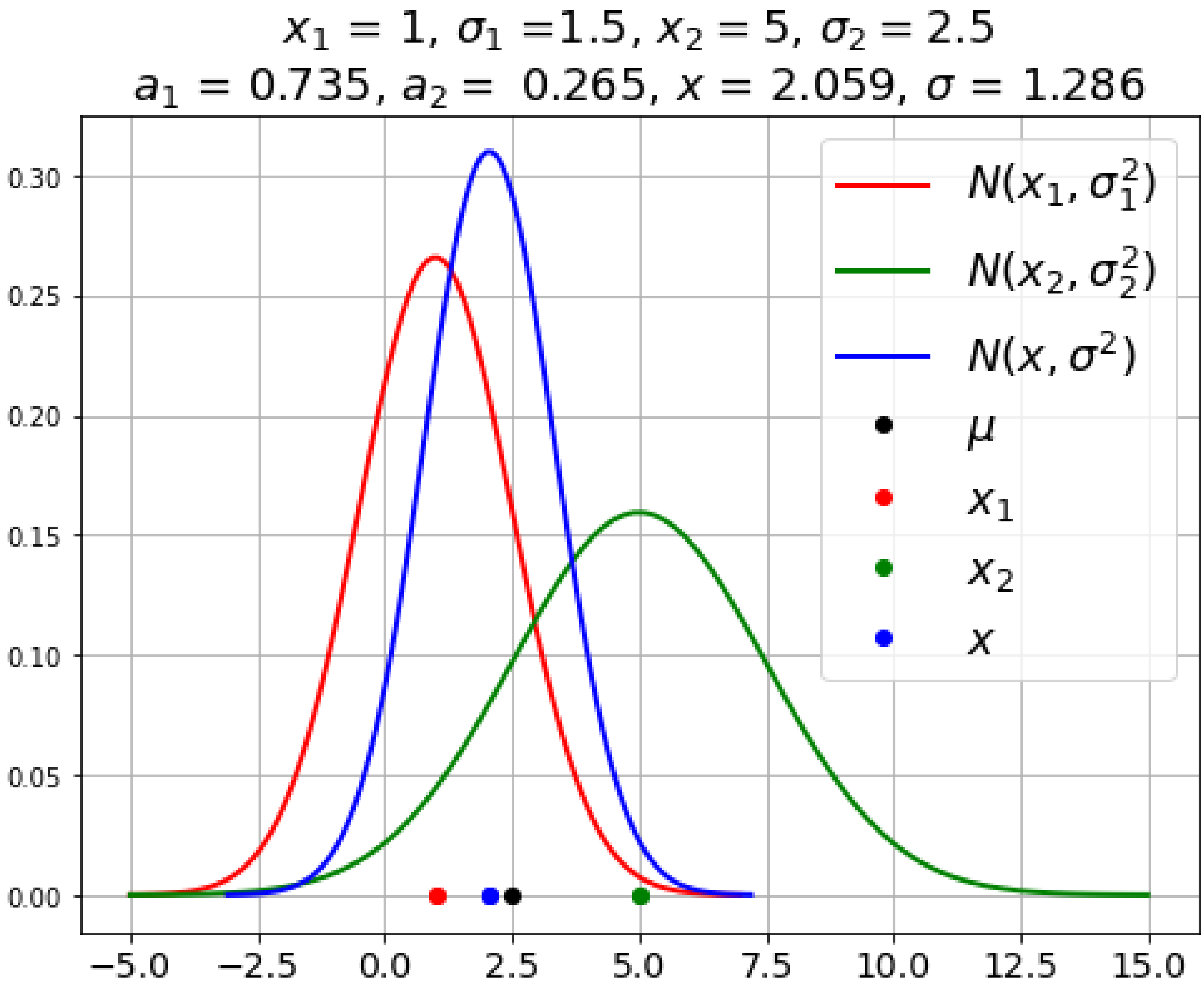
i.e $\mu \sim N(x, \sigma^2)$

Variance σ^2 is smallest (most accurate estimate for μ), when

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

Then

$$\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$



Note: if $K = a_2$ ("Kalman gain"), then
$x = a_1x_1 + a_2x_2 = x_1 + K(x_2 - x_1)$
and $\sigma^2 = (1 - K)\sigma_1^2$

Note:

$$a_1 = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}, \quad a_2 = \frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}, \quad \sigma^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}$$

Similarly, measurements

$$x_k \sim N(\mu, \sigma_k^2), \quad k = 1, 2, \ldots, n$$

Weighted average

$$x = a_1 x_1 + \cdots + a_n x_n, \quad a_1 + \cdots + a_n = 1$$

$x \sim N(\mu, \sigma^2)$, where

$$\sigma^2 = a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2$$

i.e $\mu \sim N(x, \sigma^2)$

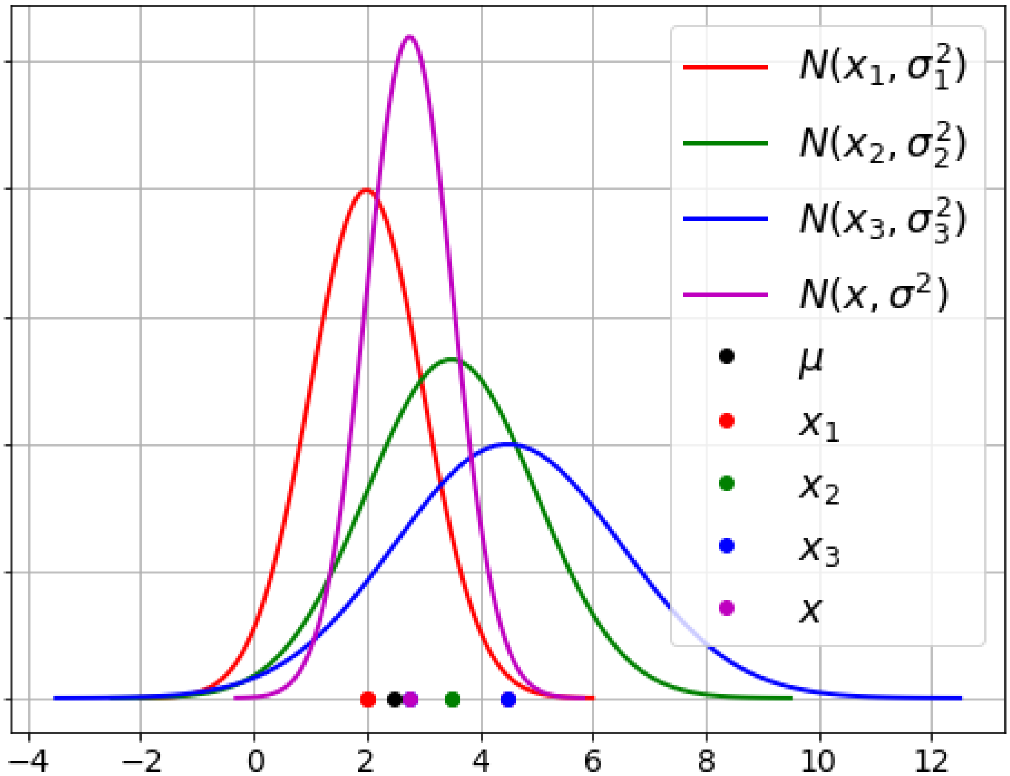
Variance σ^2 is smallest (most accurate estimate for μ), when

$$a_k = \frac{1/\sigma_k^2}{1/\sigma_1^2 + \cdots + 1/\sigma_n^2}, \quad k = 1, 2, \ldots, n$$

Then

$$\sigma^2 = \frac{1}{1/\sigma_1^2 + \cdots + 1/\sigma_n^2}$$

$x_1 = 2$, $\sigma_1 = 1$, $x_2 = 3.5$, $\sigma_2 = 1.5$, $x_3 = 4.5$, $\sigma_3 = 2.0$
$a_1 = 0.590$, $a_2 = 0.262$, $a_3 = 0.148$, $x = 2.762$, $\sigma = 0.768$



Note: If variances are equal,

$$\sigma_1 = \sigma_2 = \cdots = \sigma_n = s$$

then

$$a_1 = a_2 = \cdots = a_n = \frac{1}{n}$$

i.e x is the average and

$$\sigma^2 = \frac{s^2}{n}$$

Confidence intervals

for the mean μ and standard deviation σ of the normal distribution.

Take sample $x_1, x_2, ..., x_n \sim N(\mu, \sigma^2)$

Calculate mean and sample standard deviation

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

With 95% probability / 95 % of the samples are such that

mean μ is in the interval

$$\overline{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \ldots \quad \overline{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

standard deviation σ is in the interval

$$\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}} \cdot s \quad \ldots \quad \sqrt{\frac{n-1}{\chi^2_{\alpha/2}}} \cdot s$$

where $t_{\alpha/2}$ is critical value of the t-distribution and $\chi^2_{1-\alpha/2}, \chi^2_{\alpha/2}$ are critical values of the χ^2-distribution with degree of freedom $n-1$ and confidence level $\alpha = 0.05$.

vapaus aste k	kriittinen arvo $t_{\alpha/2}$
1	12.71
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306
9	2.262
10	2.228

vapaus aste k	kriittinen arvo $t_{\alpha/2}$
11	2.201
12	2.179
13	2.160
14	2.145
15	2.131
16	2.120
17	2.110
18	2.101
19	2.093
20	2.086

vapaus aste k	kriittinen arvo $t_{\alpha/2}$
21	2.080
22	2.074
23	2.069
24	2.064
25	2.060
26	2.056
27	2.052
28	2.048
29	2.045
30	2.042

vapaus aste k	kriittinen arvo $t_{\alpha/2}$
40	2.021
50	2.009
60	2.000
80	1.990
100	1.984
120	1.980
∞	1.960

df	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$
1	0.001	5.024
2	0.051	7.378
3	0.216	9.348
4	0.484	11.143
5	0.831	12.833
6	1.237	14.449
7	1.690	16.013
8	2.180	17.535
9	2.700	19.023
10	3.247	20.483

	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$
11	3.816	21.920
12	4.404	23.337
13	5.009	24.736
14	5.629	26.119
15	6.262	27.488
16	6.908	28.845
17	7.564	30.191
18	8.231	31.526
19	8.907	32.852
20	9.591	34.170

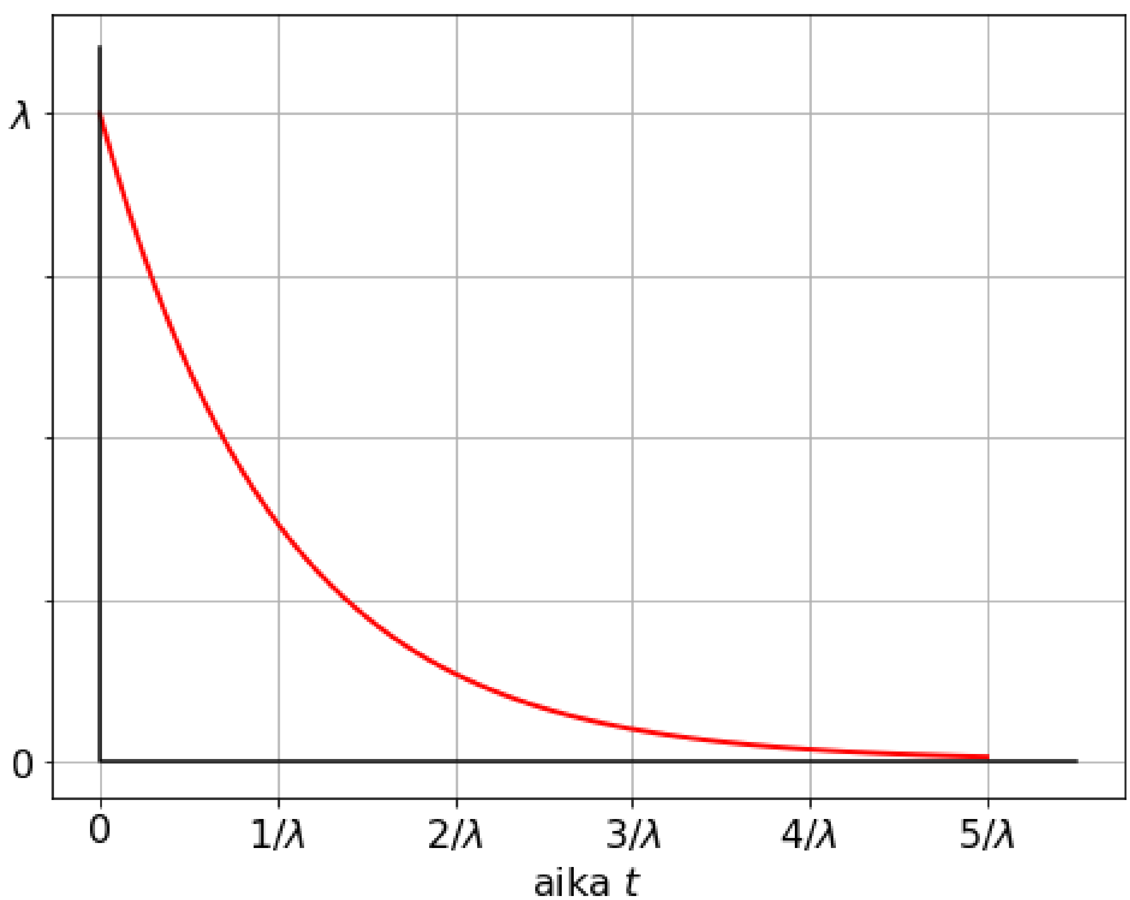
	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$
21	10.283	35.479
22	10.982	36.781
23	11.689	38.076
24	12.401	39.364
25	13.120	40.646
26	13.844	41.923
27	14.573	43.195
28	15.308	44.461
29	16.047	45.722
30	16.791	46.979

Eksponential distribution

The Exponential Distribution is typically used to model times between events or arrivals. The parameter λ is the average number of events or arrivals in a given unit time interval.

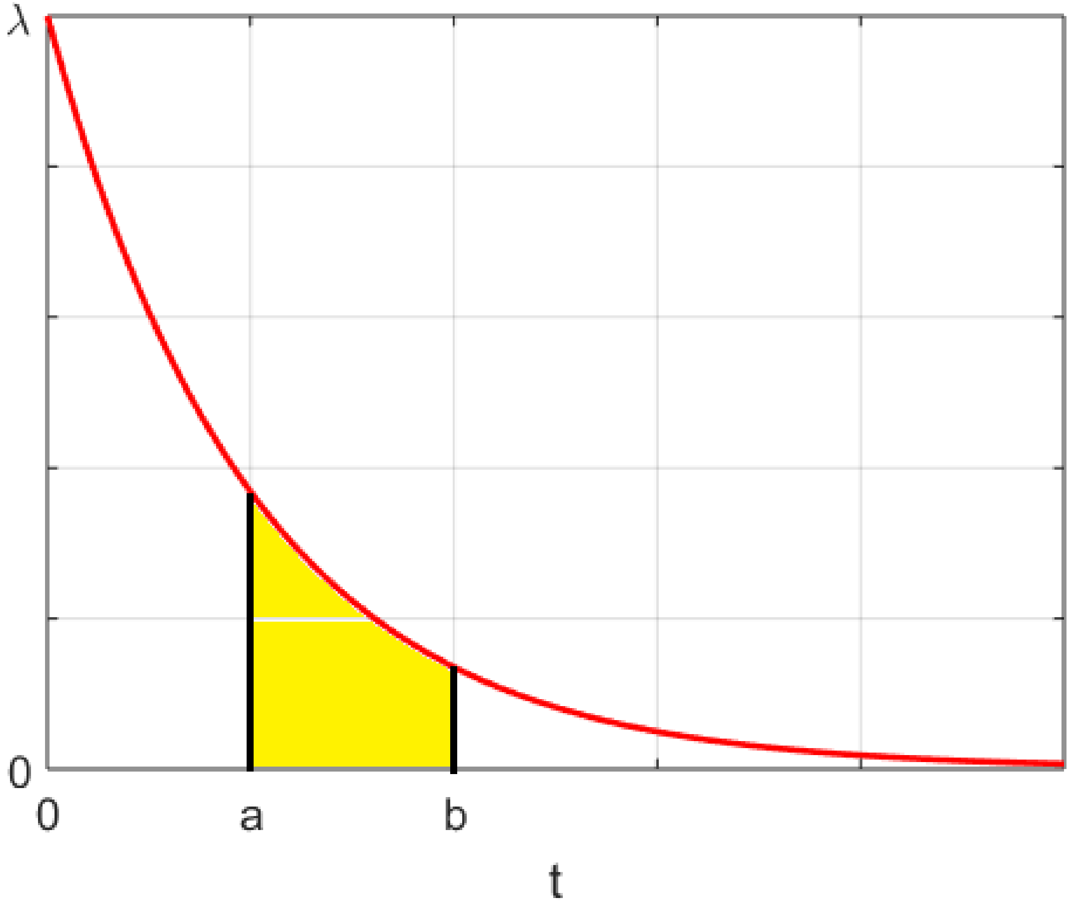
The values of X are exponentially distributed with parameter λ, $X \sim \mathrm{Exp}(\lambda)$, if the probability density function is

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0$$



i.e the values of X are in the interval $a \ldots b$ with probability

$$\int_a^b f(t)\,dt = e^{-\lambda a} - e^{-\lambda b}$$

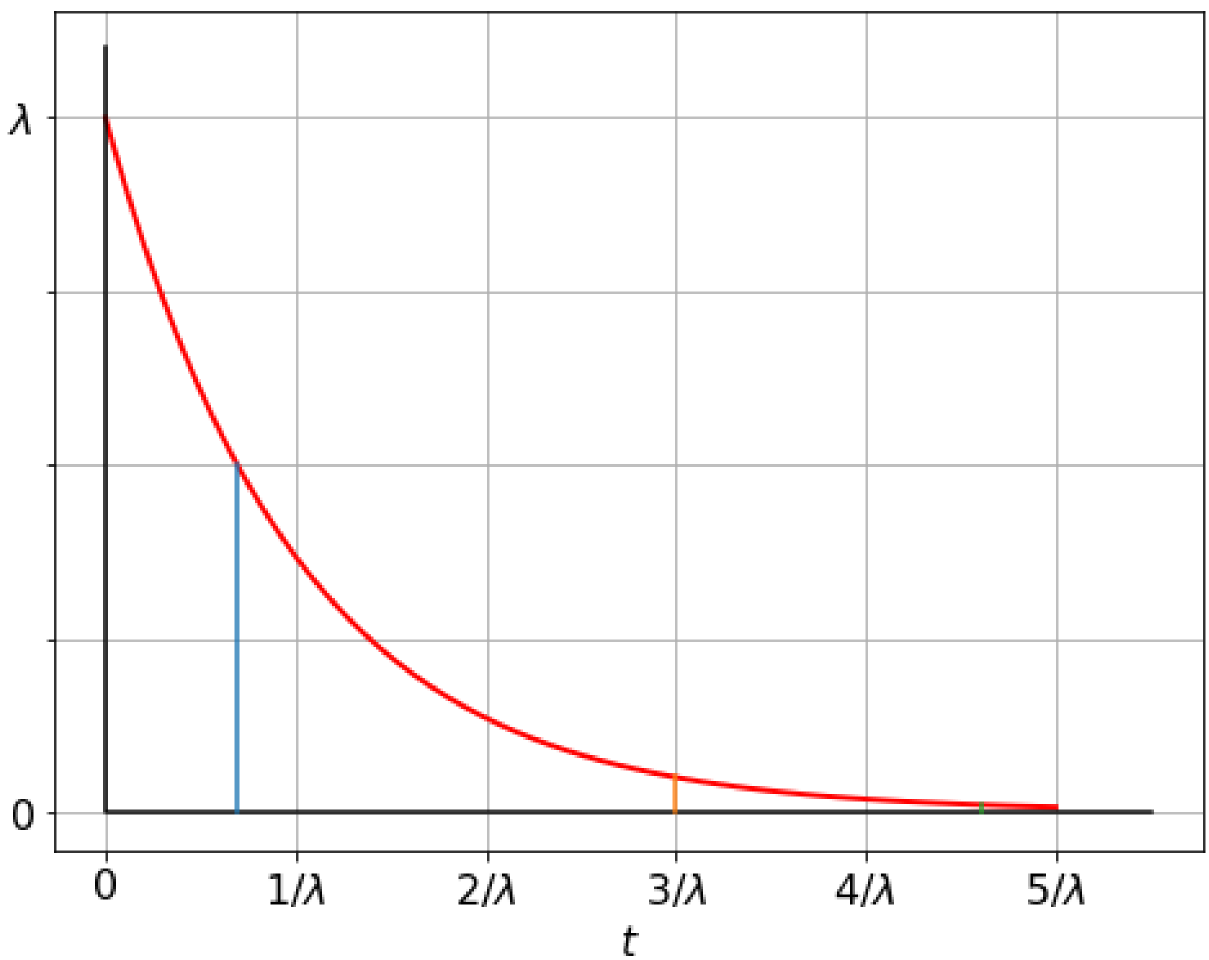


Values of X are in the interval $0 \dots b$ with probability

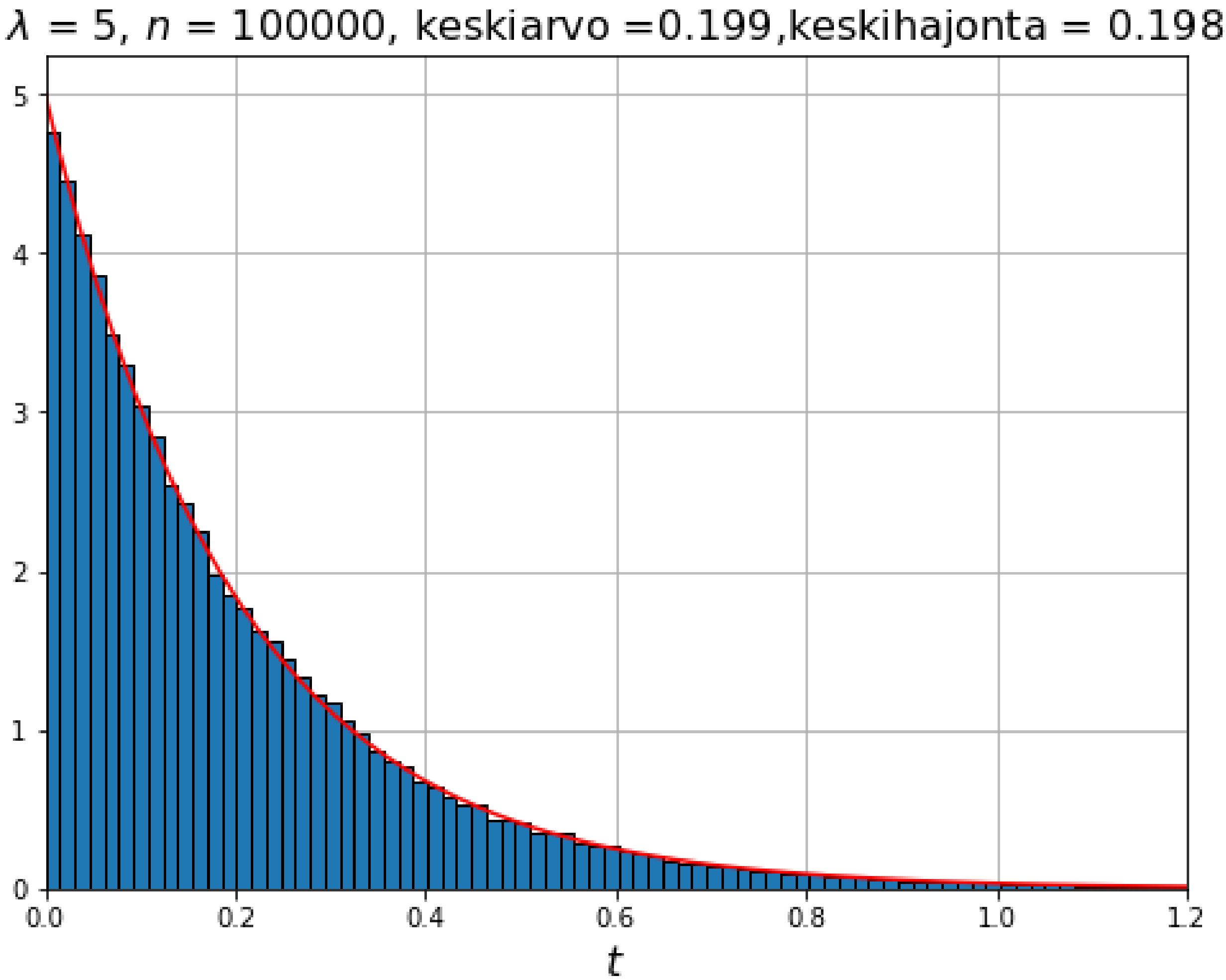
$$1 - e^{-\lambda b} = p \quad \leftrightarrow \quad b = -\frac{\log(1-p)}{\lambda}$$

i.e for example,

$$p = 0.5/0.95/0.99 \quad \leftrightarrow \quad b = \frac{0.69/3.00/4.61}{\lambda}$$

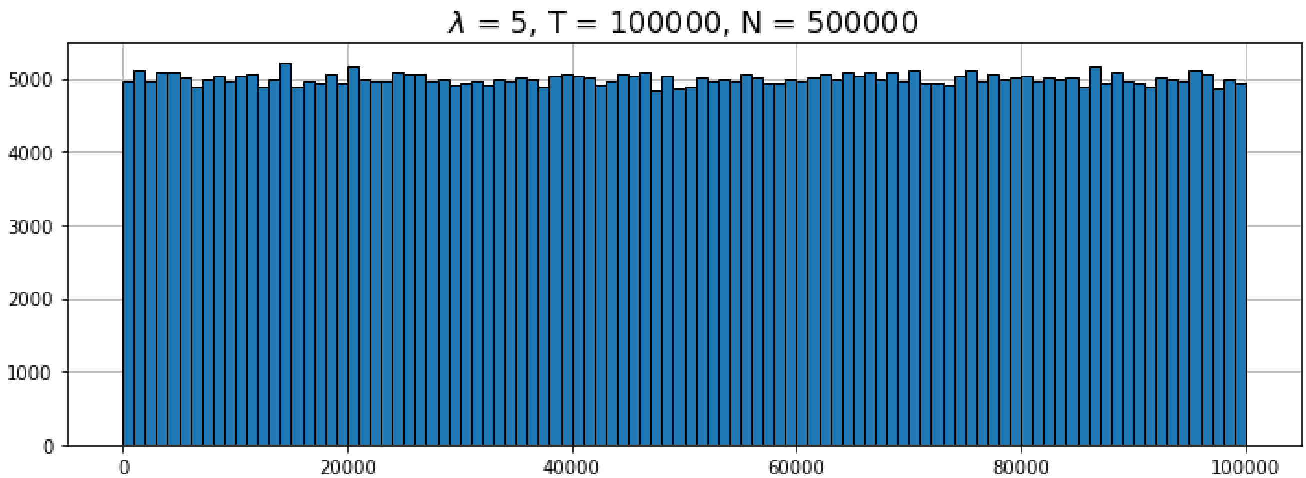


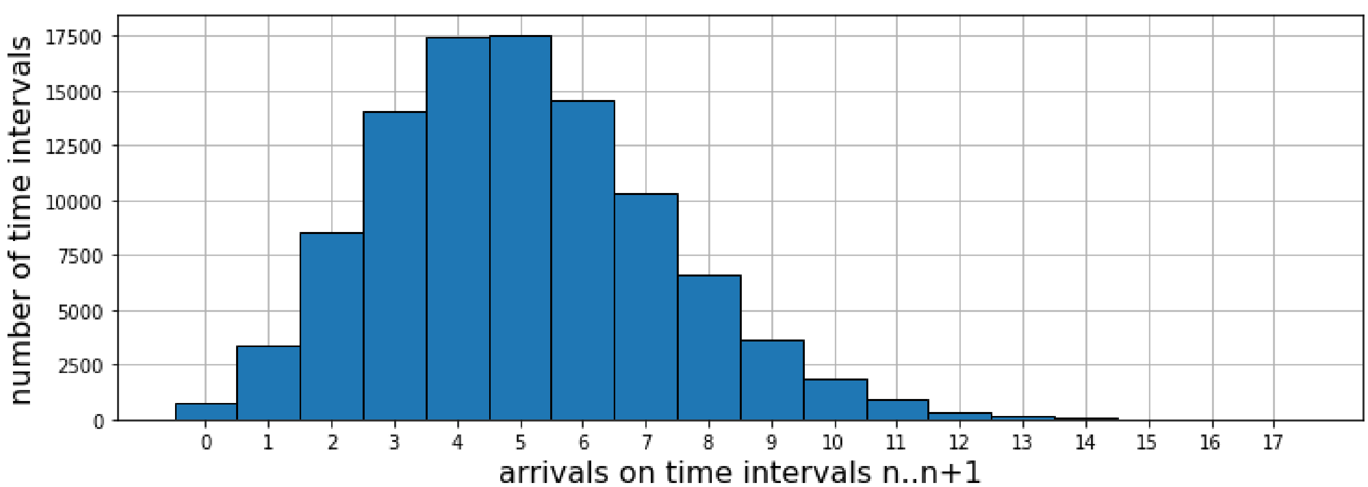
Note: If $t_1, t_2, ..., t_n$ are approximately $\text{Exp}(\lambda)$-distributed, then their mean $\approx 1/\lambda$ and standard deviation $\approx 1/\lambda$

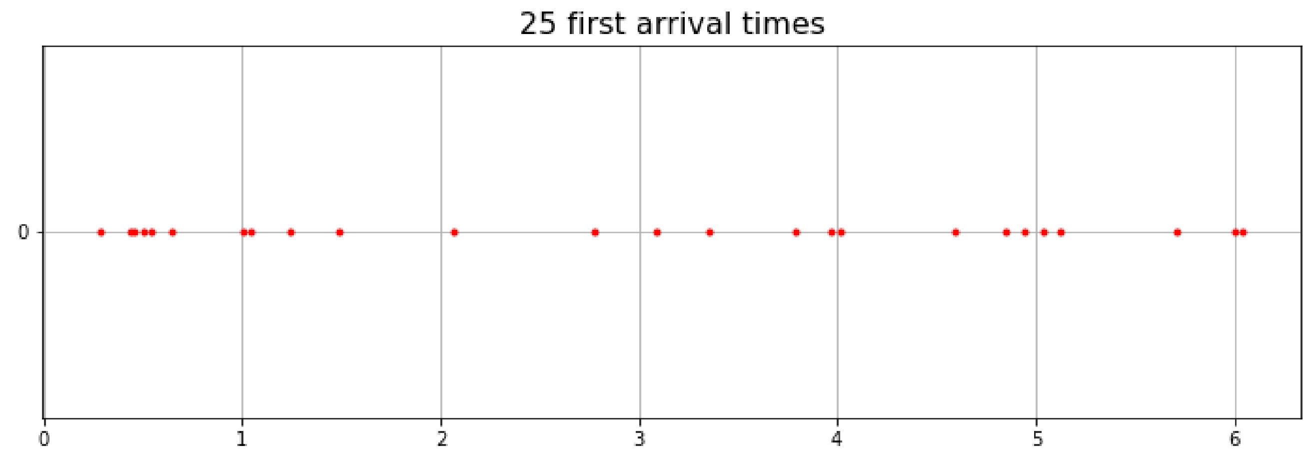


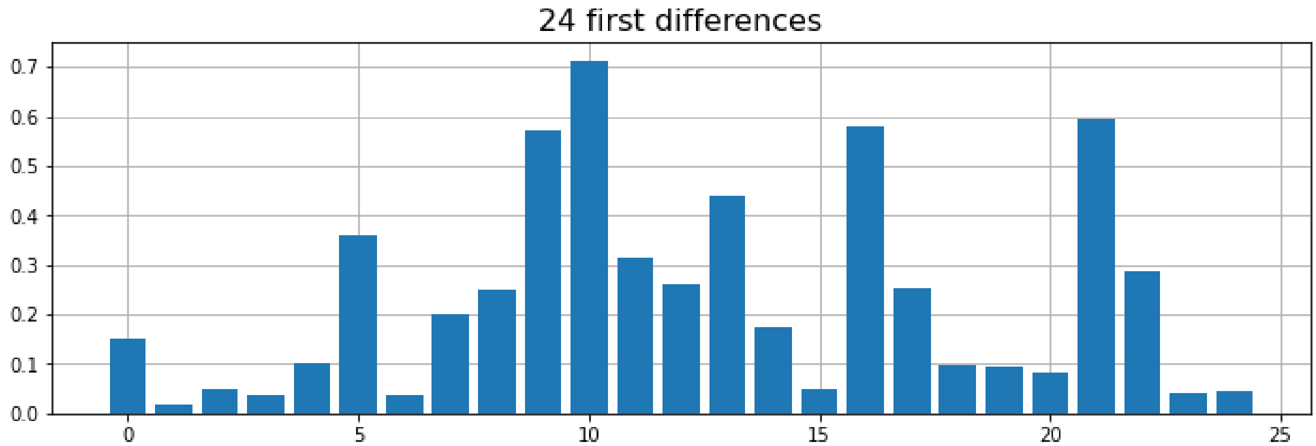
keskiarvo=mean
keskihajonta = standard deviation

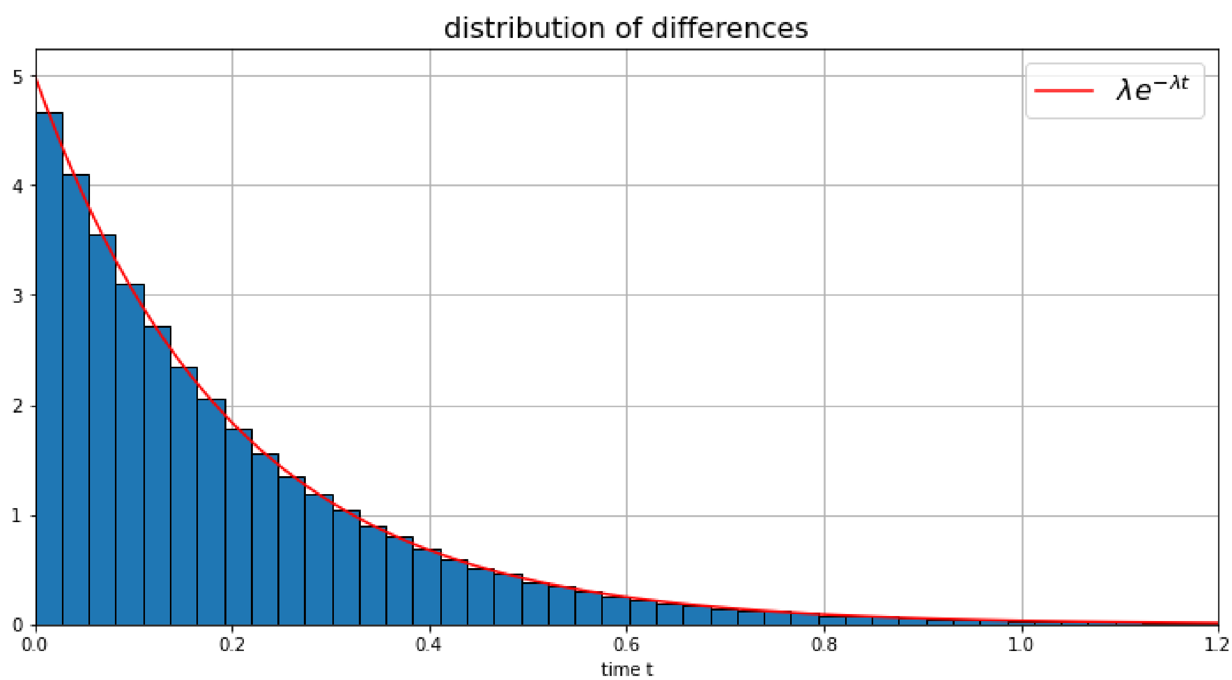
Ex: Create $N = \lambda \cdot T$ uniformly distributed arrival times between $0 \ldots T$ ($\to$ on each time interval $n \ldots n+1$ of length 1 approximately λ arrival times). Then the differences between arrival times are approximately $\mathrm{Exp}(\lambda)$-distributed.



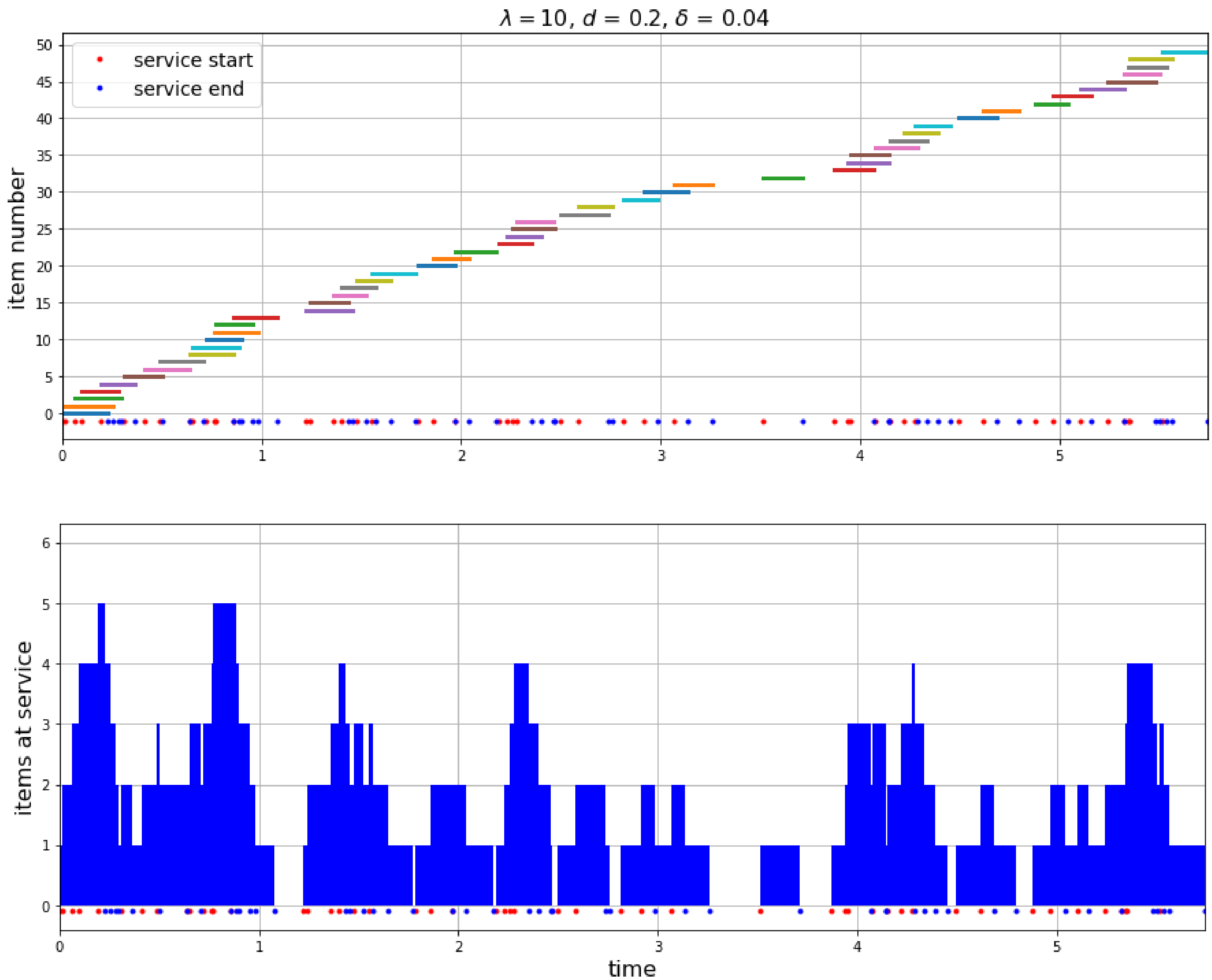


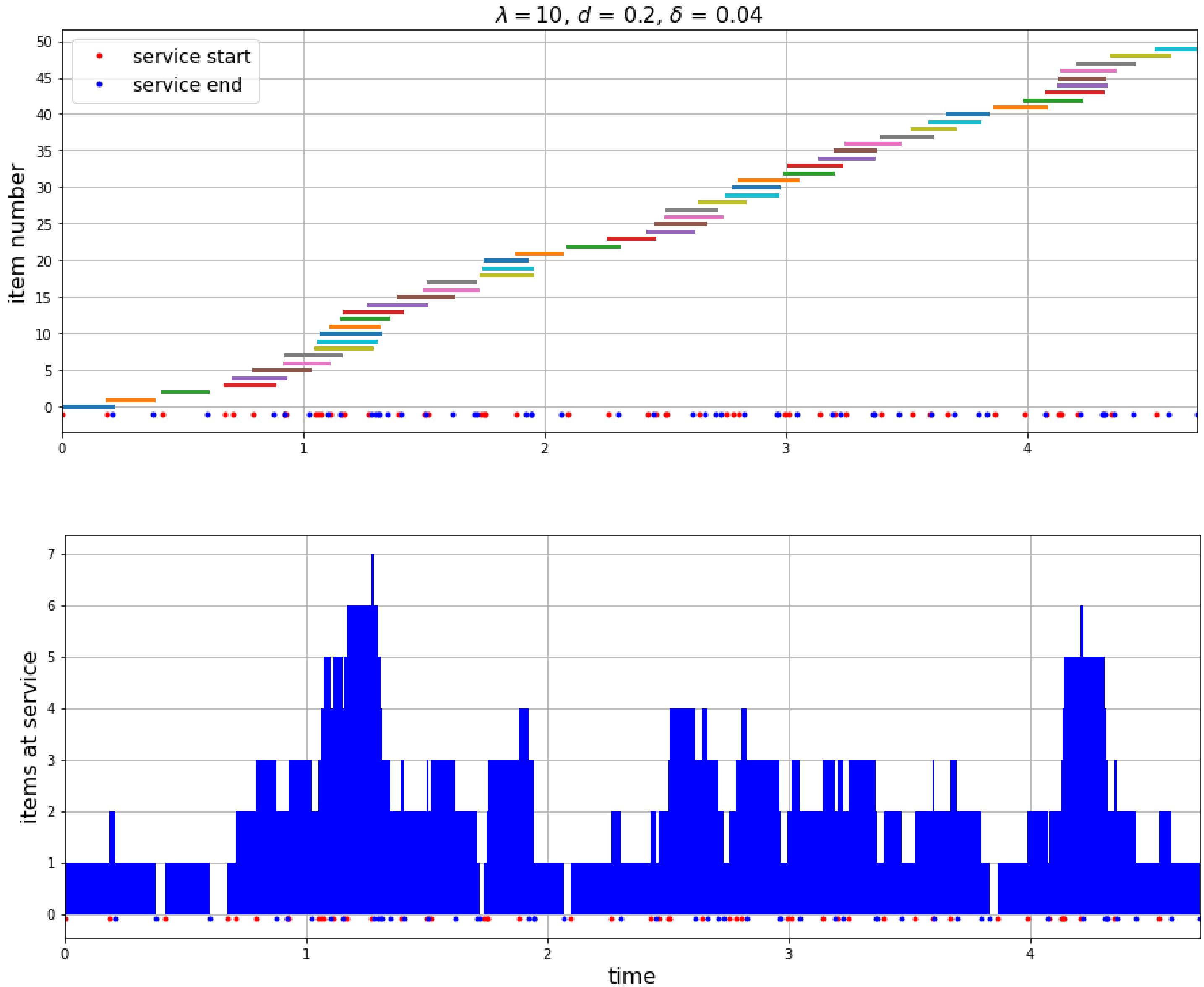

25 first arrival times
0
0
1
2
3
4
5
6


24 first differences
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25

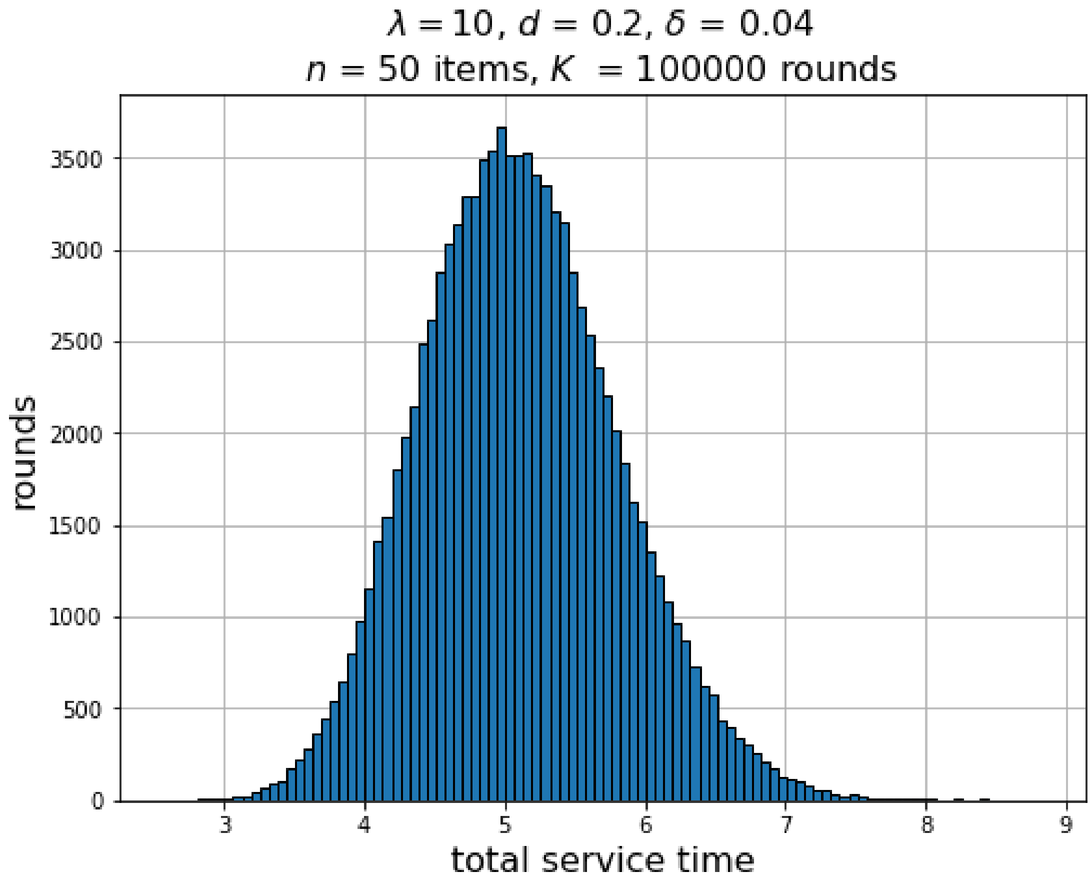

distribution of differences
$\lambda e^{-\lambda t}$
5
4
3
2
1
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
time t

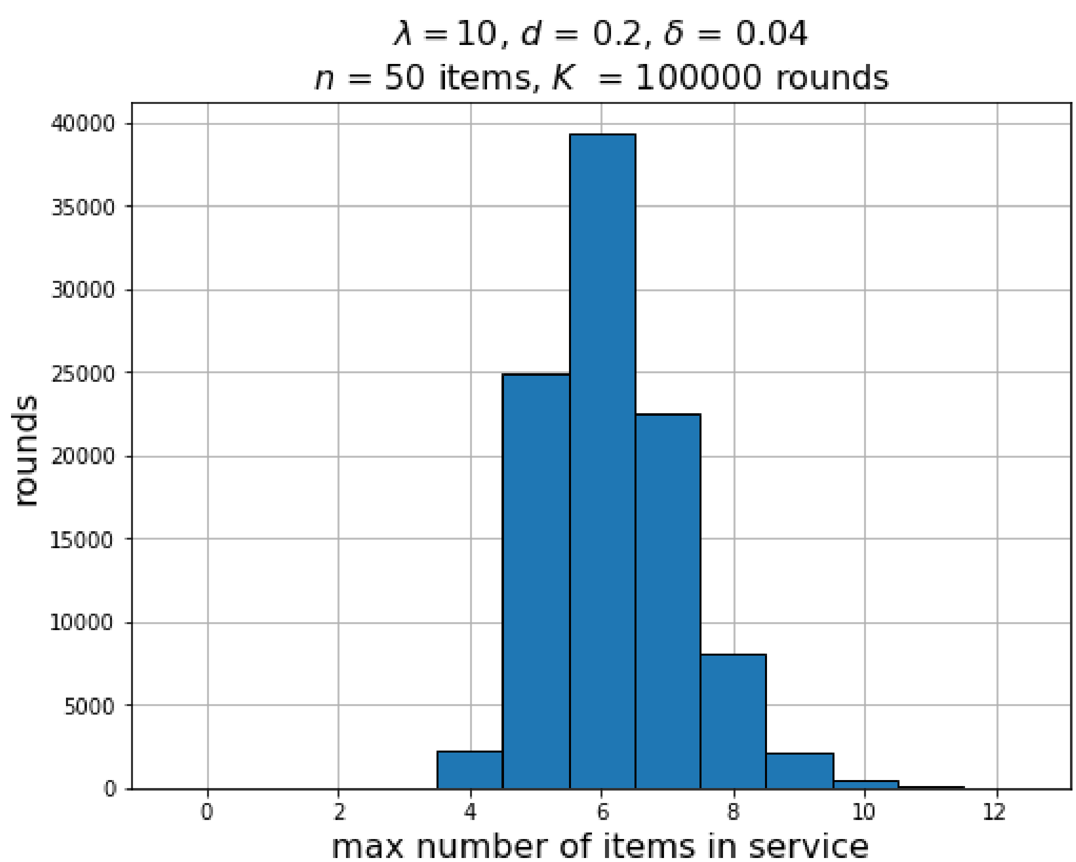
Ex: Items arrive to "service station" such that the differences of arrival times are Exp(λ)-distributed, and the service times are uniformly distributed between $d \pm \delta$.

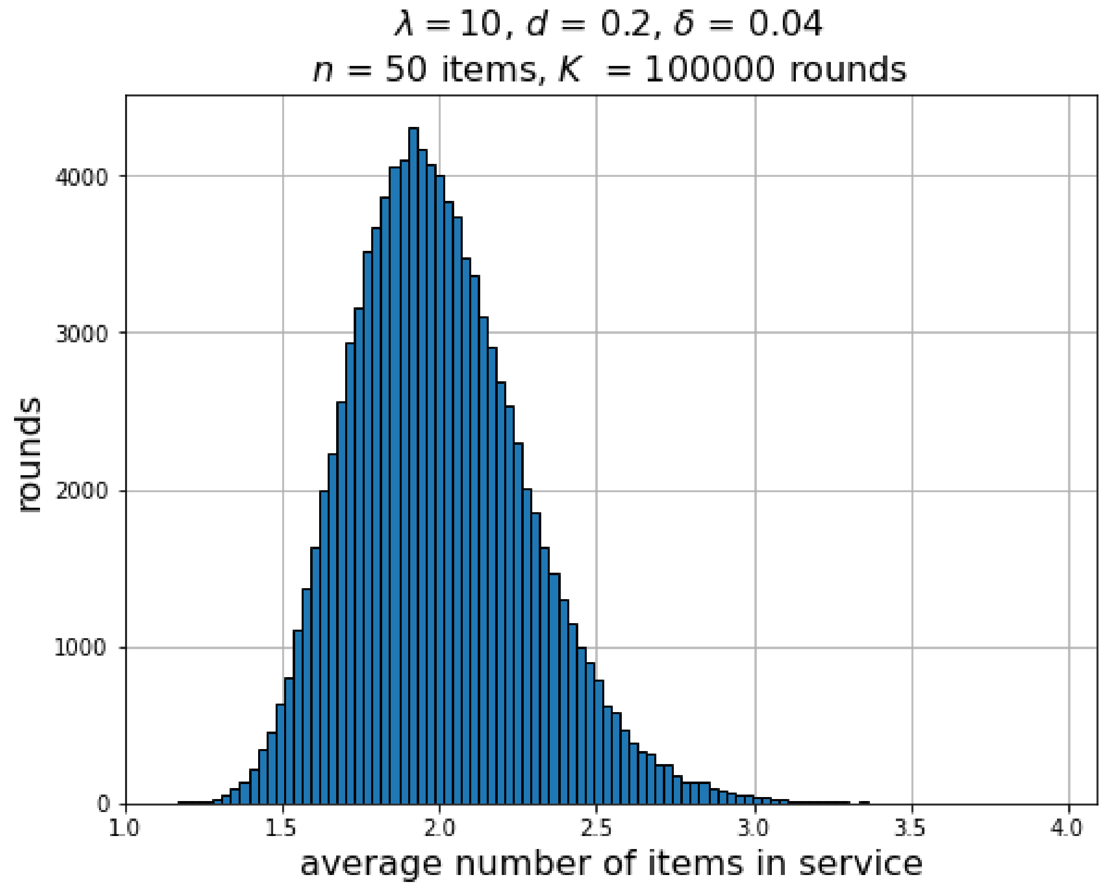




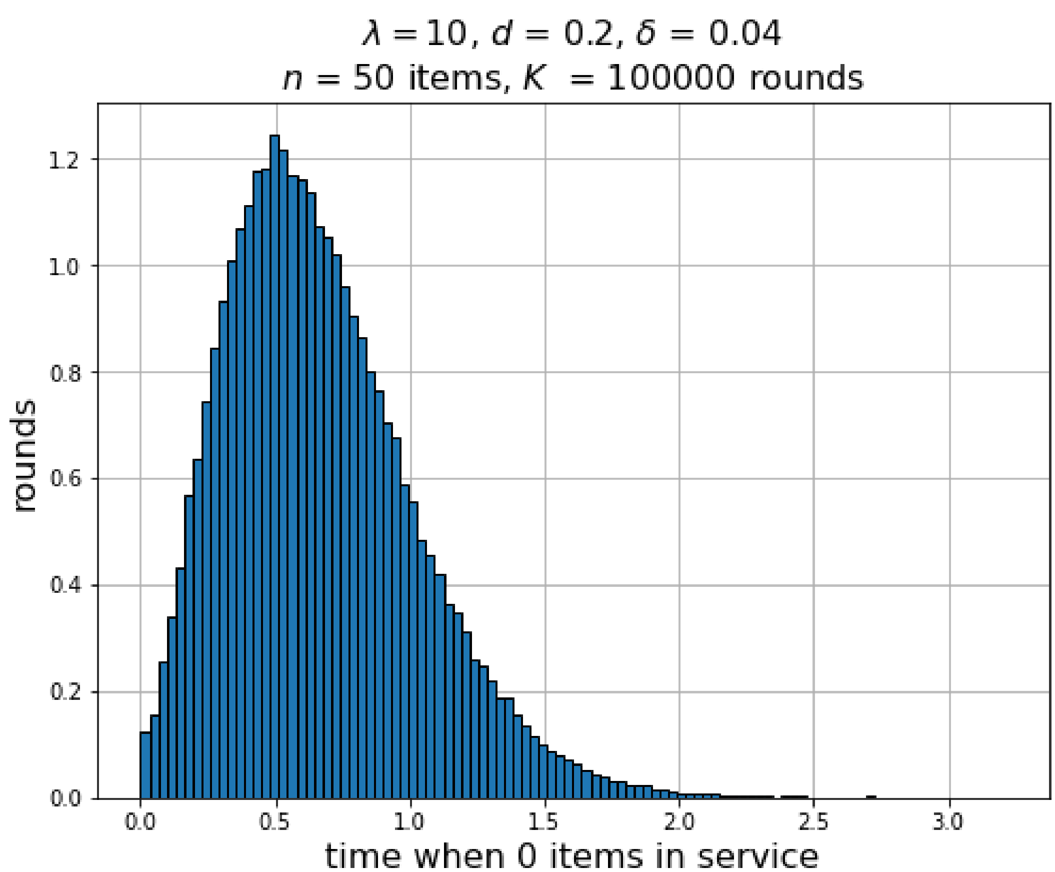
$\lambda = 10$, $d = 0.2$, $\delta = 0.04$
service start
service end
item number
items at service
time


$\lambda = 10$, d = 0.2, δ = 0.04
n = 50 items, K = 100000 rounds
rounds
total service time
0
500
1000
1500
2000
2500
3000
3500
3
4
5
6
7
8
9


$\lambda = 10$, d = 0.2, δ = 0.04
n = 50 items, K = 100000 rounds
rounds
max number of items in service
0
5000
10000
15000
20000
25000
30000
35000
40000
0
2
4
6
8
10
12



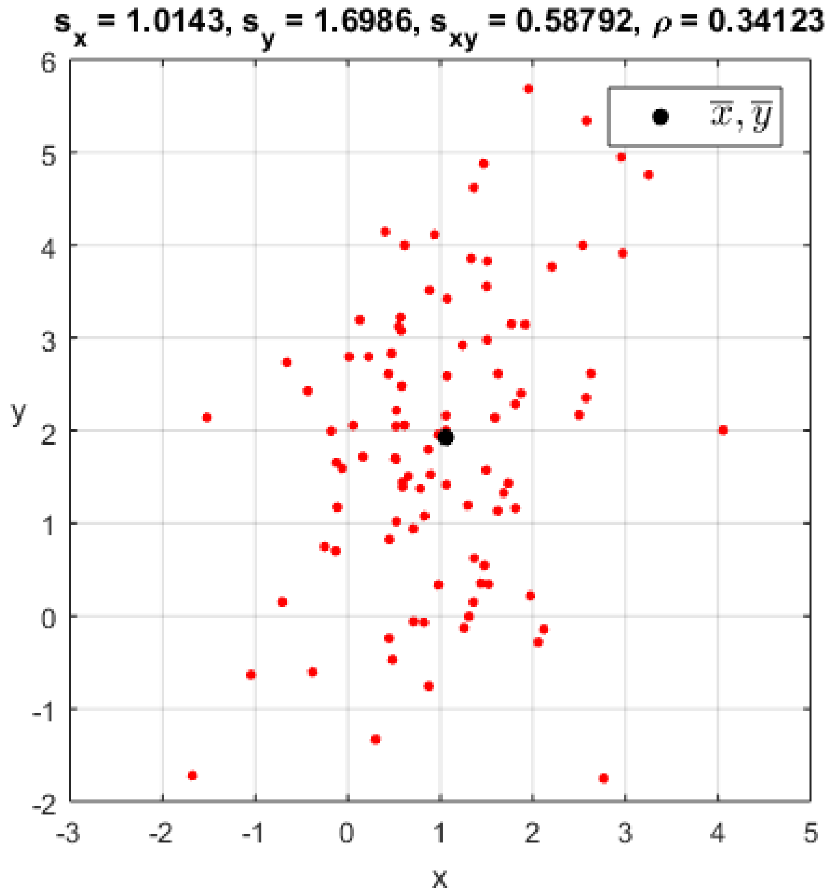
$\lambda = 10$, d = 0.2, δ = 0.04
n = 50 items, K = 100000 rounds
rounds
average number of items in service
0
1000
2000
3000
4000
1.0
1.5
2.0
2.5
3.0
3.5
4.0



$\lambda = 10$, d = 0.2, δ = 0.04
n = 50 items, K = 100000 rounds
rounds
time when 0 items in service
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

2D-data

Points $[x_1, y_1], [x_2, y_2], \ldots, [x_n, y_n]$



Means of coordinates $\overline{x}, \overline{y}$, standard deviations s_x, s_y, variances

$$s_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2, \quad s_y^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - \overline{y})^2$$

Covariance

$$s_{xy} = s_{yx} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})$$

measures the correlation between coordinates.

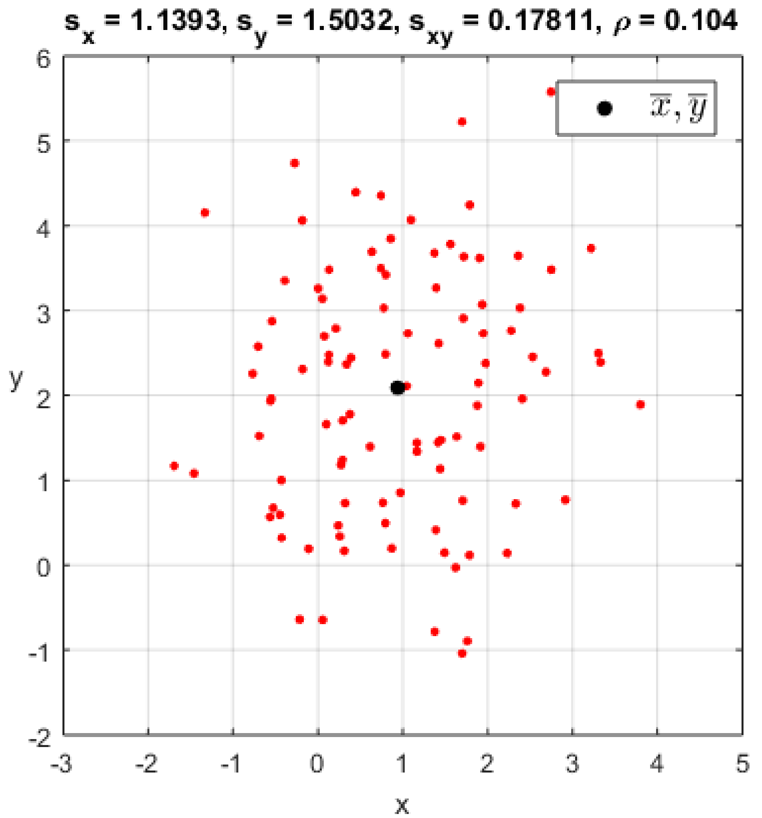
Correlation coefficient

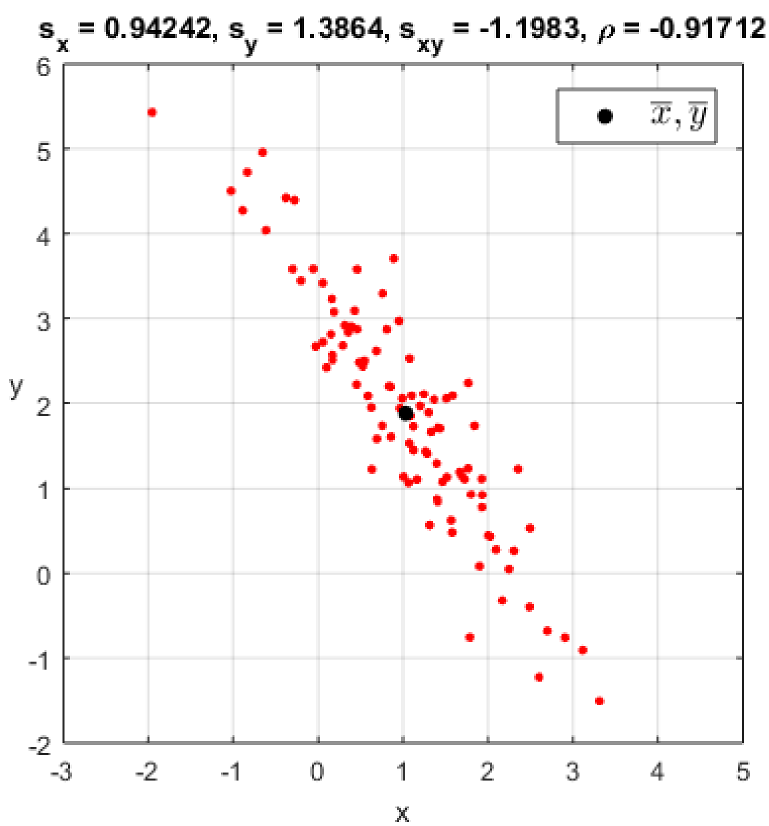
$$\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \text{ is between } -1 \ldots 1.$$

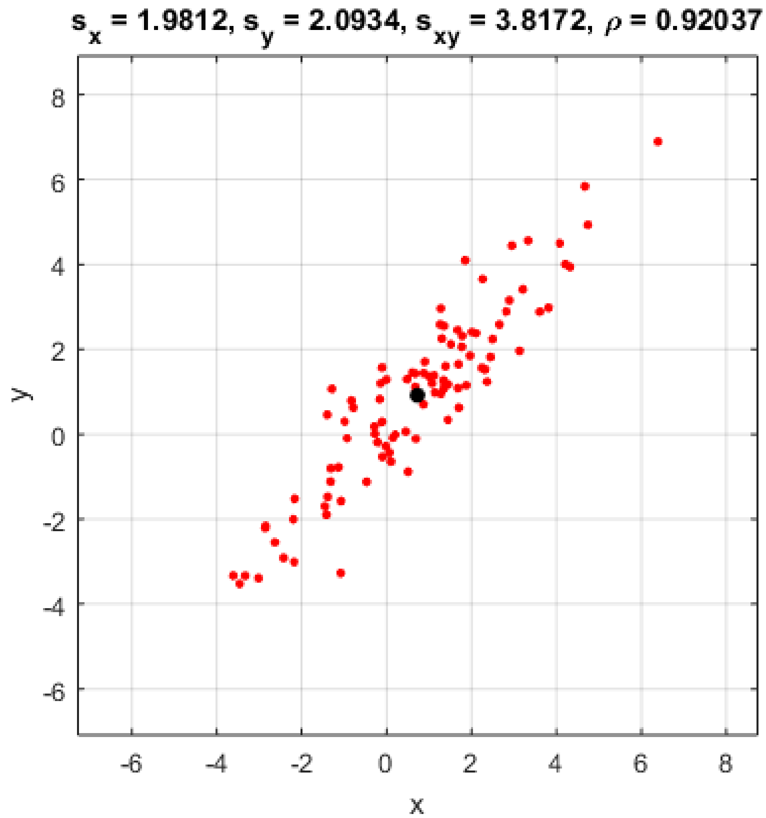
If $\rho \approx \pm 1$, then the points are approximately on a line

Covariance matrix

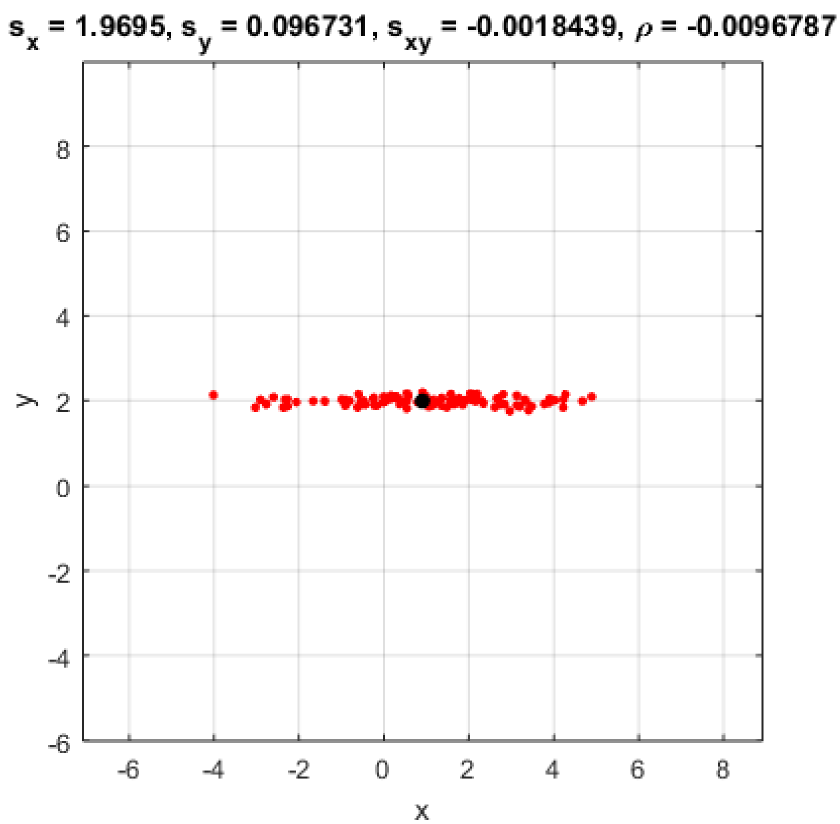
$$S = \begin{bmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{bmatrix}$$


s_x = 1.1393, s_y = 1.5032, s_xy = 0.17811, ρ = 0.104
x̄, ȳ
y
x
6
5
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
4
5


s_x = 0.94242, s_y = 1.3864, s_xy = -1.1983, ρ = -0.91712
x̄, ȳ
y
x
6
5
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
4
5



s_x = 1.9812, s_y = 2.0934, s_{xy} = 3.8172, ρ = 0.92037
y
x



s_x = 1.9695, s_y = 0.096731, s_{xy} = -0.0018439, ρ = -0.0096787
y
x

2D normal distribution

(Bivariate Gaussian)

Values of $X = \begin{bmatrix} x \\ y \end{bmatrix}$ (for example, x=position, y=velocity)

are normally distributed with mean $\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and

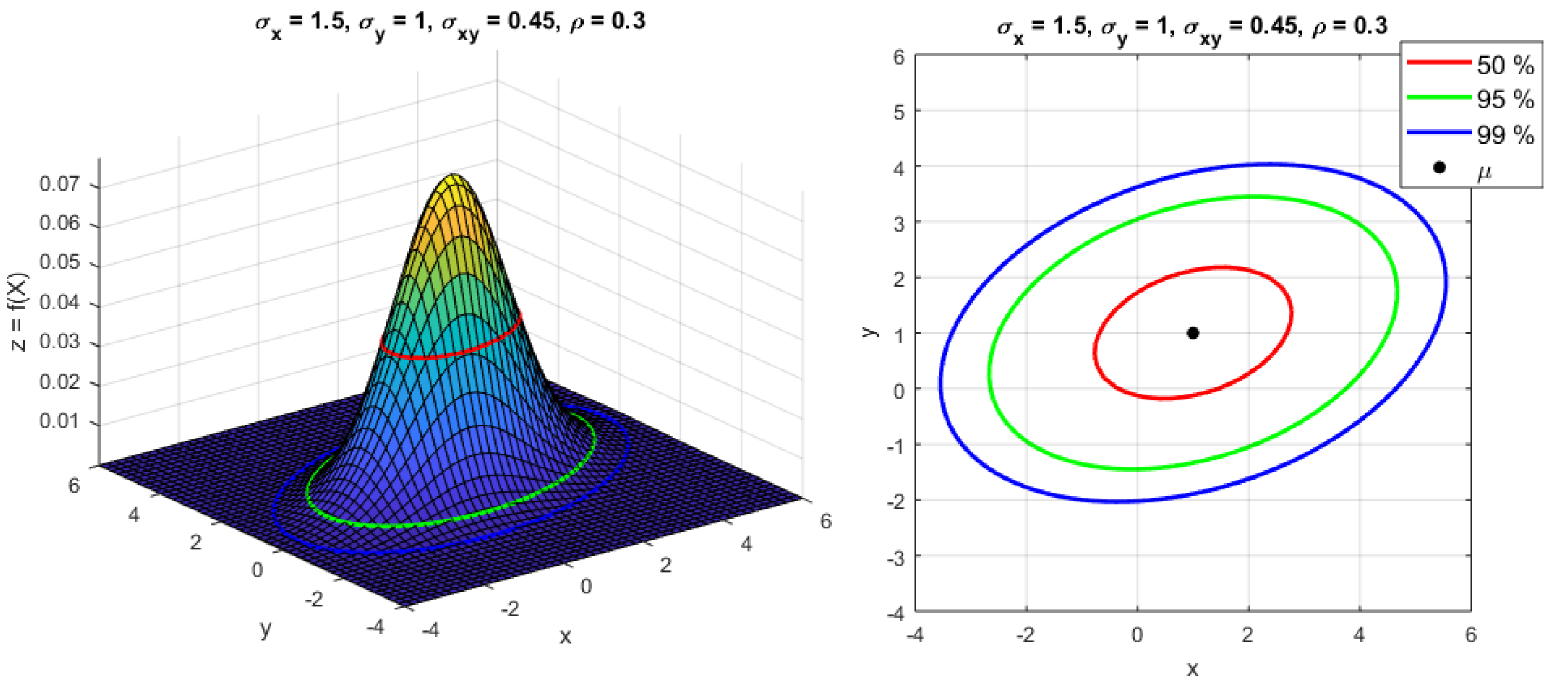
covariance matrix $\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$, $X \sim N(\boldsymbol{\mu}, \Sigma)$,

if the probability density function is

$$f(X) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(X-\boldsymbol{\mu})^T \Sigma^{-1} (X-\boldsymbol{\mu})\right)$$

μ_x, μ_y are the means and σ_x^2, σ_y^2 variances of the coordinates, $\sigma_{xy} = \rho\sigma_x\sigma_y$ their covariance, ρ is the correlation coefficient, $|\Sigma|$ is the determinant and Σ^{-1} the inverse of Σ.

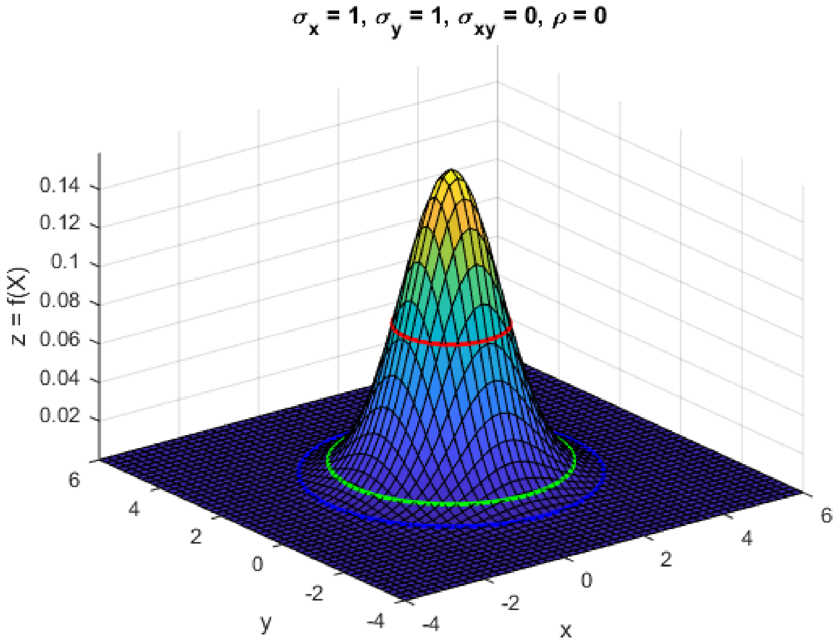
$$X - \boldsymbol{\mu} = \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}, \ (X - \boldsymbol{\mu})^T = [x - \mu_x, y - \mu_y]$$

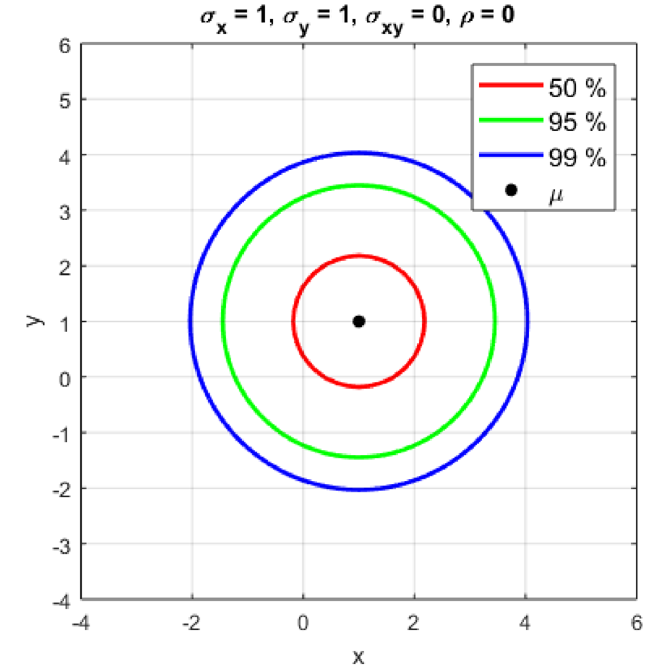


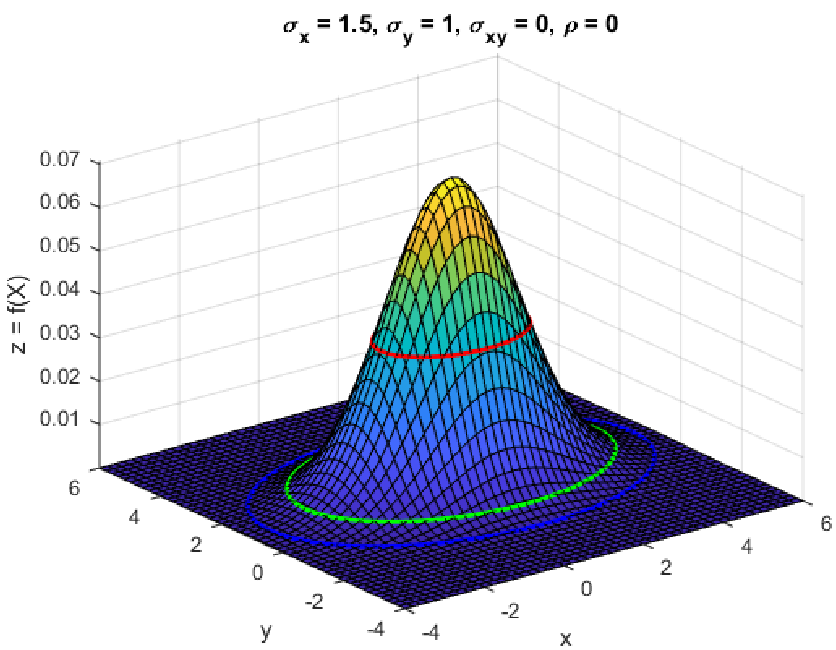
I.e, the probability that the value of X is on a certain region of the plane is the volume between the region and the surface $z = f(X)$.

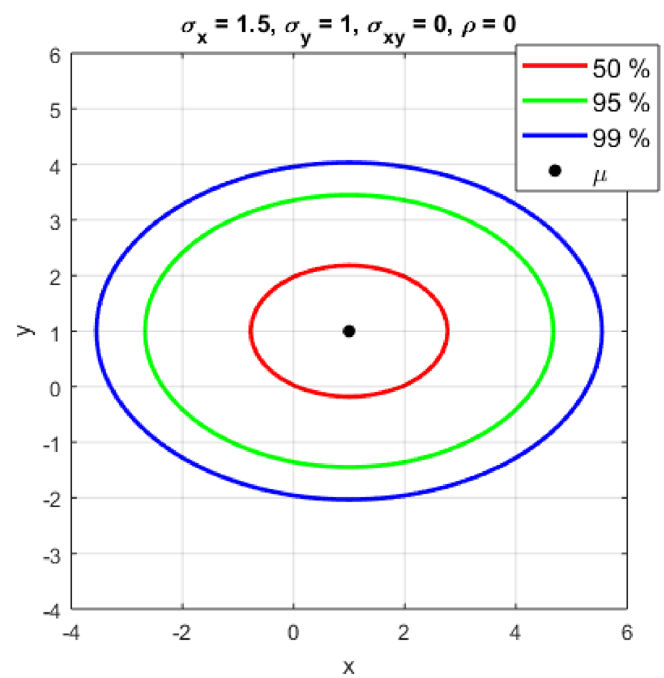
In particular 50/95/99 % of the values of X are inside the ellipse

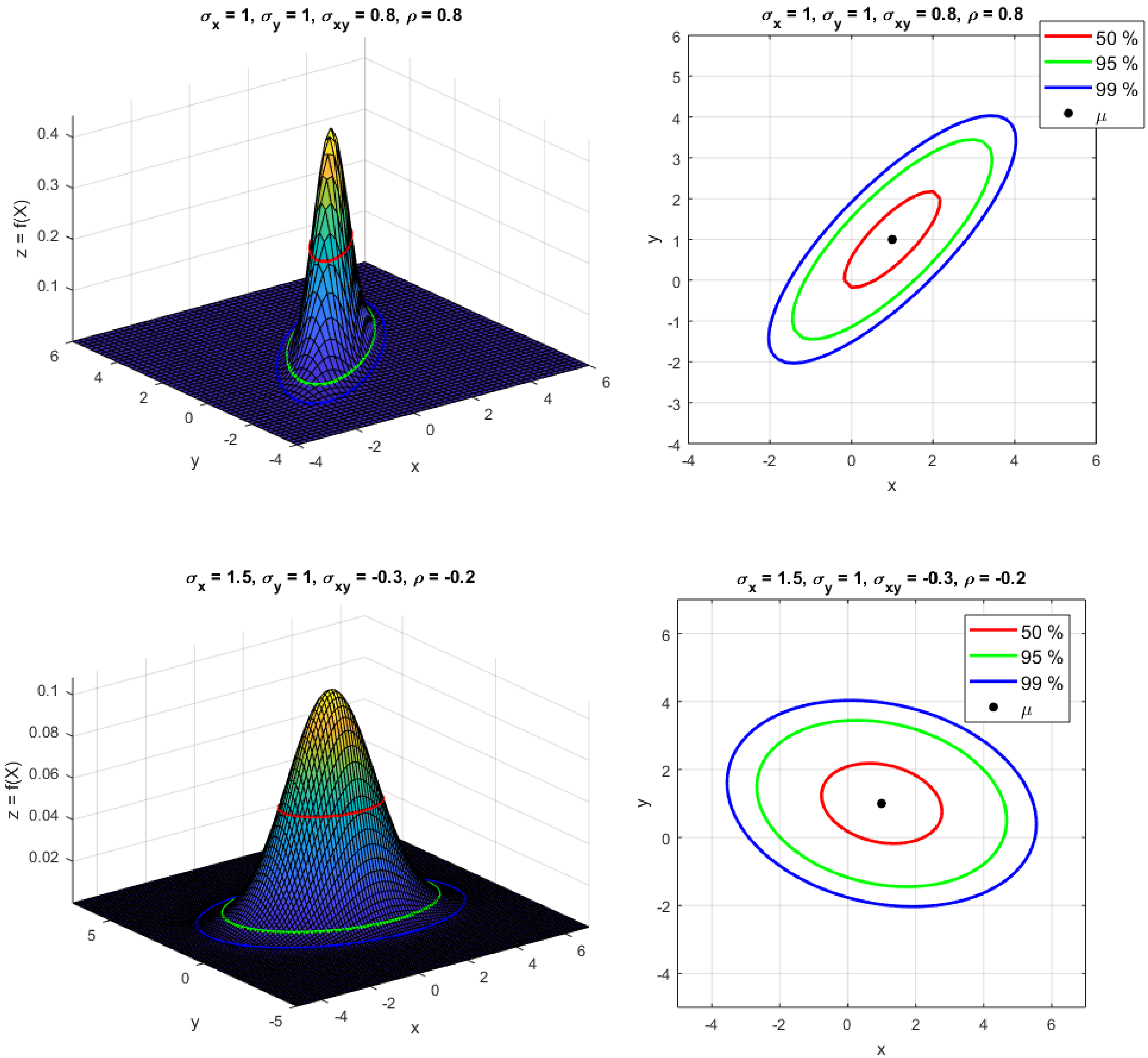
$$(X - \boldsymbol{\mu})^T \Sigma^{-1} (X - \boldsymbol{\mu}) = 1.4/6.0/9.2$$


σx = 1, σy = 1, σxy = 0, ρ = 0
z = f(X)
0.14
0.12
0.1
0.08
0.06
0.04
0.02
6
4
2
0
-2
-4
y
-4
-2
0
2
4
6
x


σx = 1, σy = 1, σxy = 0, ρ = 0
50 %
95 %
99 %
μ
y
x

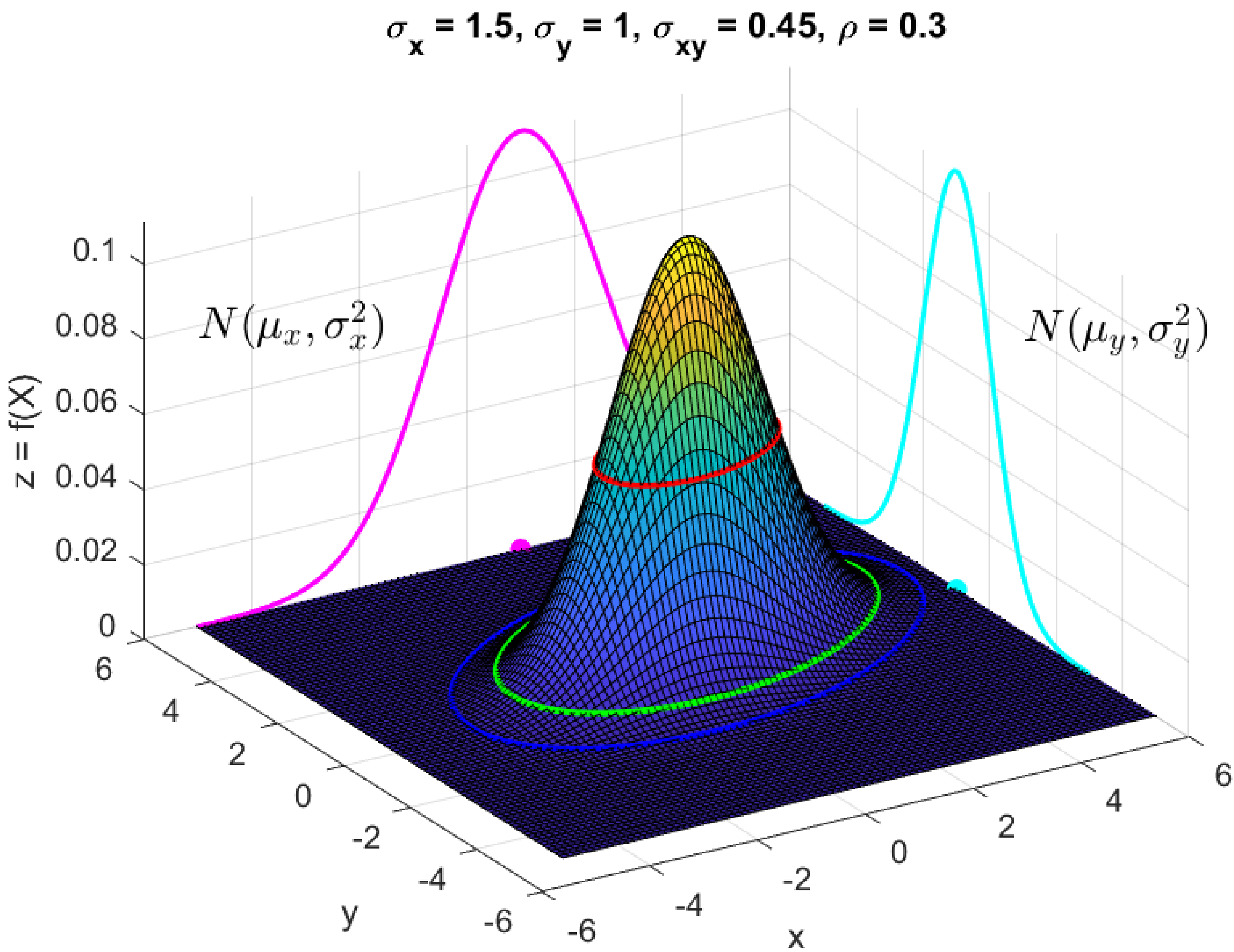

σx = 1.5, σy = 1, σxy = 0, ρ = 0
z = f(X)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
y
x


σx = 1.5, σy = 1, σxy = 0, ρ = 0
50 %
95 %
99 %
μ
y
x


σx = 1, σy = 1, σxy = 0.8, ρ = 0.8
z = f(X)
x
y
σx = 1, σy = 1, σxy = 0.8, ρ = 0.8
50 %
95 %
99 %
μ
x
y
σx = 1.5, σy = 1, σxy = -0.3, ρ = -0.2
z = f(X)
x
y
σx = 1.5, σy = 1, σxy = -0.3, ρ = -0.2
50 %
95 %
99 %
μ
x
y

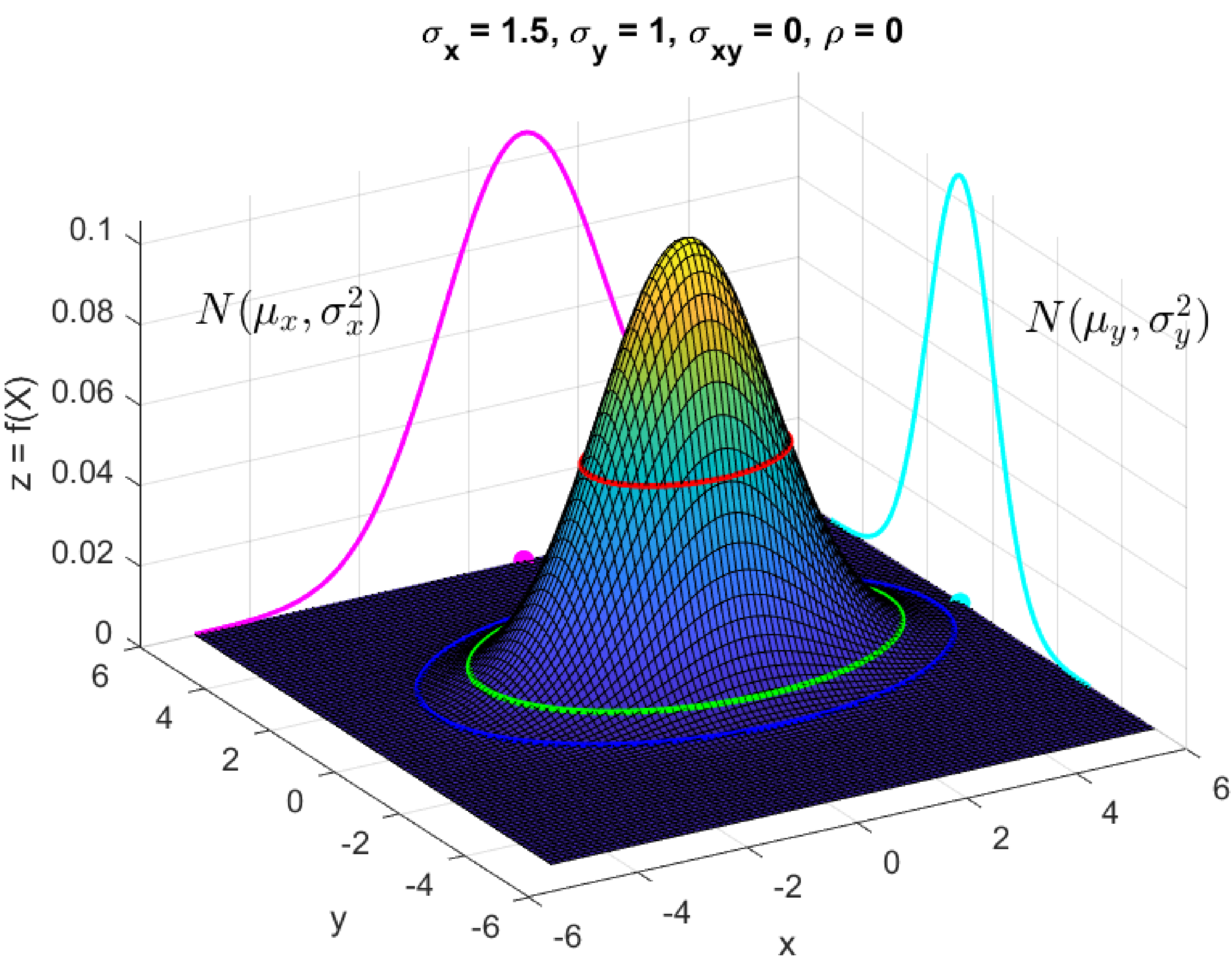
Note: If $X \sim N(\boldsymbol{\mu}, \Sigma)$ is 2D-normally distributed, then the coordinates are 1D-normally distributed

$$x \sim N(\mu_x, \sigma_x^2) \text{ and } y \sim N(\mu_y, \sigma_y^2)$$



If coordinates $x \sim N(\mu_x, \sigma_x^2)$ and $y \sim N(\mu_y, \sigma_y^2)$ are 1D-normally distributed and their covariance $\sigma_{xy} = 0$ (i.e their correlation $= 0$), then $X \sim N(\boldsymbol{\mu}, \Sigma)$, where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{ja} \quad \Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$



If points $X_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$, $k = 1, 2, ..., n$, ovat are approximately $N(\boldsymbol{\mu}, \Sigma)$-distributed,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

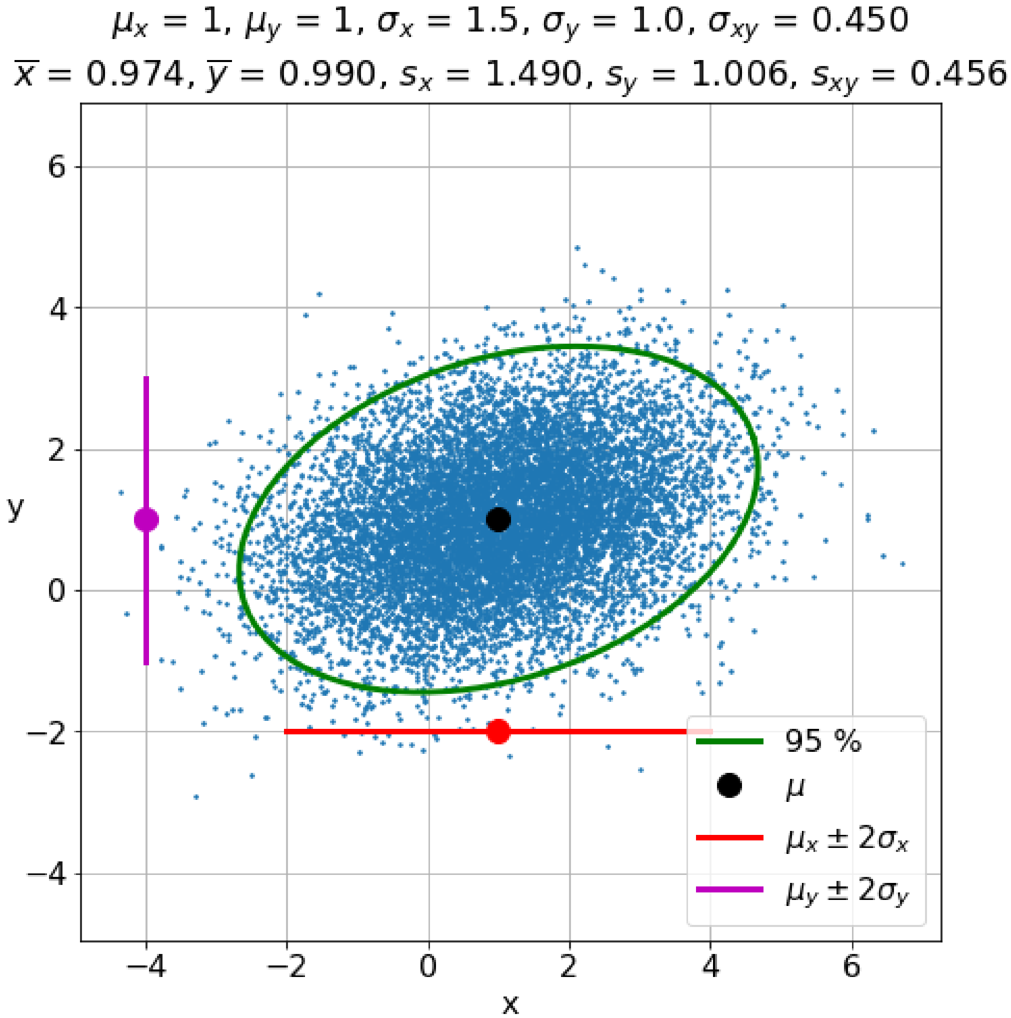
then the coordinates x_k and y_k are approximately $N(\mu_x, \sigma_x^2)$- and $N(\mu_y, \sigma_y^2)$-distributed and

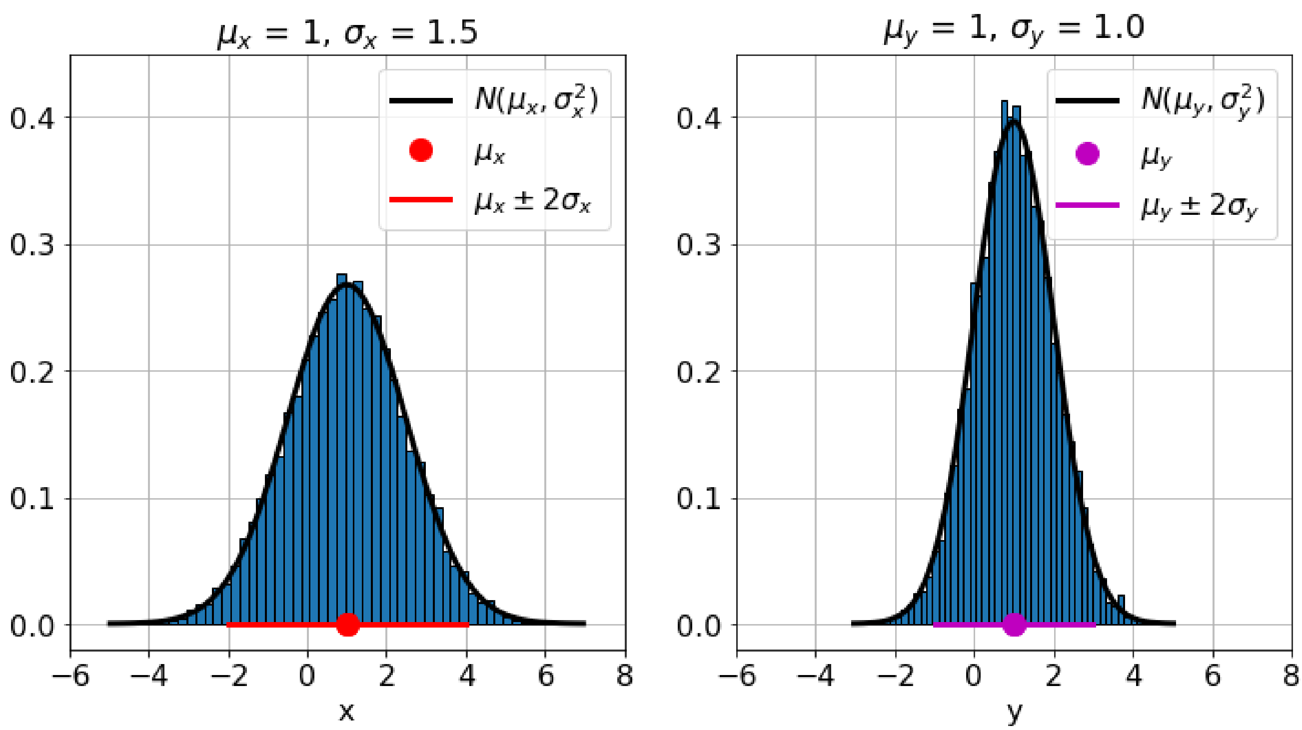
$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \approx \mu_x, \quad \overline{y} = \frac{1}{n}\sum_{k=1}^{n} y_k \approx \mu_y$$

$$s_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2 \approx \sigma_x^2$$

$$s_y^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - \overline{y})^2 \approx \sigma_y^2$$

$$s_{xy} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y}) \approx \sigma_{xy}$$


$\mu_x = 1$, $\mu_y = 1$, $\sigma_x = 1.5$, $\sigma_y = 1.0$, $\sigma_{xy} = 0.450$
$\overline{x} = 0.974$, $\overline{y} = 0.990$, $s_x = 1.490$, $s_y = 1.006$, $s_{xy} = 0.456$
95 %
μ
$\mu_x \pm 2\sigma_x$
$\mu_y \pm 2\sigma_y$
x
y


$\mu_x = 1$, $\sigma_x = 1.5$
$N(\mu_x, \sigma_x^2)$
μ_x
$\mu_x \pm 2\sigma_x$
x
$\mu_y = 1$, $\sigma_y = 1.0$
$N(\mu_y, \sigma_y^2)$
μ_y
$\mu_y \pm 2\sigma_y$
y

If coordinates x_k and y_k are approximately $N(\mu_x, \sigma_x^2)$- and $N(\mu_y, \sigma_y^2)$-distributed and their covariance is ≈ 0, then

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \approx \mu_x, \quad \overline{y} = \frac{1}{n}\sum_{k=1}^{n} y_k \approx \mu_y$$

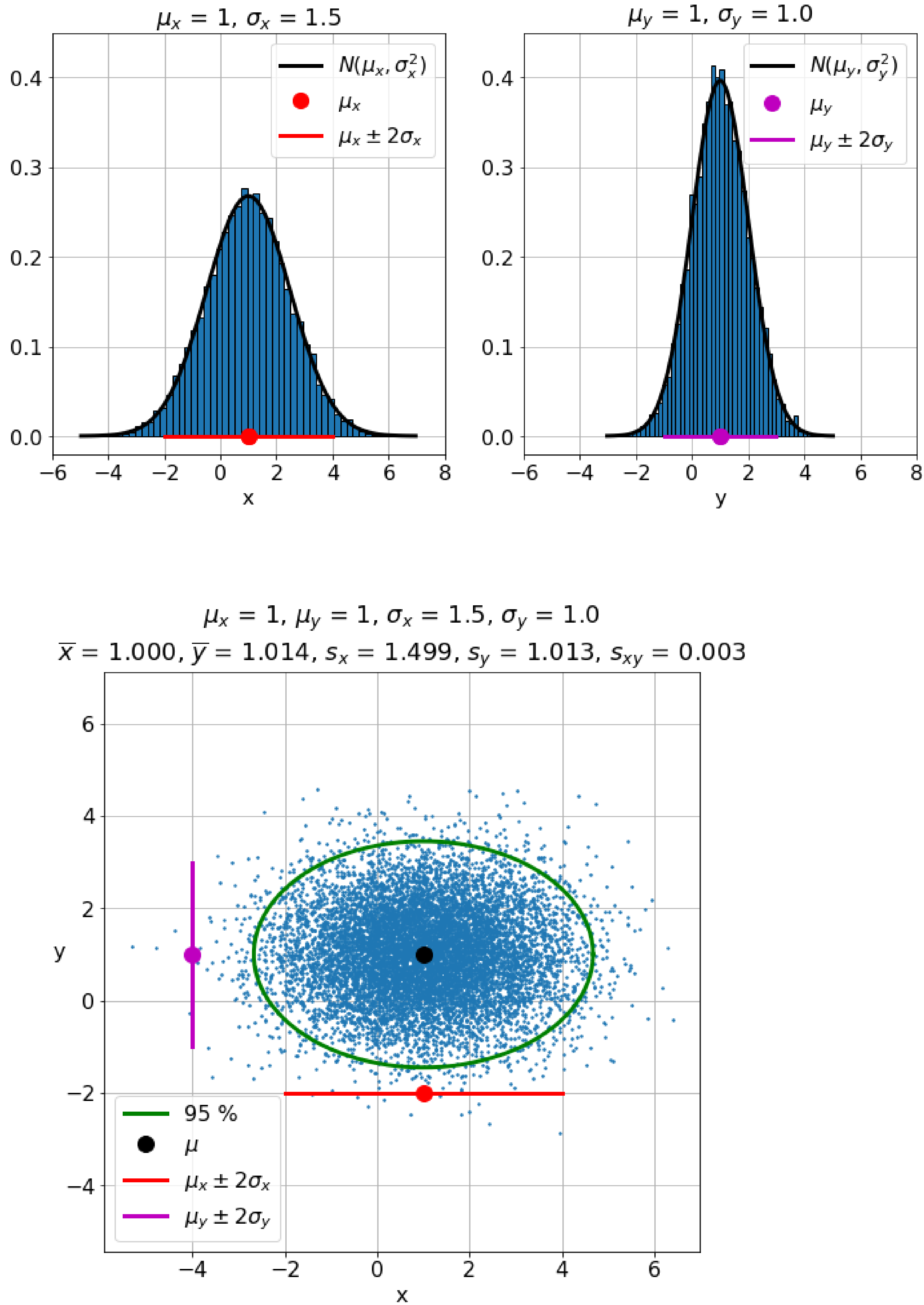
$$s_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2 \approx \sigma_x^2$$

$$s_y^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - \overline{y})^2 \approx \sigma_y^2$$

$$s_{xy} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y}) \approx 0$$

and points $X_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$ are approximately $N(\boldsymbol{\mu}, \Sigma)$ -distributed

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$


μ_x = 1, σ_x = 1.5
$N(\mu_x, \sigma_x^2)$
μ_x
$\mu_x \pm 2\sigma_x$
x
μ_y = 1, σ_y = 1.0
$N(\mu_y, \sigma_y^2)$
μ_y
$\mu_y \pm 2\sigma_y$
y
μ_x = 1, μ_y = 1, σ_x = 1.5, σ_y = 1.0
$\overline{x}$ = 1.000, $\overline{y}$ = 1.014, s_x = 1.499, s_y = 1.013, s_{xy} = 0.003
95 %
μ
$\mu_x \pm 2\sigma_x$
$\mu_y \pm 2\sigma_y$
x
y

Ex. If true value is $\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$

and measurements

$$x \sim N(\mu_x, \sigma_x^2) \text{ and } y \sim N(\mu_y, \sigma_y^2)$$

are independent (i.e their correlation $= 0$), then

$$X = \begin{bmatrix} x \\ y \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma)$$

where

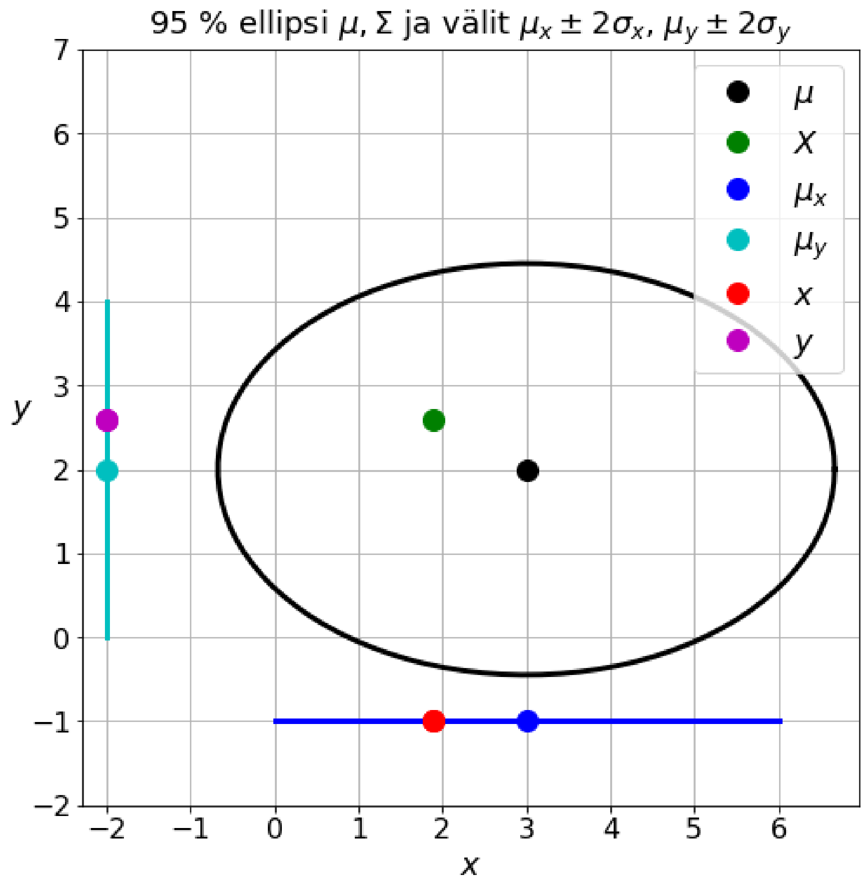
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

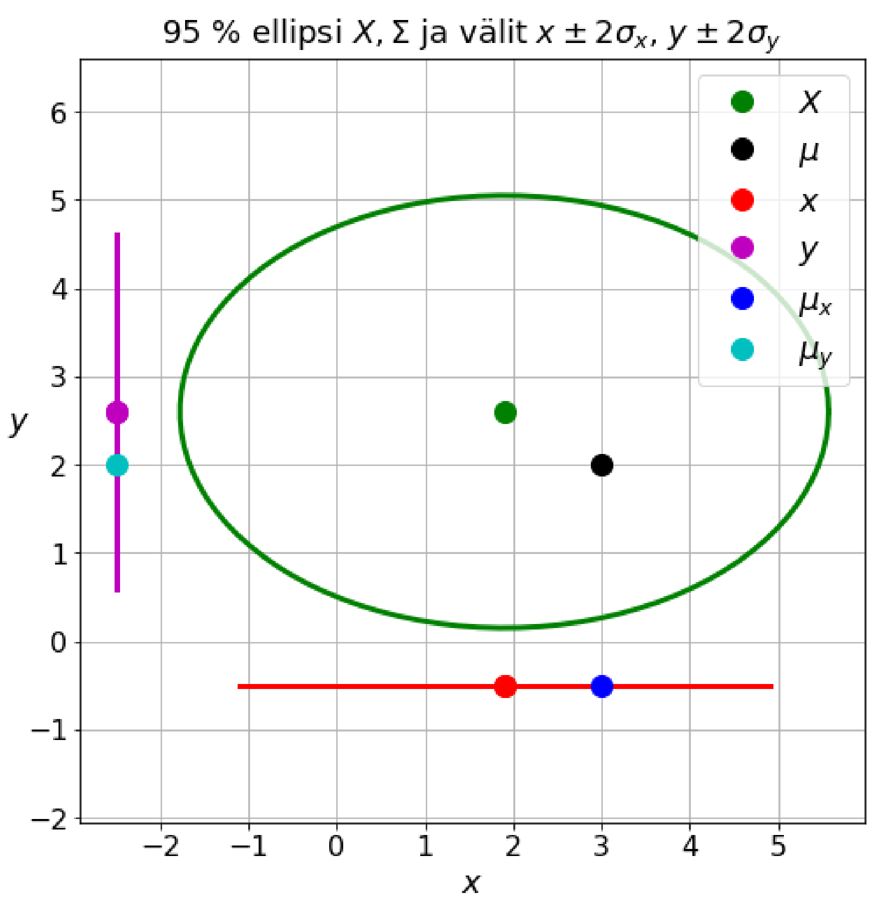
i.e

$$\mu_x \sim N(x, \sigma_x^2), \quad \mu_y \sim N(y, \sigma_y^2)$$

and

$$\boldsymbol{\mu} \sim N(X, \Sigma)$$


95 % ellipsi μ, Σ ja välit $\mu_x \pm 2\sigma_x$, $\mu_y \pm 2\sigma_y$
μ
X
μ_x
μ_y
x
y
x
y


95 % ellipsi X, Σ ja välit $x \pm 2\sigma_x$, $y \pm 2\sigma_y$
X
μ
x
y
μ_x
μ_y
x
y

Esim. True value $\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$.

Measurements

$$X_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma_1), \quad \Sigma_1 = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 y_1} \\ \sigma_{x_1 y_1} & \sigma_{y_1}^2 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma_2), \quad \Sigma_2 = \begin{bmatrix} \sigma_{x_2}^2 & \sigma_{x_2 y_2} \\ \sigma_{x_2 y_2} & \sigma_{y_2}^2 \end{bmatrix}$$

give estimates $\boldsymbol{\mu} \sim N(X_1, \Sigma_1)$ and $\boldsymbol{\mu} \sim N(X_2, \Sigma_2)$

Weighted average

$$X = \begin{bmatrix} x \\ y \end{bmatrix} = A_1 X_1 + A_2 X_2, \quad A_1 + A_2 = I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is also normally distributed, $X \sim N(\boldsymbol{\mu}, \Sigma)$, where

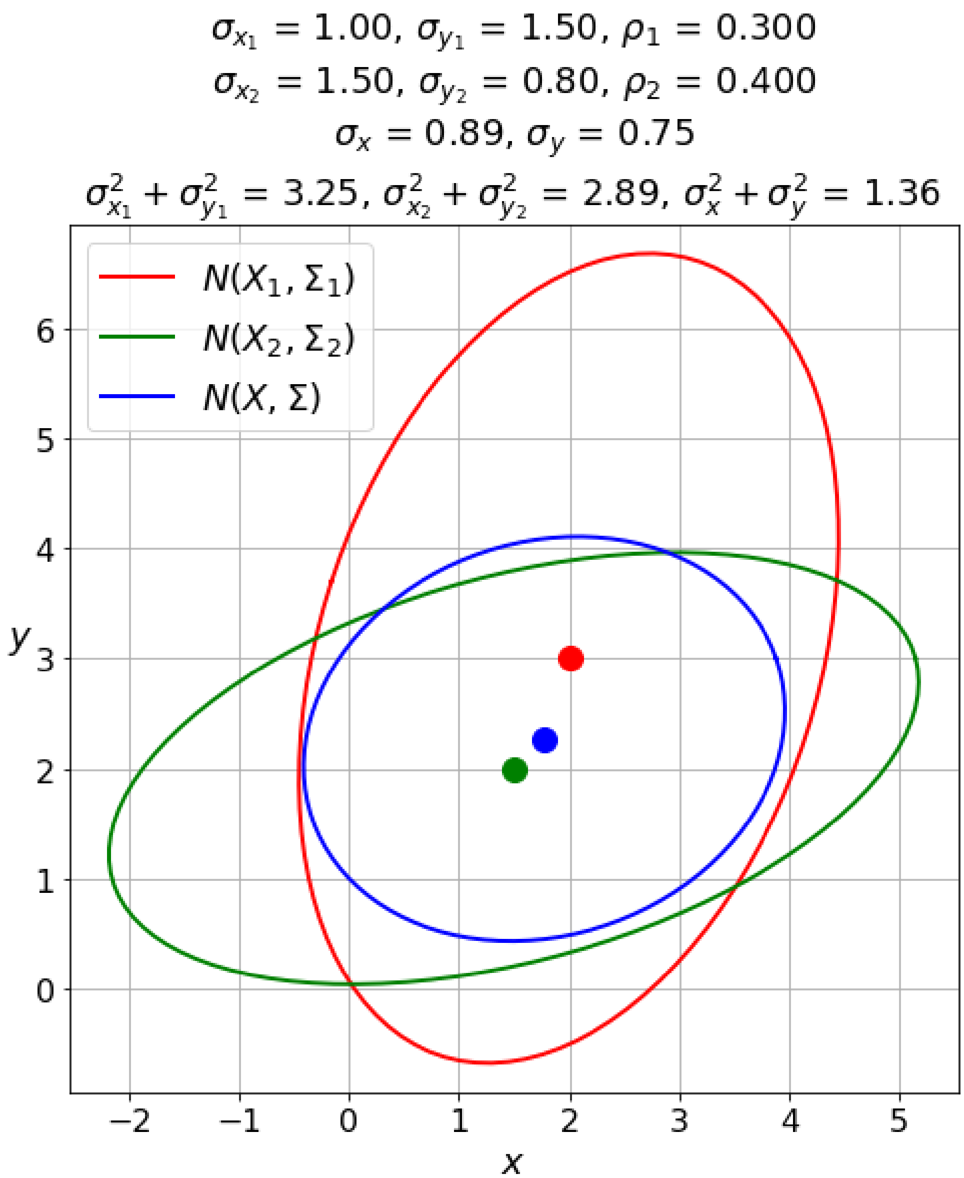
$$\Sigma = A_1 \Sigma_1 A_1^T + A_2 \Sigma_2 A_2^T = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

i.e gives an estimate $\boldsymbol{\mu} \sim N(X, \Sigma)$

Sum of variances $\sigma_x^2 + \sigma_y^2$ is smallest
(= most accurate etimate for $\boldsymbol{\mu}$ = smallest uncertainty ellipse), when

$$A_1 = \Sigma_2(\Sigma_1 + \Sigma_2)^{-1}, \quad A_2 = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$$

Then $\Sigma = \Sigma_2(\Sigma_1 + \Sigma_2)^{-1}\Sigma_1$



Note: If $K = A_2 = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$ ("Kalman gain"), then
$X = X_1 + K(X_2 - X_1)$ and
$\Sigma = (I_2 - K)\Sigma_1$

Note: If there is no correlation between the coordinates x and y i.e covariances $\sigma_{x_1 y_1} = \sigma_{x_2 y_2} = 0$, then (compare 1D)

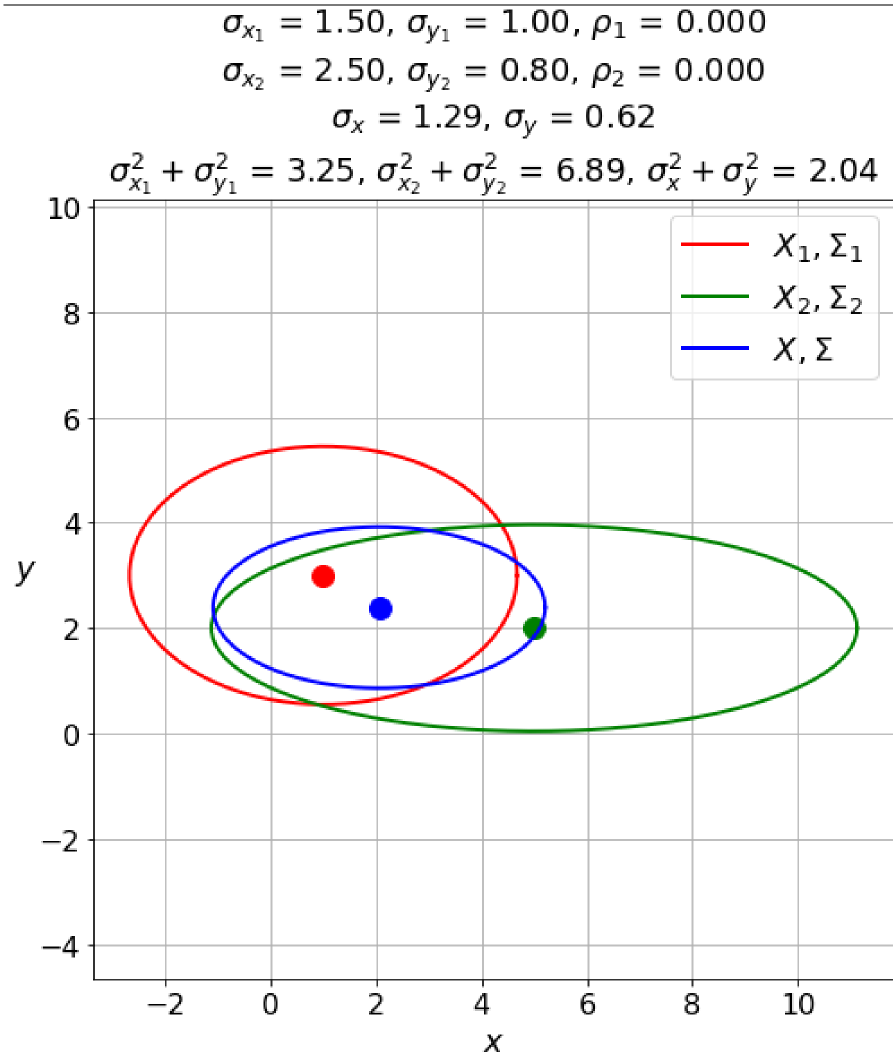
$$A_1 = \begin{bmatrix} a_x & 0 \\ 0 & a_y \end{bmatrix} \quad \text{ja} \quad A_2 = \begin{bmatrix} 1 - a_x & 0 \\ 0 & 1 - a_y \end{bmatrix}$$

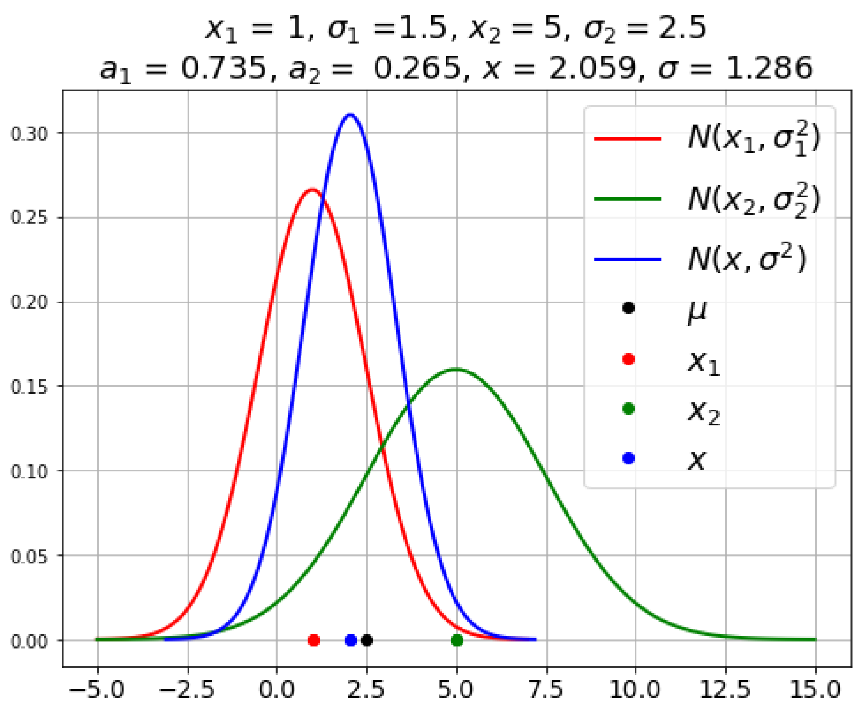
where

$$a_x = \frac{\sigma_{x_2}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2} \quad \text{and} \quad a_y = \frac{\sigma_{y_2}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2}$$

$$x = a_x x_1 + (1 - a_x)x_2, \quad y = a_y y_1 + (1 - a_y)y_2$$

$$\sigma_x^2 = \frac{\sigma_{x_1}^2 \sigma_{x_2}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2}, \quad \sigma_y^2 = \frac{\sigma_{y_1}^2 \sigma_{y_2}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2}, \quad \sigma_{xy} = 0$$


$\sigma_{x_1} = 1.50$, $\sigma_{y_1} = 1.00$, $\rho_1 = 0.000$
$\sigma_{x_2} = 2.50$, $\sigma_{y_2} = 0.80$, $\rho_2 = 0.000$
$\sigma_x = 1.29$, $\sigma_y = 0.62$
$\sigma_{x_1}^2 + \sigma_{y_1}^2 = 3.25$, $\sigma_{x_2}^2 + \sigma_{y_2}^2 = 6.89$, $\sigma_x^2 + \sigma_y^2 = 2.04$
X_1, Σ_1
X_2, Σ_2
X, Σ
y
x


$x_1 = 1$, $\sigma_1 = 1.5$, $x_2 = 5$, $\sigma_2 = 2.5$
$a_1 = 0.735$, $a_2 = 0.265$, $x = 2.059$, $\sigma = 1.286$
$N(x_1, \sigma_1^2)$
$N(x_2, \sigma_2^2)$
$N(x, \sigma^2)$
μ
x_1
x_2
x