Technische Universität Berlin

Department of Geodesy and Geoinformation Science

Chair of Geodesy and Adjustment Theory

Prof. Dr.-Ing. Frank Neitzel

Adjustment Theory

- Lecture Notes -

Last update: 17.10.2011

Content

1. Introduction to error analysis and statistcs

Statistics is the art to acquire data to visualize to analyse and to interpret in order to achieve new knowledge

Basic task of statistics
To describe, to estimate and to decide
To conclude with respect to the whole population

2. Basic notations from calculations of probability and statistics

2.1. Measurement process and mathematical model

"Statistics, as branch of applied mathematics, is the science of occurrences or events which is based on random variables".

Observations can be regarded as random occurrences (events)

Measurement process: includes the parameter value and the description of the kind of getting those measurements.

Assumptions: an unlimited number of repetitions of the measuring process under equal conditions will be possible.

The parameter value is a stochastic quantity (random error)

Using a mathematical model the relations between random variables will be mathematically described. Based upon the choice of the mathematical model the accordance with the reality can be better or worse.

Modelling errors (systematic errors): Modelling errors can be reduced by modification of the mathematical model, if the measurement process is sufficiently well described.

Aim: Choice of a mathematical model where the modelling errors which will be irrelevant with respect to the task to be achieved. (The model has to be as simple as possible but not simpler (Einstein))

Gross errors, blunders, mistakes: For the following considerations it will be assumed that no gross errors will be existent in the data.
Problem: The decision, if there is really a mistake.

2.2. Frequency, relative frequency, relative sum of frequencies

Frequency n_i = Number of measurements, which show the measurement value x_i
(among the $n = \sum n_i$ measurements carried out)

Relative frequency:

$$r_i = \frac{n_i}{n} \quad \text{with} \quad \sum r_i = 1$$

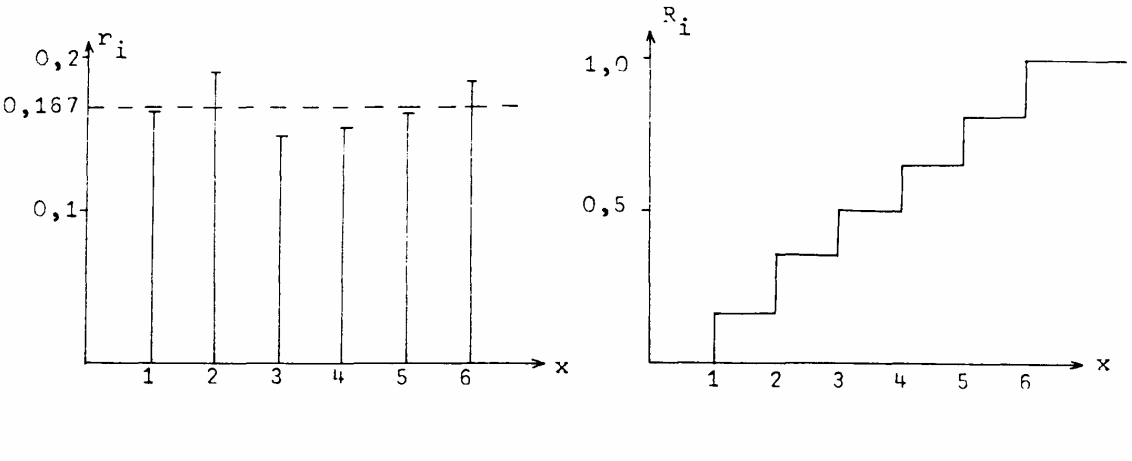
Relative sum of frequencies:

$$R_i = \sum_{\lambda=1}^{i} \frac{n_\lambda}{n} \qquad 0 \le R_i \le 1$$

Fact from experience:

The relative frequency converges stochastically to a boundary value, if the assumption of equal „measurement conditions“is rue.

Graphical representation:



relative frequency sum of relative frequencies

Relative Frequency for the occurrence „to get 2 when throwing dice“ if the number of spot checks n will increase:

n	observed number	theoretical number (the dice is assumed to be ideal)	Δ	r_i (observed)	r_i (theoret.) (ideal dice)
60	9	10	- 1	0,150	0,167
120	22	20	+ 2	0,183	
180	35	30	+ 5	0,194	
240	43	40	+ 3	0,179	
300	52	50	+ 2	0,173	
360	68	60	+ 8	0,189	
420	74	70	+ 4	0,176	
480	82	80	+ 2	0,171	
540	88	90	- 2	0,163	
600	96	100	- 4	0,160	
660	103	110	- 7	0,156	
720	109	120	- 11	0,151	
780	118	130	- 12	0,151	
840	130	140	- 10	0,155	
900	143	150	- 7	0,159	
960	150	160	- 5	0,161	
1020	170	170	0	0,167	
1080	186	180	+ 6	0,172	
1140	197	190	+ 7	0,173	
1200	207	200	+ 7	0,172	

2.3. Probability

To each event A a probability P (A) can be assigned to
(P = probability).

$$0 \leq P(A) \leq 1$$

Definition of probability after Bernoulli/Laplace (Bernoulli 1713!, Laplace 1812-1820):

$$P = \frac{Number\ of\ \ favorable\ cases}{Number\ of\ \ possible\ cases}$$

Assumption: the individual cases are equally possible
(Problem of Idealisation - Mathematical Model)

Definition of probability of v. Mises (1931) (statistical probability):

$$P_i = \lim_{N\to\infty} \frac{f_i}{N} = \lim_{N\to\infty} r_i \qquad \text{(stochastical convergence)}$$

f_i = frequency of the event
N = total number
r_i = relative frequency

Idealization:

The experiments were carried out „under equal conditions“, i.e. all experiments must belong to the same population of events.
v. Mises' definition leads to mathematical problems with respect to lim-idea.

Composed events:

With $\overline{A}$ the event „not A“ named.
$A \cup \overline{A} = E$ (A or $\overline{A}$)

E will represent the certain event, in the same way any event A, B:

$A \cup B = C$ (A or B) (sum)
(Different notation: A + B = C)

Event C is assumed to be either (A or B) or (A and B) at the same time.

Strange events: Events which mutually exclude each others.

Example: Dice A = Getting an even number when throwing dice
B = Getting a 3

Partially mutually covered events:

Example: Dice A = Getting a number < 3 = 1, 2
B = Getting an even number = 2, 4, 6

The portion which is common to both events will be called:

$D = A \cap B$ (as well A as B, „and“) (quantity of intersection, product)
(Also: $A \cdot B = D$)

For the example above: D = getting a 2

Axiomatic reasoning of calculation of probability by Kolmogoroff (1933) (no contradiction to older definitions)

1. If A and B are events, then $\overline{A}$, $A \cup B$, $A \cap B$, $\overline{B}$ are also events.

2. to each event a real number may be assigned $P(A) \geq 0$.

3. The certain event E has the probability $P(E) = 1$.

4. If A and B exclude each other, then $P(A \cup B) = P(A) + P(B)$.

5. If A_1, A_2 … are events which can never happen all together at the same time, then:

$$\lim_{n \to \infty} P(A_1 \cap A_2 \cap \ldots A_n) = 0$$

From these axioms the following relations can be derived:

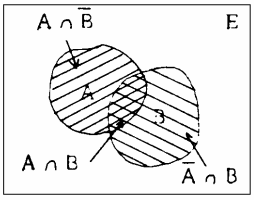
The impossible event 0 has the probability 0. $P(0) = 0$

$$0 \leq P(A) \leq 1$$

For events which are strange to each other in pairs A_1, A_2 … we get:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots A_n) = P(A_1) + P(A_2) + \ldots P(A_n)$$

Relationship between probabilities of the events $A \cup B$ and $A \cap B$ if those are events which do not exclude each other:



Events:

$$A \cup B = (A \cap \overline{B}) \cup (A \cap B) \cup (\overline{A} \cap B)$$
$$A = (A \cap \overline{B}) \cup (A \cap B)$$
$$B = (A \cap B) \cup (\overline{A} \cap B)$$

Probabilities:

$$P(A \cup B) = P(A \cap \overline{B}) + P(A \cap B) + P(\overline{A} \cap B) \quad | +$$
$$P(A) = P(A \cap \overline{B}) + P(A \cap B) \quad | -$$
$$P(B) = P(A \cap B) + P(\overline{A} \cap B) \quad | -$$

$$P(A \cup B) - P(A) - P(B) = - P(A \cap B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example:

A: Throwing dice and getting < 3:

1, 2; $P(A) = \frac{2}{6}$

B: getting an even number:

2, 4, 6; $P(B) = \frac{3}{6}$

$A \cup B$: getting an even number or a number < 3:

1, 2, 4, 6; $P(A \cup B) = \frac{4}{6}$

$A \cap B$: getting an even number < 3:

2; $P(A \cap B) = \frac{1}{6}$

$$P(A \cup B) = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{4}{6}$$

Axiomatic fixes how to calculate with probabilities and not what probability means. It is sufficient to use the interpretation of frequency. For a series of experiments the relative frequency approaches the probability of a getting a defined result, if the conditions of the experiments stay unchanged (stochastical convergence). If $P(A) = p_i$, then with N trials the result A will be achieved in approximately Np_i cases („theoretical number“).

2.4. conditional probability, stochastical independence

Assumption: $P(A) \neq 0$; $P(B) \neq 0$

P(A/B) = Probability of the becoming true of event A under the assumption that event B was already true.

P(B/A) = Probability of the becoming true of event B under the assumption that event A was already true.

Definition of the conditional probability:

$$P(A / B) = \frac{P(A \cap B)}{P(B)} \qquad P(B / A) = \frac{P(A \cap B)}{P(A)}$$

Example: Dice A = 2 B = even number = 2, 4, 6

$$P(A \cap B) = \frac{1}{6} \; ; \; P(A) = \frac{1}{6} \; ; \; P(B) = \frac{1}{2}$$

$$P(A / B) = \frac{1}{6} \cdot \frac{2}{1} = \frac{1}{3} \; ; \; P(B / A) = \frac{1}{6} \cdot \frac{6}{1} = 1$$

Definition of stochastical independence:

$P(A/B) = P(A)$ and $P(B/A) = P(B)$

The knowledge that event B happened before does not give any indication that event A will happen or not happen.

Resulting from the definition of conditional probability:

$$P(A \cap B) = P(A/B)\, P(B) = P(B/A)\, P(A)$$

For stochastically independent events, we get:

$$P(A \cap B) = P(A)\, P(B)$$

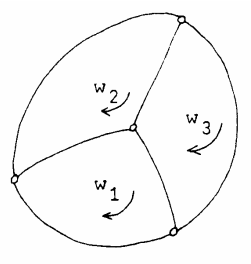
Example:

Levelling net with three closures:

The sign of the misclosures will be regarded as randomly occurred event.

Pre-Assumpton:

Positive and negative signs are equally frequent the signs in the closures are stochastically independent.



Population (if the pre-assumptions are true):

	w_1	w_2	w_3
1	+	+	+
2	+	+	-
3	+	-	+
4	+	-	-
5	-	+	+
6	-	+	-
7	-	-	+
8	-	-	-

$p_i = \text{const}$

Elementary events:

A	(w_1 positive)	P(A) = 0,5
$\overline{A}$	(w_1 negative)	P(A) = 0,5
B	(w_2 positive)	P(B) = 0,5
$\overline{B}$	(w_2 negative)	P(B) = 0,5
C	(w_3 positive)	P(C) = 0,5
$\overline{C}$	(w_3 negative)	P(C) = 0,5

Some combined events:

$B \cap C$ (w_2 and w_3 positive)

$P(B \cap C) = P(B) \times P(C) = 0{,}25$
(Due to the assumption of stochastically independence)

$B \cup C$ (w_2 or w_3 or both positive)

$P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0{,}75$

B/C (w_2 positive under assumption, that w_3 positive)

$$P(B / C) = \frac{P(B \cap C)}{P(C)} = \frac{0{,}25}{0{,}5} = 0{,}5 = P(B)$$

The events B and C are stochastically independent.

$B \cap C/C$ (w_2 and w_3 positive under assumption, that w_3 positive)

$$P(B \cap C / C) = \frac{P(B \cap C \cap C)}{P(C)} = \frac{P(B \cap C)}{P(C)} = \frac{0{,}25}{0{,}5} = 0{,}5 \neq P(B \cap C)$$

The combined event $B \cap C$ is dependent on C.

$B \cap C/\overline{C}$ (w_2 and w_3 positive under assumption, that w_3 negative)

$$P(B \cap C / \overline{C}) = \frac{P(B \cap C \cap \overline{C})}{P(C)} = \frac{0}{0{,}5} = 0$$

The event is impossible.

$A \cap B \cap C$ (all w are positive)

$P(A \cap B \cap C) = P(A \cap B) \times P(C) = P(A) \times P(B) \times P(C) = 0{,}125$

$(A \cap B \cap C) \cup \overline{A} \cap \overline{B} \cap \overline{C}$ (all three misclosures show the same sign)

$P((A \cap B \cap C) \cup \overline{A} \cap \overline{B} \cap \overline{C}) = P(A \cap B \cap C) + P(\overline{A} \cap \overline{B} \cap \overline{C}) =$

$0{,}125 + 0{,}125 = 0{,}25$

$A \cap B \cap C / B \cap C$ (all misclosures are positive under the assumption, that w_2 and w_3 are positive)

$$P(A \cap B \cap C / B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{0{,}125}{0{,}25} = 0{,}5$$

$(A \cap B \cap C) \cup (\overline{A} \cap \overline{B} \cap \overline{C})/C$ (All three misclosures show the same sign under the assumption that w_3 is positiv)

$$P\{[(A \cap B \cap C) \cup (\overline{A} \cap \overline{B} \cap \overline{C})] / C\} = \frac{P\{[(A \cap B \cap C) \cup (\overline{A} \cap \overline{B} \cap \overline{C})] \cap C\}}{P(C)}$$

$$= \frac{P(A \cap B \cap C)}{P(C)} = 0{,}25$$

$$= P\{(A \cap B \cap C) \cup (\overline{A} \cap \overline{B} \cap \overline{C})\}$$

(The events „same sign of all misclosures“ and „w_3 positiv“ are stochastically independent.)

2.5. Stochastical variable, distribution, density function of probability

An event A can be the result of a stochastic experiment (i.e. throw dice) and it can be characterized by a number. In addition the description of the experiments has to be attached to it.

Examples:

a) Dice: l_i = number and description about the arrangement of the numbers on the dice

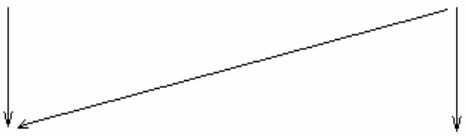
b) observations: l_i = measurement value
Description of the process of measurements, corrections

Characteristics of random events

deterministic events - random events

qualitative characteristics — quantitative characteristics numbers

i.e.. type of car, profession, observer — i.e.. age, income, measurements of all kind



discrete characteristics — continuous characteristics features (all values within a given region are possible)

A subdivision can not be applied to practical problems. The transistions are flowing.

i.e. the weight of a person is a continuous characteristic. Due the limited accuracy of the scale, the weight can only be recorded in classes, the measuring value is discrete.
On the other hand the income of a person is a discrete value. In practical applications discrete characteristics of that type can be treated like continuous characteristics.

The variable assigned to the experiment is called stochastic variable.

There is a distinction between a random variable X and the value x_i, which results for the variable from the experiment i.

i.e. with the experiment of throwing a dice, the resulting number is X

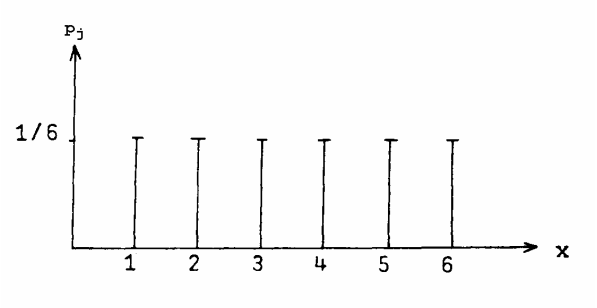
$$P(X = 5) = \frac{1}{6} \quad ; \qquad P(X > 2) = \frac{2}{3}$$

For discrete variables:

$$P(X = x_j) = p_j$$

$$\sum p_j = 1$$

Example: Dice



p_j can be interpreted as a point wise mass distribution.

For continuous variables:

It only makes sense to ask for the probability that an event is within a given interval.

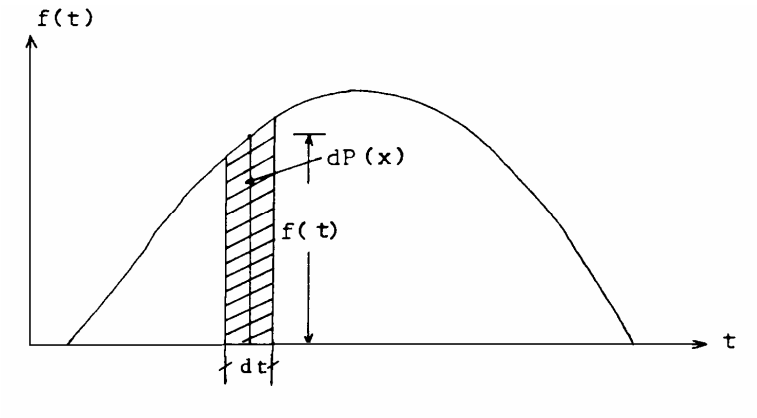
$$P(a < X < b) = \int_a^b f(t)dt$$

f(t) is called density function

$$P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} f(t)dt = 1$$

Differential at position t = x (corresponds to p_j for discrete distribution)

$$dP(x) = P(t < x < t+dt) = f(t)dt$$



f(t) can be seen as continuous mass distribution („probabilistic density function“).

2.6. Distributive function

discrete case:

$F(x) = P(X \leq x)$
(Summation curve of p_j between $-\infty$ to X = x inclusive)

contiuous distributions:

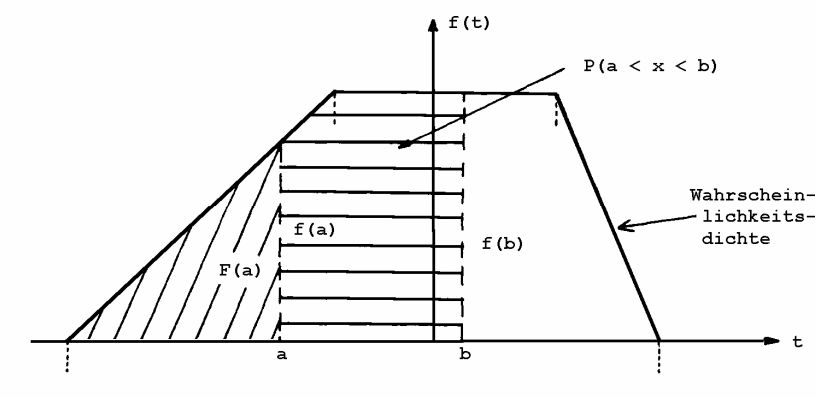
$$F(x) = \int_{-\infty}^{x} f(t)dt = P(-\infty \leq X \leq x)$$

$F(-\infty) = 0$; $F(+\infty) = 1$; $F'(x) = f(x)$

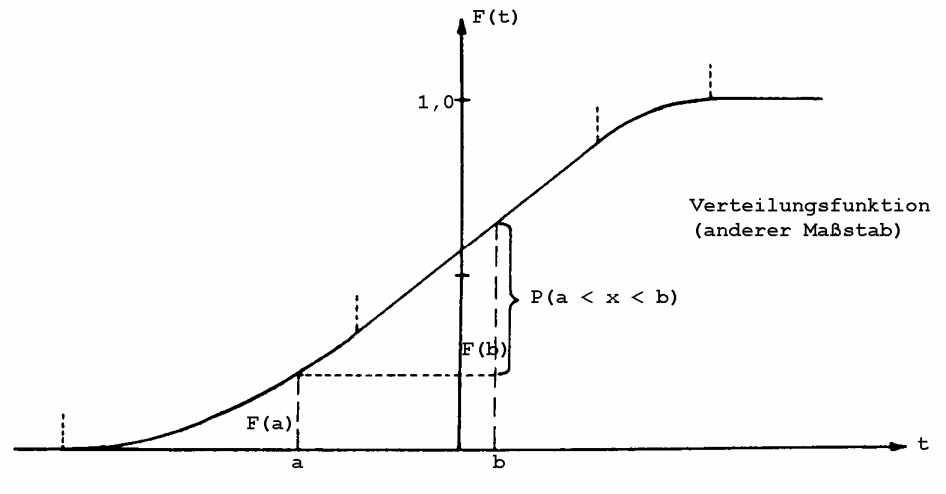
$P(X<b) = P(X<a) + P(a<X<b)$

$F(b) = F(a) + P(a<X<b)$

$$P(a < X < b) = F(b) - F(a) = \int_{a}^{b} f(t)dt$$ probability that X lies in the Intervall a …… b.



f(t) and F(t) in different scale



2.7 characteristic parameters for any density function

2.7.1. Population mean, Expectation value

Population mean: (mean value, „true value“ for continuous distribution)
(scale for position of the distribution, also called expectation value).

Definition of the Expectation value:

<u>For discrete distributions</u>

$$\xi = E(X) = \sum x_i p_i$$

<u>For continuous distributions</u>

$$\xi = E(X) = \int_{-\infty}^{+\infty} \mathrm{x} \cdot \mathrm{f(x)dx} = \int_{-\infty}^{+\infty} \mathrm{x} \cdot \mathrm{dF(x)}$$

According to v. Mises definition of probability:

$$p_i = \lim_{n\to\infty} \mathrm{r_i} = \lim_{n\to\infty} \frac{n_i}{n}$$

We get:

$$\xi = \lim_{n\to\infty} \sum \frac{n_i x_i}{n} = \lim_{n\to\infty} \frac{[l]}{n}$$

If a random variable will be observed very often, it is expected that the arithmetic mean comes very close to the population mean, the true value.

<u>Generalization:</u>

$$E\{g(X)\} = \sum g(x_i) p_i$$

$$E\{g(X)\} = \int_{-\infty}^{+\infty} \mathrm{g(x)} \cdot \mathrm{f(x)dx} = \int_{-\infty}^{+\infty} \mathrm{g(x)} \cdot \mathrm{dF(x)}$$

Example:

g(X) = aX + b (linear function)

$$E(aX+b) = \int_{-\infty}^{+\infty} (ax+b)\cdot f(x)dx = a\cdot \underbrace{\int_{-\infty}^{+\infty} x\cdot f(x)dx}_{E(X)} + b\cdot \underbrace{\int_{-\infty}^{+\infty} f(x)dx}_{1}$$

E(aX+b) = a(EX) + b

A linear function which i soften used, is:

E = ξ -X = E(X) - X (deviation of the population mean = „true“ error)

E(ε) = ξ - E(X) = 0

For nonlinear functions g(X) as a rule, the knowledge of the density function is necessary in order to calculate the population mean value E{g(X)}.

Special nonlinear functions:

The expectation values of the functions $g(X) = X^{\nu}$ with ν = 0, 1, 2, … are called “moments“.

The expectation values of the functions $(-\varepsilon)^{\nu} = (X-\xi)^{\nu}$ with ν = 0, 1, 2, … are called “central moments“.

$E(\varepsilon^0) = 1$
$E(\varepsilon^1) = 0$
$E(\varepsilon^2) = E\{(\xi - X)^2\}$
⋮

2.7.2. Variance, Dispersion, Standard deviation

A measure for the width of the distribution is the variance.

$\sigma^2 = E\{(X-E(X))^2\}$

From that definition it follows:

$$\sigma^2 = E\{(X-\xi)^2\} = E(\varepsilon^2) = \int_{-\infty}^{+\infty} \varepsilon^2 f(\varepsilon)dE$$

$$= E(X^2 - 2X\xi + \xi^2) = E(X^2) - 2\xi\cdot \underbrace{E(X)}_{\xi} + \xi^2$$

$\sigma^2 = E(X^2) - \xi^2$

Variance = (expectation value of X^2) minus (square of the expectation values of X)
Applying v. Mises definition of probability:

$$p_i = \lim_{n\to\infty} r_i = \lim_{n\to\infty} \frac{n_i}{n}$$

We get:

$$\sigma^2 = E(\varepsilon^2) = \lim_{n\to\infty} \frac{n_i}{n} \varepsilon_i^2 = \lim_{n\to\infty} \frac{1}{n} [\varepsilon^2]$$

σ^2 = Variance, dispersion $D^2(X)$

σ = Standard deviation, mean square error

σ refers to one element of the population.

In general:

$$\sigma_g^2 = E\{(g(X) - E(g(X)))^2\}$$

For g(X) = aX + b we get:

$$\sigma^2_{aX+b} = E\{(aX+b - a\xi - b)^2\} = a^2 E\{(X - \xi)^2\} = a^2\sigma^2$$

(„law of error propagation“)

2.8. Interaction of several random variables

2.8.1. distributive function and density function

For an experiment it is of interest to describe the simultaneous happening of several events A, B,…, whose characteristics can be described by the stochastic variables X, Y.

For two variables:

Two-dimensional continuous density distribution f(x,y):

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1$$

Distributive function:

$F(x,y) = P(X<x \cap Y<y)$

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t,z)dtdz$$

If only one characteristic of the population has to be evaluated, we regard:

$F_1(x) = F(x,\infty) = P(X<x \cap Y<\infty) = P(X<x)$

$F_2(y) = F(\infty,y) = P(X<\infty \cap Y<y) = P(Y<y)$

$F_1(x)$ and $F_2(y)$ are called boundary distributions.

If A and B are stochastically independent, then X and Y are called mutually independent.

For <u>independent</u> variables:

$F(x,y) = P(X<x \cap Y<y) = P(X<x) \times P(Y<y)$

$F(x,y) = F_1(x) \times F_2(y)$

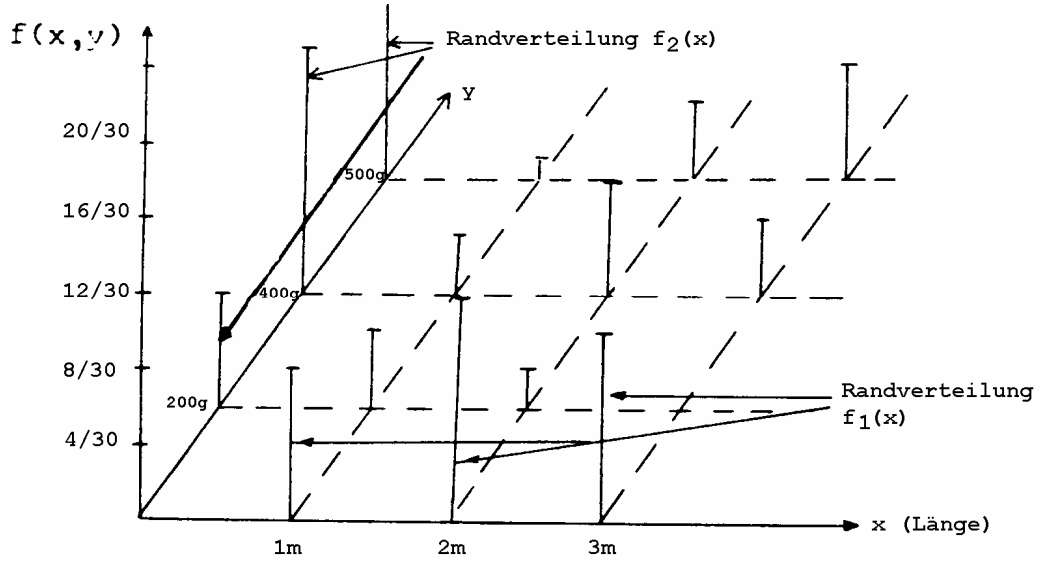
$f(x,y) = f_1(x) \times f_2(y)$

<u>Example:</u> Instruments with different lengths and weights

Characteristic I : lengths 1m, 2m, 3m (x)
Characteristic II : weight 200g, 400g, 600g (y)

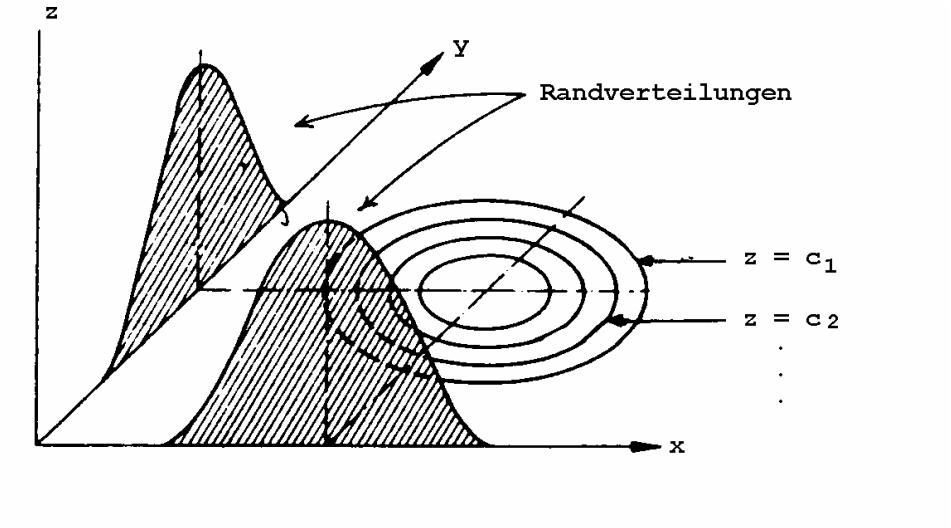
I / II	1m	2m	3m	$f_2(y)$
200g	4	2	0	6/30
400g	3	6	4	13/30
600g	1	4	6	11/30
$f_1(x)$	8/30	12/30	10/30	

$f(x,y) \neq f_1(x) \times f_2(y)$, because x and y are not independent of each other.



Density function f(x,y)

For a continuous two dimensional distribution the density can be represented as spatial surface. The spatial surface may be visualized as „contour-lines“ = lines of equal probability.



2.8.2. Expectation value of a function of several random variables

Definition:

$$E\{g(X,Y)\} = \int_{-\infty}^{+\infty} g(x,y)dF(x,y) = \int_{-\infty}^{+\infty} g(x,y)\cdot f(x,y)dxdy$$

Expectation value for the sum resp. the difference of two variables:

$$g(X,Y) = X \pm Y$$

$$E(X \pm Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x \pm y)dF(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xdF(x,y) \pm \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ydF(x,y)$$

$$= \quad E(X) \quad \pm \quad E(Y)$$

In any case, also if X and Y are not stochastically independent, it follows:

$$\mathbf{E(X \pm Y) = E(X) \pm E(Y)}$$

Expectation value for the product of functions of two variables:

$$g(X,Y) = g(X) \times h(Y)$$

$$E\{g(X)\cdot h(Y)\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)\cdot f(x,y)dxdy$$

For stochastically independent variables the density f(x,y) is split in density $f_1(x)f_2(y)$. Only then the double integral can be represented by the product of two integrals.

$$E\{g(X)\cdot h(Y)\} = \underbrace{\int_{-\infty}^{+\infty} g(x)\, f_1(x)\mathrm{dx}}_{\mathrm{E\{g(X)\}}} \cdot \underbrace{\int_{-\infty}^{+\infty} \mathrm{h(y)}\, f_2(y)\mathrm{dy}}_{E\{h(Y)\}}$$

For stochastically independence of X and Y it holds:

$$\mathrm{E\{g(X) \times h(Y)\} = E\{g(X)\} \times E\{h(Y)\}}$$

The inverse of this expression is not generally valid.

If $\mathrm{E\{g(X)\}\times E\{h(Y)\} = E\{g(X)\times h(Y)\}}$ is true it is not guaranteed that X and Y are independent of each other:

1. Example:

X	Y = aX	X+Y	X×Y
x_1	ax_1	$(1+a)x_1$	ax_1^2
x_2	ax_2	$(1+a)x_2$	ax_2^2
.	.	.	.
.	.	.	.
.	.	.	.
x_n	ax_n	$(1+a)x_n$	ax_n^2
E(X)	E(Y) = aE(X)	E(X+Y) = (1+a)E(X)	$E(X\times Y) = aE(X^2)$
E(X)	E(Y) = aE(X)	E(X)+E(Y) = (1+a)E(X)	E(X)×E(Y) = E(X)×aE(X) = $a\{E(X)\}^2 \neq aE(X^2)$

2. Example: (box with equal number of balls -1 and +1)

X		$Y = X^2$		$X \times Y = X^3$	
i	p	i	p	i	p
-1	0,5	+1	0,5	-1	0,5
+1	0,5	+1	0,5	+1	0,5
E(X) = 0		E(Y) = +1		E(XY) = 0 = E(X)×E(Y)	

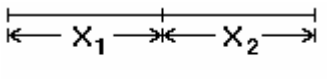
Although X and Y are not stochastically independent

2.8.3. Variance for functions of random variables

2.8.3.1. Linearisation of nonlinear functions

Examples:

1. Measurement of two parts of a distance



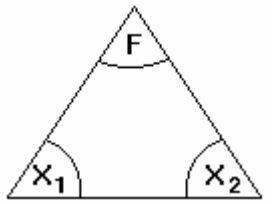
$F = X_1 + X_2$

measured: two parts

searched: total distance

(linear function of random variables X_1 and X_2)

2. third angle in triangle



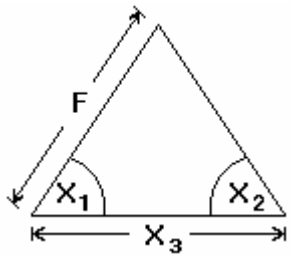
$F = 200^{g} - X_1 - X_2$

measured: two angles in a triangle

searched: third angle

(linear function of random variables X_1, X_2 and the constant value 180^{0})

3. Triangle



$$F = X_3 \frac{\sin X_2}{\sin(X_1 + X_2)}$$

measured: 2 angles (X_1, X_2) and distance (X_3)

searched: distance F

(nonlinear function of the random variables X_1, X_2 and X_3)

Nonlinear functions F may be linearized, neglecting the terms of higher order, if F can be continuously differenciated in the area of consideration.

This condition will always be assumed, except in cases where this will be explicitly excluded.

$F = f(X_1, X_2, \ldots)$
$\sigma_1, \sigma_2, \ldots$

F can be differenciated and it is assumed to be continuously differenciatable in the respective area.

$$F = f(x_1^{\,0}, x_2^{\,0}, \ldots) + (\frac{\partial f}{\partial x_1})_0 (X_1 - x_1^{\,0}) + (\frac{\partial f}{\partial x_2})_0 (X_2 - x_2^{\,0}) + \ldots + \text{elements of higher order.}$$

$x_1^{\,0}, x_2^{\,0}, \ldots$ are approximate values, which will be chosen in such a way that the terms of higher order will vanish.

$(\frac{\partial f}{\partial x_i})_0$ = partial derivative of the x_i at position $x_1^{\,0}, x_2^{\,0}, \ldots$ = constant = f_i

$$F = f_0 + f_1 (X_1 - x_1^{\,0}) + f_2 (X_2 - x_2^{\,0}) + \ldots$$

$$F = f_0 + f_1 X_1 + f_2 X_2 + \ldots$$

Die differential quotient $(\frac{\partial f}{\partial x_i})$ can be replaced by the difference quotient.

Difference quotients $(\frac{\Delta f}{\Delta x_i})$ can be directly calculated.

General form of the linear function of n random variables X_i:

$$F = f_0 + f_1 X_1 + f_2 X_2 + \ldots + f_n X_n$$

2.8.3.2. Variance of a linear function of correlated random variables X.

First, we restrict ourselves to the random variables X_1 and X_2:

$$F = f_0 + f_1 X_1 + f_2 X_2$$

The values f_0, f_1 and f_2 can be defined by the mathematical model which was chosen before.

$E(X_1) = \xi_1$ $\qquad$ $E(X_2) = \xi_2$

$E\{(X_1 - \xi_1)^2\} = \sigma_1^{\,2}$ $\qquad$ $E\{(X_2 - \xi_2)^2\} = \sigma_2^{\,2}$

$E(F) = \phi = f_0 + f_1 \xi_1 + f_2 \xi_2$

$\sigma_F^{\,2} = E\{(F - \phi)^2\}$ (see definition 2.7.1.)

$\sigma_F^{\,2} = E\{[f_0 + f_1X_1 + f_2X_2 - E(f_0 + f_1X_1 + f_2X_2)]^2\}$

$\sigma_F^{\,2} = E\{[f_1 (X_1 - \xi_1) + f_2 (X_2 - \xi_2)]^2\}$

$\sigma_F^{\,2} = E\{f_1^{\,2}(X_1 - \xi_1)^2\} + E\{f_2^{\,2}(X_2 - \xi_2)^2\} + 2 f_1 f_2 E\{(X_1 - \xi_1)(X_2 - \xi_2)\}$

$\sigma_F^{\,2} = f_1^{\,2} E\{(X_1 - \xi_1)^2\} + f_2^{\,2} E\{(X_2 - \xi_2)^2\} + 2 f_1 f_2 E\{(X_1 - \xi_1)(X_2 - \xi_2)\}$

$= f_1^{\,2} E(\varepsilon_1^{\,2}) + f_2^{\,2} E(\varepsilon_2^{\,2}) + 2 f_1 f_2 E(\varepsilon_1 \varepsilon_2)$

$= f_1^{\,2} \sigma_1^{\,2} + f_2^{\,2} \sigma_2^{\,2} + 2 f_1 f_2 E(\varepsilon_1 \varepsilon_2)$

$E(\varepsilon_1 \varepsilon_2) = E\{(X_1 - \xi_1)(X_2 - \xi_2)\} = \text{Cov}(X_1, X_2)$

General definition of the covariance of two variables X_i and X_k (see chapter 2.8.4.)

Error propagation for correlated observables.

$$\sigma_F^2 = f_1^2 \sigma_1^2 + 2 f_1 f_2 \text{ Cov } (X_1, X_2) + f_2^2 \sigma_2^2$$

Extension for more than two random variables

$$F = f_0 + f_1 X_1 + f_2 X_2 + f_3 X_3 + \ldots$$

$$\begin{aligned}\sigma_F^2 = f_1^2 \sigma_1^2 &+ 2 f_1 f_2 \text{ Cov } (X_1, X_2) + 2 f_1 f_3 \text{ Cov } (X_1, X_3) + \ldots \\ &+ f_2^2 \sigma_2^2 \qquad\qquad + 2 f_2 f_3 \text{ Cov } (X_2, X_3) + \ldots \\ &\qquad\qquad + f_3^2 \sigma_3^2 \qquad\qquad + \ldots\end{aligned}$$

In order to calculate the variance σ_F^2 , in addition to the variances σ_i^2 and σ_2^2 all covariances Cov (X_i, X_k) have to be known.

Representation of variance σ_F^2 as a quadratic form with matrix notations:

$$F - f_0 = f_1 X_1 + f_2 X_2 + \ldots = f^T X$$

with: $f^T = (f_1, f_2, \ldots , f_n)$ (vector of known coefficients of random variables X_i)
$X^T = (X_1, X_2, \ldots , X_n)$ (vector of random variables)
Covariances of the random variables X_i:

$$C_{xx} = \begin{pmatrix} \sigma_1^2 & Cov(X_1, X_2) & Cov(X_1, X_3) & \cdots \\ Cov(X_2, X_1) & \sigma_2^2 & Cov(X_2, X_3) & \cdots \\ Cov(X_3, X_1) & Cov(X_3, X_2) & \sigma_3^2 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

Falk's grid

$$\begin{pmatrix} \sigma_1^2 & Cov(X_1, X_2) & Cov(X_1, X_3) \\ Cov(X_2, X_1) & \sigma_2^2 & Cov(X_2, X_3) \\ Cov(X_3, X_1) & Cov(X_3, X_2) & \sigma_3^2 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

$$\begin{pmatrix} f_1 & f_2 & f_3 \end{pmatrix} \begin{pmatrix} \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} \sigma_F^2 \end{pmatrix}$$

$$\sigma_F^2 = f^T C_{xx} f$$

2.8.3.3. Variance of a linear function of stochastically independent random variables X_i

For the special case of stochastically independent observables it holds: (see 2.8.2.):

$$\begin{aligned}\text{Cov } (X_i, X_k) &= E \{ (X_i - \xi_i) (X_k - \xi_k) \} = E(X_i - \xi_i) \times E(X_k - \xi_k) \\ &= E (\varepsilon_i) \times E (\varepsilon_k) = 0\end{aligned}$$

This corresponds to v. Mises's definition with the expression:

$$\lim_{n\to\infty} \frac{1}{n}[\varepsilon_i \varepsilon_k] = 0$$

Covariance matrix for stochastically independent random variables:

$$C_{xx} = \begin{pmatrix} \sigma_1^2 & & & & 0 \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \ddots & \\ 0 & & & & \sigma_n^2 \end{pmatrix}$$

Random variables which show **a covariance of zero** are called **uncorrelated**.
Stochastically independent variables are always uncorrelated. The inverse conclusion is not true in general (see example 2 in 2.8.2.).

„Special law of error propagation for linear functions of independent observables":

$$F - f_0 = f_1 X_1 + f_2 X_2 + \ldots = f^T X$$

with $f^T = (f_1, f_2, \ldots, f_n)$
$X^T = (X_1, X_2, \ldots, X_n)$

and C_{xx} as diagonal matrix we get:

$$\sigma_F^2 = f^T C_{xx} f$$
$$\sigma_F^2 = f_1^2 \sigma_1^2 + f_2^2 \sigma_2^2 + \ldots + f_n^2 \sigma_n^2 = \sum_{\lambda=1}^{n} (f_\lambda \sigma_\lambda)^2$$

Hint: Because the property Cov (X_i, X_k) = 0 has been applied, the equation above is not only valid for stochastically independent, but also for uncorrelated random variables in general.

Basic examples which are frequently mixed up:

1. n stochastically independent random variables are added. (n-times measured) added (see example: Measurement of two positions)

$$F = x_1 + x_2 + \ldots + x_n$$
$$\sigma_F^2 = 1\,\sigma_1^2 + 1\,\sigma_2^2 + \ldots + 1\,\sigma_n^2$$
$$\sigma_1 = \sigma_2 = \ldots = \sigma_n = \sigma$$
$$\sigma_F^2 = n\,\sigma^2$$
$$\sigma_F = \sigma\sqrt{n}$$

2. However: a random variable will be multiplied by n (see example: Indirect measurement of a distance)

$F = n \times x_1 \qquad \sigma_1 = \sigma$
$\sigma_F^2 = n^2 \times \sigma^2$
$\mathbf{\sigma_F = n\,\sigma}$

3. arithmetic mean of n random variables (n-times repeated)

$$F = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

$$\sigma_F^2 = (\frac{1}{n})^2 \sigma_1^2 + (\frac{1}{n})^2 \sigma_2^2 + \ldots + (\frac{1}{n})^2 \sigma_n^2 \qquad \text{with } \sigma_i = \sigma \;\Rightarrow$$

$$\sigma_F^2 = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = n \cdot \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma_F = \frac{\sigma}{\sqrt{n}}$$

Remark: if function F is non linear in X_i, it has to be linearized. Instead of X_i, dX_i resp. ΔX_i are used.

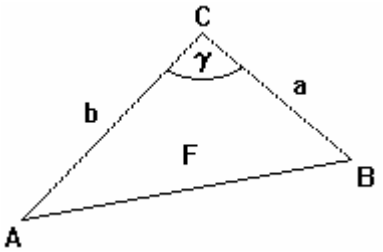
Essential remark:

All parts belonging to a differential dx_i have to be added, prior to applying the formula for σ_F^2.

However:

If several measurements x_i, x_k show equal standard deviations $\sigma_i = \sigma_k$,
only in equation $\sigma_F^2 = \ldots$ .the standard deviations may be substituted by $\sigma_i = \sigma_k$

Example: calculation of the area of a triangle



given: $a \pm \sigma_a$, $b \pm \sigma_b$, $\gamma \pm \sigma_\gamma$

unknown: $F \pm \sigma_F$

$$C = \begin{pmatrix} \sigma_a^2 & & 0 \\ & \sigma_b^2 & \\ 0 & & \sigma_\gamma^2 \end{pmatrix}$$

$$F = \frac{1}{2} a \cdot b \cdot \sin\gamma$$

$$dF = \frac{1}{2} b \cdot \sin\gamma \cdot da + \frac{1}{2} a \cdot \sin\gamma \cdot db + \frac{1}{2} a \cdot b \cdot \cos\gamma \cdot d\gamma$$

Check the dimensions: dγ in analytical unit:

$$dF = \frac{F}{a}da + \frac{F}{b}db + \frac{F}{\tan\gamma}d\gamma$$

$$\sigma_F{}^2 = \left(\frac{F}{a}\right)^2 \sigma_a{}^2 + \left(\frac{F}{b}\right)^2 \sigma_b{}^2 + \left(\frac{F}{\tan\gamma}\right)^2 \sigma_\gamma{}^2 \quad \text{mit} \quad \sigma_\gamma = \frac{\sigma_\gamma{}^{cc}}{\rho^{cc}}$$

For the special case $\sigma_a{}^2 = \sigma_b{}^2 = \sigma_s{}^2$ may only now be summed up.

$$\sigma_F{}^2 = \left[\left(\frac{F}{a}\right)^2 + \left(\frac{F}{b}\right)^2\right]\sigma_s{}^2 + \left(\frac{F}{\tan\gamma}\right)^2\left(\frac{\sigma_\gamma{}^{cc}}{\rho^{cc}}\right)^2$$

$$\sigma_F = \sqrt{\left[\left(\frac{F}{a}\right)^2 + \left(\frac{F}{b}\right)^2\right]\sigma_s{}^2 + \left(\frac{F}{\tan\gamma}\right)^2\left(\frac{\sigma_\gamma{}^{cc}}{\rho^{cc}}\right)^2}$$

2.8.4. Covariance and correlation coefficient of two random variables

2.8.4.1. Definitions and mathematical relations

$$\text{Cov}\,(X_i, X_k) = E\,\{\,(X_i - \xi_i)\,(X_k - \xi_k)\,\}$$

$$= E\,(X_i X_k) - \underbrace{\xi_i\, E\,(X_k)}_{\xi_i\,\xi_k} - \underbrace{\xi_k\, E\,(X_i)}_{\xi_i\,\xi_k} + \xi_i\,\xi_k$$

$$\text{Cov}\,(X_i, X_k) = E\,(X_i \times X_k) - \xi_i \times \xi_k$$

For i = k the definition equation of the variance results:

$$\text{Cov}\,(X_i, X_i) = E\,\{\,(X_i - \xi_i)^2\,\} = \sigma_i^2$$

$$\text{Cov}\,(aX_i, bX_k) = E\,\{\,(aX_i - a\xi_i)\,(bX_k - b\xi_k)\,\} = ab\ \text{Cov}\,(X_i, X_k)$$

Therefore the covariance is not an adequate measure for the correlation.

Correlation coefficient:

$$\rho_{i,k} = \frac{Cov(X_i, X_k)}{\sigma_i \sigma_k}$$

$$\text{Cov}\,(X_i, X_k) = \rho_{ik} \times \sigma_i \times \sigma_k$$

$$\rho_{aX_i, bX_k} = \frac{a \cdot b \cdot Cov(X_i, X_k)}{a \cdot \sigma_i \cdot b \cdot \sigma_k} = \rho_{i,k}$$

The r correlation coefficient is independent of the choice of the random variables.

For $X_k = bX_i$ we get:

$$\sigma_k^2 = b^2 \sigma_i^2 \quad , \text{Cov}(X_i, X_k) = b\,\sigma_i^2$$

$$\rho_{i,k} = \frac{b \cdot \sigma_i^2}{b \cdot \sigma_i \cdot \sigma_i} = +1$$

For $X_k = -bX_i$ we get in the same way:

$$\rho_{ik} = -1$$

The correlation coefficient is a measure of the linear dependency of two random variables.

For the function $F = \frac{1}{\sigma_i} X_i \pm \frac{1}{\sigma_k} X_k$ it holds:

$$\sigma_F^2 = \left\{ \frac{1}{\sigma_i^2}\sigma_i^2 + \frac{1}{\sigma_k^2}\sigma_k^2 \pm 2\frac{1}{\sigma_i \sigma_k} \underbrace{\text{Cov}(X_i, X_k)}_{\rho_{i,k} * \sigma_i * \sigma_k} \right\} \geq 0$$

$$\sigma_F^2 = 2 \pm 2\,\rho_{ik} > 0$$

The condition can only be fulfilled for $-1 \leq \rho_{ik} \leq +1$.

2.8.4.2. Calculation of covariances and covariance matrices

First we restrict ourselves to two random variables X_1 and X_2. For further notation and with respect to individual steps refer to 2.8.3.2.

$$F = f_0 + f_1X_1 + f_2X_2$$
$$G = g_0 + g_1X_1 + g_2X_2$$

$\text{Cov}(F,G) = E\{(F-E(F))\,(G-E(G))\}$ (according to the definition of the covariance)

The constant values f_0 and g_0 don't affect the calculations to follow.

$$\text{Cov}(F,G) = E\{[f_1(X_1 - \xi_1) + f_2(X_2 - \xi_2)] \times [g_1(X_1 - \xi_1) + g_2(X_2 - \xi_2)]\}$$
$$\text{Cov}(F,G) = E\{f_1g_1(X_1 - \xi_1)^2 + f_1g_2(X_1 - \xi_1)\,(X_2 - \xi_2) + f_2g_1(X_1 - \xi_1)\,(X_2 - \xi_2) + f_2g_2\,(X_2 - \xi_2)^2\}$$
$$\text{Cov}(F,G) = f_1g_1\,E(X_1 - \xi_1)^2 + (f_1g_2 + f_2g_1)\,E\{(X_1 - \xi_1)\,(X_2 - \xi_2)\} + f_2g_2\,E(X_2 - \xi_2)^2$$
$$\text{Cov}(F,G) = f_1g_1\,\sigma_1^2 + (f_1g_2 + f_2g_1)\,\text{Cov}(X_1, X_2) + f_2g_2\,\sigma_2^2$$

Extension to n random variables X_i :

$$\begin{aligned}\mathrm{Cov}(F,G) = {} & f_1 g_1 \sigma_1^2 + f_1 g_2 \mathrm{Cov}(X_1, X_2) + f_1 g_3 \mathrm{Cov}(X_1, X_3) + \dots \\ & + f_2 g_1 \mathrm{Cov}(X_1, X_2) + f_2 g_2 \sigma_2^2 + f_2 g_3 \mathrm{Cov}(X_2, X_3) + \dots \\ & + f_3 g_1 \mathrm{Cov}(X_1, X_3) + f_3 g_2 \mathrm{Cov}(X_2, X_3) + f_3 g_3 \sigma_3^2 + \dots \\ & + \dots\dots\end{aligned}$$

The covariance expression of Cov (F,G) can be interpreted as bilinear form, based on the covariance matrix C_{xx} of the random variables X_i as form matrix.

$$\mathrm{Cov}(F,G) = f^T C_{xx}\, g = g^T C_{xx}\, f$$

For the special case of independent random variables X_i we get the result of 2.8.3.3.:

C_{xx} is a diagonal matrix in the case:

$$C_{xx} = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix}$$

and therefore the covariance of two functions of <u>independent</u> random variables is:

$$\mathrm{Cov}(F,G) = f_1 g_1 \sigma_1^2 + f_2 g_2 \sigma_2^2 + \dots + f_n g_n \sigma_n^2 = \sum_{i=1}^{n} f_i g_i \sigma_i^2$$

Remark which is valid for all f_i and g_i :

For $f_i = g_i$, the covariance turns into the variance σ_F^2 .
The covariance matrix C_{FF} of several functions of any random variables:

$$\begin{array}{lllllllll} F_I & = f_{I,1} X_1 & + f_{I,2} X_2 & + f_{I,3} X_3 & + \dots & + f_{I,n} X_n & = f^T{}_I\, X \\ F_{II} & = f_{II,1} X_1 & + f_{II,2} X_2 & + f_{II,3} X_3 & + \dots & + f_{II,n} X_n & = f^T{}_{II}\, X \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ F_u & = f_{u,1} X_1 & + f_{u,2} X_2 & + f_{u,3} X_3 & + \dots & + f_{u,n} X_n & = f^T{}_u\, X \end{array}$$

The roman number in the index indicates the number of the function (here $F_I \equiv F$ and $F_{II} \equiv G$ of the calculation formula).
The elements of the covariance matrix C_{FF} can be calculated individually according to the given formula.

Using:

$$f^T{}_i = (f_{i,1}\, ,\, f_{i,2}\, ,\, \dots\, ,\, f_{i,n})$$

we get

$$\sigma_{Fi}^2 = f^T{}_i\, C_{XX}\, f_i \qquad \text{and} \qquad \mathrm{Cov}(F_i, F_k) = f^T{}_i\, C_{XX}\, f_k = f^T{}_k\, C_{XX}\, f_i$$

If the vectors f_i are arranged in the matrix

$$F_{u,n} = \begin{pmatrix} f_{I,1} & f_{I,2} & \cdots & f_{I,n} \\ \vdots & \vdots & & \vdots \\ f_{u,1} & f_{u,2} & \cdots & f_{u,n} \end{pmatrix} = \begin{pmatrix} f^T{}_I \\ \vdots \\ f^T{}_u \end{pmatrix}$$

The covariance matrix C_{FF} of the Functions $F_I \ldots F_u$ can be calculated using matrix calculus as follows:

$f = Fx \quad f^T = (F_I , F_{II} , \ldots , F_u)$

$$\underset{(u,u)}{C_{FF}} = \underset{(u,n)}{F} \cdot \underset{(n,n)}{C_{XX}} \cdot \underset{(n,u)}{F^T} = \begin{pmatrix} f^T{}_I C_{xx} f_I & f^T{}_I C_{xx} f_{II} & \cdots & f^T{}_I C_{xx} f_u \\ f^T{}_{II} C_{xx} f_I & f^T{}_{II} C_{xx} f_{II} & \cdots & f^T{}_{II} C_{xx} f_u \\ \vdots & \vdots & & \vdots \\ f^T{}_u C_{xx} f_I & f^T{}_u C_{xx} f_{II} & \cdots & f^T{}_u C_{xx} f_u \end{pmatrix}$$

Generalisation of the covariance matrix C_{GG} , if the functions $G_I \ldots G_r$
are functions of $F_I \ldots F_u$
and those are functions of $X_1 \ldots X_n$.

$g = G \times f$
$f = F \times x$ $\qquad C_{FF} = F\, C_{XX}\, F^T$
$C_{GG} = G\, C_{FF}$ $\qquad G^T = G\, F\, C_{XX}\, F^T\, G^T$

Way of calculation:

1. The covariance matrix C_{XX} is known based on theoretical consideration or from prior calculations. C_{FF} will be calculated and then C_{GG}.

2. The values X are independent random variables based on the mathematical model. The C_{XX}-matrix is then diagonal. If only matrix C_{GG} is to search for, the calculation of matrix C_{FF} is not necessary explicitly. However the matrix product G×F has to be calculated. (Going back to stochastically independent variables)

2.9. Efficient estimation

The estimation x of the expectation value ξ has to be determined in such a way that for any function:

$F = f_0 + f_1 X_1 + f_2 X_2 + \ldots = f_0 + f^T X$

The following condition is valid:

$\sigma_F^2 = f^T C_{XX} f \rightarrow$ minimum

This demand corresponds to:

σ_X^2 = minimum

and

$[(x - l_i)^2]$ = minimum.

The deviations v_i of the observables l_i from the estimation x will be called residuals“ v_i:

$l_i + v_i = x$

$v_i = x - l_i$

Due to this, also the following condition holds:

[vv] = minimum

(„Method of least squares“).

2.9.1. Maximum - Likelihood - Method = Method of the largest probability

This method has as a prerequisite that the distribution of the measurements is known. (i.e. a normal distribution)

3. Estimation of empirical variances

3.1. From true errors ε_i

Empirical values of the standard deviation („mean square error“) can be calculated according to:

$$m^2 = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\varepsilon_i^2}{n} = \frac{[\varepsilon\varepsilon]}{n}$$

Even if the requirement $n \to \infty$ is given up, the equations are only in exceptional cases suited to calculate an empirical value for the standard deviation because the true values of the observations and therefore the ε's are unknown in general.

In exceptional cases the expectation value is known, i.e. if a precise distance for calibration purposes has been determined with much higher precision than the actual measurement which has to be evaluated. Then the empirical variance of the measurements can be calculated from true errors ε_i = **precise value – measurement value.**

A random variable is completely defined by its distributive function. It is often sufficient to use characteristic parameters instead which can be calculated from the distributive function. Most essential is the so called **expectation value of the random variable E(x)**. It could be determined as sum of all possible values of x if the probability is taken into account:

$$E(x) = \int_{-\infty}^{+\infty} x * f(x)dx$$

For functional values of equal probability $p_i = 1/n$, $E(x)$ will be the **arithmetic mean** of all possible values of the random variable.

The expectation value does not give any information how much the individual random variables deviate from each other. The second characteristic value of any distributive function is a parameter to describe the distribution that called **variance**.

Here the squares of all deviations of the individual variables from the expectation value will be taken into account, and the probability is considered again.

$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - E(x))^2 * f(x)dx$$

$$E(\varepsilon) = 0$$

$$E(\varepsilon\varepsilon) = E(x - \xi)^2 = \sigma^2 = \int_{-\infty}^{+\infty} \varepsilon^2 f(\varepsilon)d\varepsilon$$

For values ε of equal probability and for a sufficiently large number n of measurement variables, the variance becomes the arithmetic mean of all deviations squared:

$$\sigma^2 = \lim_{n\to\infty} \frac{[\varepsilon\varepsilon]}{n} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{\varepsilon_i^2}{n}$$

m^2 therefore is an estimation of σ^2 under the pre-assumption that n is sufficiently large.

$$m^2 = \frac{[\varepsilon\varepsilon]}{n}$$

The following problem exists:

How can true errors be determined?

3.2. From apparent errors v_i (residuals)

$$\sigma^2 = E(\varepsilon\varepsilon) = \frac{1}{n}\sum \varepsilon^2 \approx \int_1^n \varepsilon^2 f(\varepsilon)d\varepsilon \qquad f(\varepsilon)d\varepsilon = \text{probability}$$

$$\sigma^2 = E(\varepsilon\varepsilon) = \frac{1}{n}\sum \varepsilon^2 \qquad \text{n = large !}$$

estimation : m^2 n small !

problem : ε unknown !

$$[\varepsilon] = 0 \qquad E(\varepsilon) = 0$$

$$\varepsilon = x - E(x) \qquad \textit{uncorrelated}$$

$$y = \textit{estimation for } E(x) \textit{ i.e. } y = \frac{[x]}{n} \; n \textit{ small!}$$

$$v = y - x \qquad (\textit{apparent error})$$

$$[v] = ny - [x] = 0 \quad \rightarrow \qquad y = \frac{[x]}{n}$$

with

$$\varepsilon_i = x_i - E(x) \qquad \textit{and} \qquad v_i = y - x_i$$

$$\begin{aligned}
\varepsilon_1 &= -v_1 + (y - E(x)) \qquad E(\varepsilon_1 \varepsilon_1) = \sigma^2 \\
\varepsilon_2 &= -v_2 + (y - E(x)) \qquad E(\varepsilon_2 \varepsilon_2) = \sigma^2 \\
&\vdots \\
\varepsilon_n &= -v_n + (y - E(x)) \qquad E(\varepsilon_n \varepsilon_n) = \sigma^2
\end{aligned}$$

$$\overline{E(\varepsilon_1 \varepsilon_1 + \varepsilon_2 \varepsilon_2 + \ldots + \varepsilon_n \varepsilon_n) = \; n\sigma^2}$$

Or:

$$\begin{aligned}
\varepsilon_1 \varepsilon_1 &= v_1 v_1 - 2v_1 (y - E(x)) + (y - E(x))^2 \\
\varepsilon_2 \varepsilon_2 &= v_2 v_2 - 2v_2 (y - E(x)) + (y - E(x))^2 \\
&\vdots \\
\varepsilon_n \varepsilon_n &= v_n v_n - 2v_n (y - E(x)) + (y - E(x))^2
\end{aligned}$$

$$\overline{[\varepsilon\varepsilon] = [vv] - 2(y - E(x))\underbrace{[v]}_{0} + n((y - E(x))^2)}$$

$$[\varepsilon\varepsilon] = [vv] + n((y - E(x))^2)$$

$$E(\varepsilon\varepsilon) = n\sigma^2 = E([vv]) + nE((y - E(x))^2)$$

$$y - E(x) = \frac{[x]}{n} - E(x) = \frac{[x - E(x)]}{n} = \frac{[\varepsilon]}{n}$$

$$E(y - E(x))^2 = E\left(\frac{[\varepsilon]^2}{n^2}\right) = \frac{E(\varepsilon\varepsilon)}{n^2} = \frac{n\sigma^2}{n^2} \qquad \textit{stochastically independent}$$

$$E(\varepsilon_i \varepsilon_k) = 0 i \neq k$$

$$E[vv] = [vv]$$

Therefore :

empirical variance m^2 *for* σ^2

$$nm^2 = [vv] + n\frac{nm^2}{n^2}$$

From the residuals of several stochastically independent observations of the same type and the same accuracy the empirical variance can be calculated as follows:

$$\Rightarrow [vv] = (n-1)m^2 \quad or \quad \underline{\underline{m^2 = \frac{[vv]}{n-1}}}$$

The formula above gives approximate values for m , which becomes more and more accurate if only independent observations were used, and then it can be assumed that no systematic errors are present. The number of redundant observations is visible (n-1).

3.3. Introduction of weights

The pre-assumption that all observations are of equal accuracy can not be hold in many cases. In those cases it not allowed just to build the mean of variables of different accuracy. The observational values will therefore be assigned as specific weight in the computations.

$f(\varepsilon_i)d\varepsilon_i \neq \frac{1}{n}$ but in a variable way i.e. <u>*probability*</u> $\left(\frac{p_i}{[p]}\right)$

$$E(\varepsilon_i) = \int_{-\infty}^{+\infty} \varepsilon_i f(\varepsilon_i) d\varepsilon_i = 0$$

$$\frac{\varepsilon_i p_1 + \varepsilon_2 p_2 + ... + \varepsilon_n p_n}{\sum p} = 0 = \frac{[\varepsilon p]}{[p]}$$

$$\sigma^2 = E(\varepsilon_i^2) = \int_{-\infty}^{+\infty} \varepsilon_i^2 f(\varepsilon_i) d\varepsilon_i = \frac{[\varepsilon_i^2 p_i]}{n}$$

$$\sigma^2 = E(\varepsilon\varepsilon p) \approx \int_{-\infty}^{+\infty} \varepsilon^2 f(\varepsilon) d\varepsilon$$

$$\varepsilon = x - E(x)$$

$$v = y - x \qquad\qquad y = \frac{[xp]}{[p]}$$

$$[vp] = o$$

$$\varepsilon_1 = -v_1 + (y - E(x)) \qquad\qquad E(\varepsilon_1 \varepsilon_1 p) = \sigma^2$$

$$\vdots$$

$$\varepsilon_n = -v_n + (y - E(x)) \qquad\qquad E(\varepsilon_n \varepsilon_n p) = \sigma^2$$

$$---------- \qquad\qquad ----------$$

$$E(\varepsilon\varepsilon p) = n\sigma^2 = [vvp] - 2\underbrace{[vp]}_{0}(y - E(x)) + [p](y - E(x))^2$$

$$y - E(x) = \frac{[xp] - [p]E(x)}{[p]}$$

$$= \frac{[x_i p_i - p_i E(x)]}{[p]}$$

$$= \frac{[p_i (x_i - E(x_i))]}{[p]}$$

$$E(x_i - E(x_i))^2 = \frac{\sigma^2}{p_i}$$

$$E(y - E(x))^2 = \frac{E[p_i (x_i - E(x_i)]^2}{[p]^2} \Rightarrow \frac{[p]\sigma^2}{[p]^2}$$

$$E[p_i (x_i - E(x_i))]^2 = p_1^2 E\underbrace{(x_1 - E(x_1))^2}_{\varepsilon_1^2} + p_2^2 E\underbrace{(x_2 - E(x_2))^2}_{\varepsilon_2^2} + \quad \ldots \quad + p_n^2 E\underbrace{(x_n - E(x_n))^2}_{\varepsilon_n^2}$$

$$= p_1 \underbrace{(p_1 E(\varepsilon_1^2))}_{\sigma^2} + p_2 \underbrace{(p_2 E(\varepsilon_2^2))}_{\sigma^2} + \quad \ldots \quad + p_n \underbrace{(p_n E(\varepsilon_n^2))}_{\sigma^2}$$

$$= p_1\sigma^2 + p_2\sigma^2 + p_3\sigma^2 + \quad \ldots \quad + p_n\sigma^2$$

$$= [p]\sigma^2$$

The empirical variance of an observation of weight 1 can be derived from the residuals of differently weighted observations:

$$n\sigma^2 = [vvp] + \frac{[p]}{[p]^2}[p]\sigma^2$$

$$\underline{\underline{m^2 = \frac{[vvp]}{n-1}}}$$

3.4. Weighting of observations

If an observational value has been observed repeatedly with different accuracy, this has to be taken into account when forming the mean.

Problem:

The distance E was determined once with a precise instrument (E_M) and once by less precise measurement (E_S).
The empirical standard deviation $m_M = \pm 4$ mm,
the empirical standard deviation $m_S = \pm 9$ mm.

Determine the empirical standard deviation of the weighted mean! How should the weights be chosen?

$$E = \frac{P_M \cdot E_M + P_S \cdot E_S}{P_M + P_S} = \frac{P_M}{\sum P} \cdot E_M + \frac{P_S}{\sum P} \cdot E_S$$

$$\Rightarrow \qquad \Delta E = \frac{P_M}{\sum P} \cdot \Delta E_M + \frac{P_S}{\sum P} \cdot \Delta E_S$$

The empirical variance of E results from:

$$m_E = \pm\sqrt{\left(\frac{P_M}{\sum P}\right)^2 \cdot m_M{}^2 + \left(\frac{P_S}{\sum P}\right)^2 \cdot m_S{}^2}$$

The weights P_M and P_S should be chosen in such a way that their sum gives 1 and m_E becomes minimal, resulting in $m_E' = 0$

$$m_E = f\left(P_M, P_S\right) = \frac{1}{P_M + P_S}\sqrt{P_M{}^2 \cdot m_M{}^2 + P_S{}^2 \cdot m_S{}^2}$$

because of $\sum P = 1$, it follows:

$$m_E = \sqrt{P_M{}^2 \cdot m_M{}^2 + (1 - P_M)^2 \cdot m_S{}^2}$$

The partial derivative with respect to the weight P_M is:

$$m_E' = \frac{P_M \cdot m_M{}^2}{\sqrt{P_M{}^2 \cdot m_M{}^2 + P_S{}^2 \cdot m_S{}^2}} - \frac{(1 - P_M) \cdot m_S{}^2}{\sqrt{P_M{}^2 \cdot m_M{}^2 + P_S{}^2 \cdot m_S{}^2}} = 0$$

$$\Rightarrow \qquad P_M \cdot m_M{}^2 = P_S \cdot m_S{}^2$$

$$\Rightarrow \qquad \frac{P_M}{P_S} = \frac{m_S{}^2}{m_M{}^2} \qquad \textbf{(1)}$$

For the relations between weights the following rule holds:

$$P_M = \frac{P_S \cdot m_S{}^2}{m_M{}^2} \qquad \Rightarrow \qquad P_M \sim \frac{1}{m_M{}^2}$$

$$P_S = \frac{P_M \cdot m_M{}^2}{m_S{}^2} \qquad \Rightarrow \qquad P_S \sim \frac{1}{m_S{}^2}$$

for any constant value c it holds:

$$P_i = \frac{c}{m_i{}^2} \qquad \textbf{(2)}$$

From (13) follows:
The weights are reversely proportional to the (empirical) variances.

Variance of unit weight:
For simplicity c can be chosen in such a way that a specific observation l_0 gets the weight

$$P_0 = 1$$

According to (2) it follows:

$$P_0 = \frac{c}{m_0{}^2} = 1 \qquad \Rightarrow \qquad c = m_0{}^2 \qquad \textbf{(3)}$$

The (empirical) variance $m_0{}^2$ of that observation which gets weight $P_0 = 1$ is called variance of unit weight.

3.5. some examples of error propagation

When evaluating observations the following tasks have to be performed:

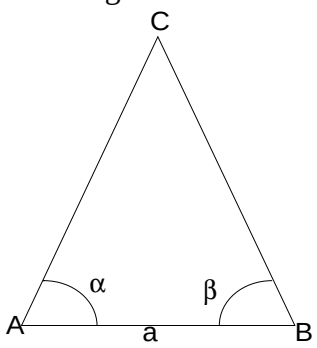
a) to derive the most probable value for a stochastic variable

b) to provide a scale for the accuracy of an individual observation

c) to estimate the accuracy of the mean value and its area of confidence

Task 1

In a triangle ABC the following values were achieved for length a and angles α, β:



a
53.56 m
53.53 m
53.57 m

α	β
65.7689 gon	71.3428 gon
65.7697gon	71.3419 gon
65.7686 gon	71.3434 gon
65.7678 gon	71.3426 gon
65.7692 gon	71.3438 gon

To be determined: -empirical variance of individual measurements
-empirical variance of the mean values

Results:

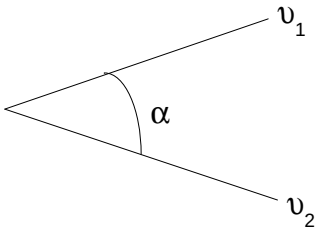
(I) square root of the empirical variance of individual measurements: 0.021 m
distance a = (53.55 ± 0.012) m

(II) square root of the empirical variance of individual measurements: 0.00071 gon
angle α = (65.7688 ± 0.00032) gon

(III) square root of the empirical variance of individual measurements: 0.00073 gon
angle β = (71.3429 ± 0.00032) gon

Task 2

For the determination of α the following variables were measured:



υ_2	υ_1
131.5173 gon	22.3723 gon
131.5168 gon	22.3718 gon
131.5165 gon	22.3725 gon
131.5176 gon	22.3728 gon
131.5171 gon	22.3717 gon

To be determined: a) empirical variance of directions υ_1 and υ_2
b) empirical variance of angle α

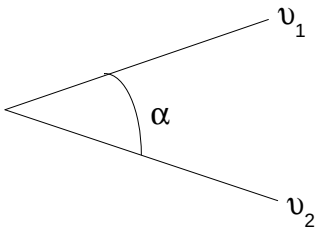
Results:

(I) direction $\overline{\upsilon_2}$ = (131.5171 ± 0.00019) gon

(II) direction $\overline{\upsilon_1}$ = (22.3722 ± 0.00021) gon

(III) angle α = (109.1449 ± 0.00028) gon

Task 3



the angle α was determined 5-times

α
40.9665 gon
40.9659 gon
40.9671 gon
40.9669 gon
40.9663 gon

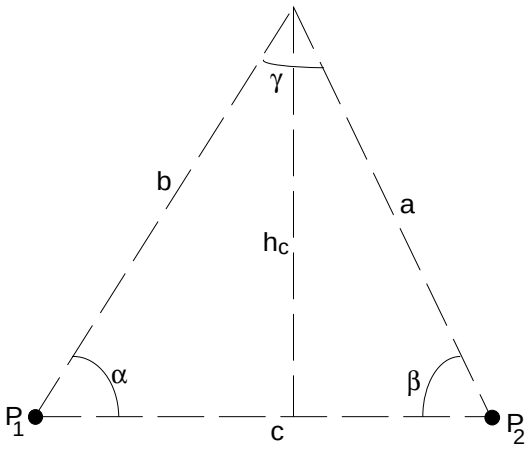
To be calculated: a) empirical variance of the angle
b) empirical variance of the directions υ_1 and υ_2 under the assumption that $m_{x\upsilon 1} = m_{x\upsilon 2}$.

Results:

(I) angle $\alpha = (40.9665 \pm 0.00021)$ gon

(II) empirical variance of the directions $m_{x\upsilon 1} = m_{x\upsilon 2} = 0.00015$ gon

Task 4



The coordinates of two points P_1 and P_2 are given stochastically independent of each other:

$x_1 = 540\ 7681{,}45 \pm 0{,}02$ m
$x_2 = 540\ 7753{,}21 \pm 0{,}04$ m

$y_1 = 351\ 4892{,}84 \pm 0{,}03$ m
$y_2 = 351\ 5168{,}48 \pm 0{,}01$ m

The angles α, β are measured, values of task 1

To be determined:
I) empirical variance s of the measured quantities,
II) values for the lengths c,a and b, angle γ and height h_c and their empirical variances,
III) the area $F = \frac{1}{2} c \cdot h_c$ and the emprical variance of F.

Results:

(I) see task 1
(II) $c = (284.83 \pm 0.033)$ m
$a = (293.03 \pm 0.034)$ m
$b = (307.19 \pm 0.036)$ m
$h_C = (263.84 \pm 0.031)$ m
$\gamma = (62.8883 \pm 0.00045)$ gon
(III) $F = (37575 \pm 6)$ m^2

Task 5

The function $U = \frac{\sin\alpha \cdot \cos\beta \cdot \ln a}{a}$ is to be calculated of the following values, including their corresponding empirical variances:

α	β	a
17.6978 gon	58.1342 gon	167.32 m
17.6971 gon	58.1348 gon	167.39 m
17.6964 gon	58.1336 gon	167.48 m
17.6973 gon	58.1352 gon	167.45 m

Results:

(I) angle α = (17.6972 ± 0.00029) gon
angle β = (58.1344 ± 0.00035) gon
distance a = (167.41 ± 0.035) m

(II) Functional value U = (5.13042 E-3 ± 8.7 E-7) $\left[\frac{1}{m}\right]$

3.6. Efficient estimation

The estimation x for the expectation value E(x) should be determined in such a way that for any function

$$F = f_0 + f_1 X_1 + f_2 X_2 + \ldots = f_0 + f^T X$$

it is valid that

$$\sigma_F^2 = f^T C_{XX} f \rightarrow \text{minimal}$$

This requirement corresponds to

$$\sigma_X^2 = \text{minimum}$$

and

$$[(x - l_i)^2] = \text{minimum.}$$

The deviations v_i of the observations l_i form the estimable parameter x will be called „residuals" v_i:

$$l_i + v_i = x$$
$$v_i = x - l_i$$

Therefore also the following requirement is valid:

$$[vv] = \text{minimum}$$

(„Method of Least Squares").

4. Parametric adjustment of observations

4.1. Linear best estimation of unknown parameters

A preliminary remark

accuracy of calculated (unknown) parameters is evaluated by the main diagonal elements of the corresponding covariance matrix.

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \quad \rightarrow \quad C_x = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \cdots \\ \sigma_{21} & \sigma_2^2 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \cdots & \sigma_m^2 \end{pmatrix}$$

In order to get one quantity for all unknown parameters, the following character is used:

$$\text{tr}\,(C_X) = \text{Sum of all diagonal elements}$$

B Linear estimation, unbiased

for l observational vector for the unknown parameters x, which follow the relation

$$l + v = A\,x$$

(l is random vector; therefore x and v are also random vectors!) with the covariance-matrix C_l .

1. The unknown parameters x will be derived from l by linear relationship, therefore the following equation is valid:

   $$x = G * l \qquad \text{G has to be determined!} \qquad (1)$$

2. x has to be unbiased, therefore:

   $$x = G * l = G * (Ax - v) \qquad (2)$$

Expectation values:

$$E(x) = G\,(A * E(x) - E(v))$$

Requirement:

$$E(v) = 0 : \text{unbiased, therefore}$$

.

$$E(x) = G * A * E(x)$$

$$\rightarrow \quad G * A = E \quad \text{unit matrix} \qquad (3)$$

C Best Estimation

$$x = G * l \qquad (4)$$

Then the covariance matrix C_X reads:

$$C_X = G\, C_l\, G^T \qquad (5)$$

G has to be determined, that

$$tr\,(C_X) \rightarrow \min.$$

trace of C_X.

Evidently the trace of C_X is

$$(C_X) = \sum_j g_j^T C_l\, g_j \qquad (6)$$

boundary conditions for g_j^T from (3) **GA = E**

$\rightarrow$ $g_j^T * A = e_j^T = \{0,0,\ldots,0,1,0,\ldots,0\}$ (7)

j-th component of x will be most accurate, if the following relation holds:

$g_j^T C_l\, g_j \rightarrow \min.$ under the boundary condition $g_j^T * A = e_j^T$

Lagrange function:

$$\phi(g_j , \lambda_j) = g_j^T C_l\, g_j - 2\,(g_j^T A - e_j^T) * \lambda_j \qquad (8)$$

with $\lambda_i = \begin{pmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{mj} \end{pmatrix}$ (m number of unknowns)

$$\phi_{gj} : ? \quad 2\, g_j^T C_l - 2\, \lambda_j^T A^T = 0 \qquad (9)$$

$$\phi_{\lambda j} : ? \quad g_j^T A - e_j^T = 0 \qquad (10)$$

$\rightarrow$

$$C_l\, g_j - A\, \lambda_j = 0 \qquad \text{(I)}$$

$$A^T g_j = e_j \qquad \text{(II)}$$

Equations (I) and (II) are valid for j = 1, 2, … , m

They can be assembled in matrix formulation as follows:

Then (I) and (II) result in:

$$C_l\, G^T - A\, \Lambda = 0 \qquad \text{(I)}$$

$$A^T G^T = E \qquad \text{(II)}$$

From (I):

$$G^T = [C_l]^{-1} A \Lambda$$

into (II):

$$A^T [C_l]^{-1} A \Lambda = E$$

$$\Lambda = (A^T [C_l]^{-1} A)^{-1}$$

Substituted:

$$G^T = [C_l]^{-1} A (A^T [C_l]^{-1} A)^{-1}$$

$$\mathbf{G = (A^T [C_l]^{-1} A)^{-1} A^T [C_l]^{-1}}$$

As $\mathbf{P = \sigma_0^2 [C_l]^{-1}}$ G can also be written as:

$$\mathbf{G = (A^T P A)^{-1} A^T P}$$

Remarks:

1. Statistics: No pre-assumption was made with respect to the distribution of the random variables!

2. The only requirement is E(v) = 0 ,and a linear relationship between observations and unknown parameters is assumed.

4.2. Parametric adjustment

Between the adjusted observations $\bar{l}_1 \ldots \bar{l}_n$ and the h unknown parameters $x_1, \ldots, x_h$ with n > h the following relations hold:

$$\bar{l} = \mathbf{l + v = Ax} .$$

The adjusted observations $\bar{l}$ result from adding the residuals v to the observations. The observations are further described by the **cofactor matrix $\mathbf{Q_{ll} = (1/m_0^2) C_{ll}}$** .

According to the method of least squares a solution vector x is searched for which minimizes

$$\mathbf{v^T [Q_{ll}]^{-1} v}$$

Using $P = [Q_{ll}]^{-1}$, the weight matrix we get the following task of minimization:

$$\mathbf{\phi = v^T P v \quad ==> min}$$

Due to

$v = Ax - l$ we get:

$$\phi = (x^T A^T - l^T) P (Ax - l)$$

$$\frac{\partial \phi}{\partial x} = 0 \quad ==> \quad A^T PAx - A^T Pl = 0$$

$$A^T PAx = A^T Pl$$

The function represents a minimum because $A^T PA$ is positive definite (2. derivative > 0)

$$x = (A^T PA)^{-1} A^T Pl$$

As

$$P = m_0^2 [C_{ll}]^{-1}$$

also the following equation is valid:

$$x = (A^T [C_{ll}]^{-1} A)^{-1} A^T [C_{ll}]^{-1} l$$

Therefore:

Minimizing the sum of the squared residuals and minimizing the variances of the unknown parameters corresponds to each other!

The remaining parameters are as follows:

$$v = Ax - l = (A (A^T PA)^{-1} A^T P - E) * l$$
$$\bar{l} = Ax = A (A^T PA)^{-1} A^T Pl$$

from

$$\phi = v^T P v \quad \Rightarrow \text{ min, it follows:}$$
$$\frac{\partial \phi}{\partial x} = 0 \quad \Rightarrow \quad 2v^T P \frac{\partial v}{\partial x} = v^T PA = 0$$
$$\mathbf{A^T Pv = 0}$$

An essential matrix for parametric adjustments is:

$$A_0 = A(A^T PA)^{-1} A^T P$$

Properties:

1. Spez. unitmatrix, as
   $$A_0 \cdot A_0 = A_0$$
2. $A_0 \cdot A = E \cdot A = A$ (check yourself)
3. $A_0 v = 0$
4. Eigenvalues of A_0 are 0 or (and) 1:

for eigenvalue λ with corresponding eigenvector x the following must be valid:

$$A_0 x = \lambda x \qquad *$$

$$A_0 A_0 x = \lambda A_0 x \quad \Rightarrow \quad A_0 x = \lambda^2 x \qquad **$$

Corresponding * and ** gives:

$$\lambda x = \lambda^2 x$$

This can only be fulfilled for $\lambda = 1$ or $\lambda = 0$.

Calculation of $[vvP] = v^T Pv$ for preparing the estimation of the variance of unit weight.

a) direct use of v from the observation equations
b) Determination from different control formulas.

b1) $v^T Pv = - l^T Pv$
$[vvP] = - [lvP]$

The prove: $v^T Pv = v^T P(\underbrace{Ax - l}_{=v}) = \underbrace{v^T PAx}_{=0!} - v^T Pl$

b2) $v^T Pv = l^T Pl - x^T(A^T Pl)$
$[vvP] = [llP] - x_1 [alP] - x_2 [blP] - \ldots - x_h [hlP]$.

The prove: $\mathbf{v^T Pv = - l^T Pv = - l^T P(Ax - l) = l^T Pl - l^T PAx}$

b3) $v^T Pv = l^T Pl - (l^T PA)(A^T PA)^{-1}(A^T Pl)$
$[vvP] = [llP] - \{ [alP] Q_{11} + [blP] Q_{22} + \ldots + [hlP] Q_{hh}$
$+ 2 [alP] [blP] Q_{12} + \ldots + 2 [alP] [hlP] Q_{1h}$
$+ \ldots + 2 [h_{-1}lP] [hlP] Q_{h-1,h}\}$

The prove: $v^T Pv = l^T Pl - x^T(A^T Pl) = l^T Pl - l^T PA(A^T PA)^{-1}(A^T Pl)$
$x^T = l^T PA(A^T PA)^{-1}$

b4) $[vvP] = [llP \cdot h]$ reduction in the system of normal equations.

4.2.1. Table of linear relationships for parametric adjustment

Table I: linear relationships for parametric adjustment

	l	x	v	$\bar{l} = l+v$
l =	El			
x =	$[(A^T PA)^{-1}A^T P]l$	Ex		
v =	$[A(A^T PA)^{-1}A^T P-E]l$ = $= (A_0-E)l$		Ev $[E-A(A^T PA)^{-1}A^T P]v =$ $= (E-A_0)v$	
$\bar{l} = l+v =$	$[A(A^T PA)^{-1}A^T P]l =$ $= A_0 l$	Ax		$E\bar{l}$ $A(A^T PA)^{-1}A^T Pl$ = $= A_0 \bar{l}$

E has been used to represent the unit matrix (normally called I, but here it would be confusing)
For independent observations of equal accuracy: $\mathbf{P \rightarrow E}$!

Additional linear relations and definitions:

Extraordinary unit matrix of parametric adjustment

$$A_0 = A(A^T PA)^{-1} A^T P$$

A_0 is singular and idempotent

$$A_0 A = EA = A \text{ !}$$

.

It holds: $(E-A_0)\, v = v$; i.e. $A_0\, v = 0$ Zero results

From:

$A_0\, \bar{l} = \bar{l}$; also $(E-A_0)\, \bar{l} = 0$ $A_0\, v = 0$

Therefore: Residuals are eigenvectors of $(E-A_0)$ $(E-A_0)\, \bar{l} = 0$
Adjusted observations are eigenvectors of A_0 $A^T P v = 0$

4.3. Determination of Cofactor Matrices which result from parametric adjustment

Determination i.e. of:

$Q_{x,x}$ of unknowns x
$Q_{\bar{l}\bar{l}}$ of the adjusted observations $\bar{l} = l+v$
$Q_{v,v}$ of the residuals

The corresponding covariance matrices $C_{x,x}$, $C_{\bar{l}\bar{l}}$, $C_{v,v}$ result by multiplication with the variance of unit weight σ^2.

Solution: The parameters x, $\bar{l}$, v are expressed as linear functions of the original observations and then the law of error propagation has to be applied. Based on the relations of table I we get for Table II:

Table II: Cofactors for parametric adjustment

	l	x	v	$\bar{l} = l+v$
l	$P^{-1} = Q_{l,l}$	$A(A^TPA)^{-1} = Q_{l,x}$	$A(A^TPA)^{-1}A^T-E = Q_{l,v}$	$A(A^TPA)^{-1}A^T = Q_{l,\bar{l}}$
x	$(A^TPA)^{-1}A^T = Q_{x,l} = Q_{x,x}A^T$	$(A^TPA)^{-1} = Q_{x,x}$	0 uncorrelated!	$(A^TPA)^{-1}A^T = Q_{x,\bar{l}}$
v	$A(A^TPA)^{-1}A^T-E = Q_{v,l}$	0 uncorrelated!	$P^{-1}-A(A^TPA)^{-1}A^T = Q_{v,v} = Q_{l,l} - Q_{\bar{l},\bar{l}} = (E-A_0)P^{-1}$	0 uncorrelated!
$\bar{l} = l+v$	$A(A^TPA)^{-1}A^T = Q_{\bar{l},l} = Q_{\bar{l},\bar{l}}$	$A(A^TPA)^{-1} = Q_{\bar{l},x}$	0 uncorrelated!	$A(A^TPA)^{-1}A^T = Q_{\bar{l},\bar{l}} = Q_{l,l} - Q_{v,v} = A_0P^{-1}$

For parametric adjustment: $A_0 = A(A^T PA)^{-1} A^T P$

$$\bar{l} = l+v = [A(A^T PA)^{-1}A^T P]l = A_0 l$$

Applying error propagation:

$$Q_{\bar{l},\bar{l}} = A_0 P^{-1} A_0^T = [A(A^T PA)^{-1}A^T P]\, P^{-1}\, [PA(A^T P^{-1} A)^{-1}A^T] = A(A^T PA)^{-1}A^T .$$

Or based on $\bar{l} = l+v = Ax$, we get:

$$Q_{\bar{l},\bar{l}} = AQ_{x,x}A^T = A(A^T PA)^{-1}A^T .$$

The estimated variances of the parameters can be calculated according to:

$$m_{xi}^2 = Q_{xi,xi}\, m_0^2$$
$$m_{\bar{l}i}^2 = Q_{\bar{l}i,\bar{l}i}\, m_0^2$$

4.4. Proof of the formula m0 =sqrt ([vvP]) / (n-h) for the parametric adjustment

$l + v = Ax$	n observation equations
$(A^T PA)\, x = A^T Pl$	h normal equations
$x = (A^T PA)^{-1} A^T Pl$	solution for the h unknowns x_i

and $v = (A\,(A^T PA)^{-1} A^T P - E) * l = (A_0 - E) * l$
using $A_0 = A(A^T PA)^{-1} A^T P$.

In general the following relation holds for the trace of a matrix product:

$$tr\,(XAY) = tr\,(YXA)$$

and applying it to A_0

$$tr\, A(A^T PA)^{-1} A^T P = tr\,(A^T PA)\,(A^T PA)^{-1} = tr\, E(\text{unit matrix}) = h$$
(4.4.1.)

For regular matrix $(A^T PA)$.

A is a (n,h)-matrix,

$l = (l_1, l_2, \ldots, l_n)$ is a random vector with the elements of the random variables:

l_1 and the realisations: $l_1^1, l_1^2, \ldots, l_1^m$
l_2 and the realisations $l_2^1, l_2^2, \ldots, l_2^m$
……
l_n and the realisations $l_n^1, l_n^2, \ldots, l_n^m$

However for practical reasons only **one** realisation of the l_i will be generated and used.

Given is the symmetrical weight matrix $P = P_{ll}$ of the (correlated or uncorrelated) observations l. It holds:

$$P_{ll} = Q_{ll}^{-1} = \sigma_0^2 \, C_{ll}^{-1} \qquad (4.4.2.)$$

Imagine the experiment adjustment would be repeated m times and the mean will be calculated, i.e. the expectation values will be used, and therefore it holds: i

$$E(l) + \underbrace{E(v)}_{=0} = A \cdot E(x) \quad \rightarrow \quad E(l) = A \cdot E(x) \qquad (4.4.3.)$$

Corresponding to the expectation values of the observations the expectation values of the unknowns are calculated which are consistent:

With $E(l) = A\,E(x)$
it holds: $A^T P E(l) = (A^T P A)\,E(x)$
and therefore $E(x) = (A^T P A)^{-1} A^T P E(l)$
and

$$\underbrace{A \cdot E(x)}_{E(l)} = \underbrace{A(A^T P A)^{-1} \cdot A^T P E(l)}_{A_0 \cdot E(l)} \qquad (4.4.4.)$$

If we define in an analogous way true errors:

$$\varepsilon_q^{\,i} := l_q^{\,i} - E(l_q) \qquad \leftrightarrow \qquad l_q^{\,i} = \varepsilon_q^{\,i} + E(l_q)$$

for the i-th realisation of the random observation l_q , we can formulate for the residuals v_i :

$$\begin{aligned} v^i &= (A_0 - E)\, l^i = (A_0 - E)\,(\varepsilon^i + E(l)) \\ &= \underbrace{(A_0 - E)\,E(l)}_{=0} + (A_0 - E)\,\varepsilon^i \end{aligned} \qquad (4.4.5.)$$

due to(4.4.4.)

For the sum $v^T P v$ in the i-th realisation we get:

$$(v^i)^T P_{ll} v^i = (\varepsilon^i)^T (A_0^T - E)\, P_{ll}\, (A_0 - E)\, (\varepsilon^i)$$

And after the multiplication and rearrangement:

$$(v^i)^T P_{ll} v^i = ((\varepsilon^i)^T P_{ll}\, (\varepsilon)) - ((\varepsilon^i)^T P_{ll}\, A_0\, (\varepsilon^i)) \qquad (4.4.6.)$$

$$= \qquad (1) \qquad\qquad (2)$$

Both quadratic forms (1) and (2) can be reformulated as traces of matrix products:

$$(v^i)^T P_{ll} v^i = \mathrm{tr}\,(\varepsilon^i\,(\varepsilon^i)^T P_{ll}) - \mathrm{tr}\,(\varepsilon^i\,(\varepsilon^i)^T P_{ll}\, A_0)$$

For the expectation values we get:

$$E(v^T P v) = tr\,(\underbrace{E(\varepsilon\,\varepsilon^T)}_{C_{ll}}\,P_{ll}) - tr\,(\underbrace{E(\varepsilon\,\varepsilon^T)}_{C_{ll}}\,P_{ll}\,A_0)$$

And as $C_{ll} = Q_{ll}\,\sigma_0^2 = \sigma_0^2\,P_{ll}^{-1}$

it follows:
$$E(v^T P v) = tr\,(\sigma_0^2\,P_{ll}^{-1}\,P_{ll}) - tr\,(\sigma_0^2\,P_{ll}^{-1}\,P_{ll}\,A_0)$$
$$= \sigma_0^2\,tr\,E - \sigma_0^2\,tr\,A_0 = \sigma_0^2\,(n-h) \qquad (4.4.7.)$$

And therefore we get for the variance factor:

$$\sigma_0^2 = \frac{E(v^T P v)}{n-h} \;\hat{=}\; (4.4.7.)$$

As we have only one realisation we get an estimate for the variance of unit weight according to:

$$m_0^2 = \frac{vPv}{n-h} \qquad (4.4.8.)$$

4.5. m_0 = determined for one observation

<u>$v^T Pv$-control formula</u>:

$v = Ax - l$;
$A^T PAx = A^T Pl$
$v^T Pv = (x^T A^T - l^T)\,P\,(Ax - l) = x^T A^T PAx - l^T PAx - x^T A^T Pl + l^T Pl$
$= l^T Pl - x^T A^T Pl$

with $x = (A^T PA)^{-1}\,A^T Pl$ => $l^T Pl - l^T PA(A^T PA)^{-1}\,A^T Pl = l^T(P - PA(A^T PA)^{-1}\,A^T P)l$

$v^T Pv = l^T P(E-A_0)l$

using $A_0 = A(A^T PA)^{-1}\,A^T P$ we get:

$v_i = (A_0-E)_{ii}\,l_i = (A_0-E)_{ii}\,\varepsilon_i$

$v_i^2 p_i = \varepsilon_i^2\,p_i\,r_i$ for $r_i = diag\,(E-A_0)_{ii}$
$E(v_i^2 p_i) = E(\varepsilon_i^2)\,p_i\,r_i$
$E(\varepsilon_i^2) = \sigma_i^2 = E(\frac{v_i^2 p_i}{p_i r_i})$
Due to $m_0^2\,p_0 = m_i^2\,p_i$ and m_0 as estimation for σ_0, we get:

$$m_0^2 = \frac{v_i^2 p_i}{r_i}$$

m_0^2 = estimation for variance of unit weight, derived from one observation i from v_i

In general it holds:

$v_i = (A_0\text{-}E)_{ii}\ \varepsilon_i$

$v_i = -\ r_i \cdot \varepsilon_i$ is also valid if a blunder ∇l_i occurred in observation i
$v_i = -\ r_i \cdot \nabla l_i$

Assumption $\nabla l_i = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \nabla l_i \\ 0 \\ 0 \\ 0 \end{pmatrix}$ one observation is false

m_0^2 as function of ∇l_i :

$$m_0^2 = \frac{\nabla l_i^2 r_i^2 p_i}{r_i} = r_i\ p_i\ \nabla l_i^2$$

For given m_0 :

$$\nabla l_i^2 = \frac{m_0^2}{r_i p_i} \qquad \rightarrow \qquad \nabla l_i = \frac{m_0}{\sqrt{r_i p_i}}$$

Transition to m_0^2 of the adjustment

$$m_0^2 = \frac{v^T P v}{r}$$ r total redundancy

$$v^T P v = \sum_{i=1}^{n} v_i^2 p_i$$ for P as diagonal matrix

Therefore:

$$r \cdot m_0^2 = \sum_{i=1}^{n} r_i\ m_{0_i}^2$$

$$m_0^2 = \frac{\sum_{i=1}^{n} r_i\ m_{0_i}^2}{r}$$

m_0^2 is the weighted mean of the individual m_{0i}^2 ,

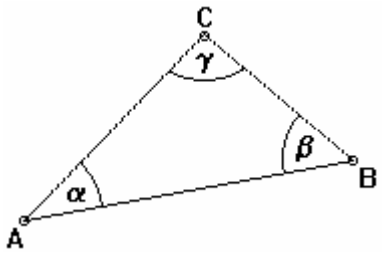
weight here: r_i , as $\sum_{i=1}^{n} r_i = r$

5. Special forms of adjustment calculus

5.1. Adjustment based on conditions of compatibility only

Example:

1. Sum of angles in a planar triangle or with spherical excess:
$(l_\alpha + v_\alpha) + (l_\beta + v_\beta) + (l_\gamma + v_\gamma) = 180° +$ spherical excess



For every „redundant“observation a consistency condition equation has to be formulated between the adjusted observations. This guarantees that the result is consistent.

Analogous to the parametric adjustment it holds:

- The adjusted observations fulfil the conditions exactly,
- The sum of the (weighted) squared residuals will be minimal. **[vvp] -> min**.

5.2. Derivation of the formula of condition adjustment

For an over determined (redundant) set of n observations with redundancy r , there are
n observations $l_1 , l_2 , \dots , l_n$ with
weights $p_1 , p_2 , \dots , p_n$

given.
Between the adjusted observations conditions exist of the form:

$$f(\bar{l}) = f(l+v) = s$$

In detail:

$$f_1 (l_1+v_1 , l_2+v_2 , \dots , l_n+v_n) = s_1$$
$$f_2 (l_1+v_1 , l_2+v_2 , \dots , l_n+v_n) = s_2$$
$$\vdots$$
$$f_r (l_1+v_1 , l_2+v_2 , \dots , l_n+v_n) = s_r$$

The condition equations have to be fulfilled observing the requirement:

$$\mathbf{v^T P v = (\bar{l} - l)^T P (\bar{l} - l) => min.}$$

Approach according to Lagrange applying „Lagrange multipliers“:

Functions , to be minimized:

$$\Phi = v^T P v - 2k^T (f(l+v) - s)$$

resp..

$$\Phi = \Phi(\bar{l}, k) = (\bar{l} - l)^T P (\bar{l} - l) - 2k^T (f(\bar{l}) - s)$$

$$\frac{\partial \Phi}{\partial l} = 2P(\bar{l} - l) - 2(\frac{\partial f(\bar{l})}{\partial \bar{l}})k = 0$$
$$\frac{\partial \Phi}{\partial k} = (f(\bar{l}) - s) = 0$$

The following nonlinear equations result:

$$\bar{l} - P^{-1}(\frac{\partial f(\bar{l})}{\partial \bar{l}})k = l \qquad \text{number n}$$
$$f(\bar{l}) = s \qquad \text{number r}$$

5.2.1. Linearization and general linear approach:

$$\bar{l}_i = l_i + v_i \quad <==> \quad \bar{l} = l + v$$

$$\frac{\partial f(\bar{l})}{\partial l} \approx \frac{\partial f(l)}{\partial l} := B$$

Then it holds:

$$f(\bar{l}) = f(l) + (\frac{\partial f(l)}{\partial l}) \cdot v \qquad \text{Linearization according to Taylor}$$

Then we get from the equations above:

$$l + v - P^{-1}B \cdot k = l \quad ==> \quad v = P^{-1}B\,k$$
$$f(l) + B^T \cdot v = s$$
$$(B^T P^{-1} B) \cdot k = s - f(l) = w$$

Nonlinear condition equations:

$$f(\bar{l}) = f(l + v) = s$$

Linearization applying Jacobi-matrix:

$$\frac{\partial f(l)}{\partial l} = B$$

Definition of misclosures:

$$s - f(l) := w$$

Normal equations:

$$(B^T P^{-1} B) \cdot k = w$$

Check of the „correlate equations“

$$v = P^{-1} B \cdot k$$

- replace the $v = P^{-1} B \cdot k$ in the linearized condition equations:

$$B^T v = w$$

- replace the adjusted observations in the original nonlinear condition equations:

$$f(l + v) = f(\bar{l}) = s$$

5.2.2. Determination of residuals and [vvp]

It holds:

$$\left.\begin{array}{l} v = P^{-1} B \cdot k \\ v^T = k^T B^T P^{-1} \end{array}\right\} \quad v^T P v = k^T B^T P^{-1} P P^{-1} B \cdot k$$

$$v^T P v = k^T \underbrace{(B^T P^{-1} B) \cdot k}_{=w} = k^T w = w^T k$$

5.2.3. Table of the linear relations for adjustment of conditions

Table I: Linear relations

	1 l	2 w	3 k	4 v	5 $\bar{l}$
1 l =	E·l				
2 w =	s - B^Tl (s - f(l))	E·w	$(B^T P^{-1} B)$k	B^Tv	
3 k =	$(B^T P^{-1} B)^{-1}$ (s-B^Tl) = -$(B^T P^{-1} B)^{-1}$ B^Tl + $(B^T P^{-1} B)^{-1}$ s	$(B^T P^{-1} B)^{-1}$ ·w	E·k	$(B^T P^{-1} B)^{-1}$·B^Tv	
4 v =	-$P^{-1}B(B^T P^{-1} B)^{-1}$ B^Tl + $P^{-1}B(B^T P^{-1} B)^{-1}$·s = -$B_0$l + $\bar{s}$	$P^{-1}B(B^T P^{-1} B)^{-1}$ ·w	P^{-1}B·k	E·v B_0·v	
5 $\bar{l}$ =	(E-$P^{-1}B(B^T P^{-1} B)^{-1}$ B)l + $\bar{s}$ = (E - B_0)l + $\bar{s}$				E·$\bar{l}$ (E - B_0)·$\bar{l}$ + $\bar{s}$

Introducing extraordinary unit matrix“ $B_0 := P^{-1}B(B^TP^{-1}B)^{-1}B^T$ it holds:

a) $B^TB_0 = B^T$
b) $(B_0)^n = B_0$ I
c) $B_0 \cdot v = v$ ==> $B_0 \cdot v = \lambda \cdot v$ possible for $\lambda = 1$

d) $\bar{l} = (E-B_0)\cdot \bar{l} + \bar{s}$ ==> $(E-B_0)\cdot \bar{l} = \lambda \cdot \bar{l} + \bar{s}$,
$\bar{l}$ are eigenvalues of $(E-B_0)$ for $\lambda = 1$, if $\bar{s}$ is zero

e) B_0 only has eigenvalues 0 and 1:
due to:

$$B_0 \cdot x = \lambda \cdot x \qquad (*)$$

$$\underbrace{B_0 \cdot B_0}_{B_0} \cdot x = \lambda \cdot \underbrace{B_0 \cdot x}_{=\lambda \cdot x} \quad \Rightarrow \quad B_0 \cdot x = \lambda^2 \cdot x \qquad (**)$$

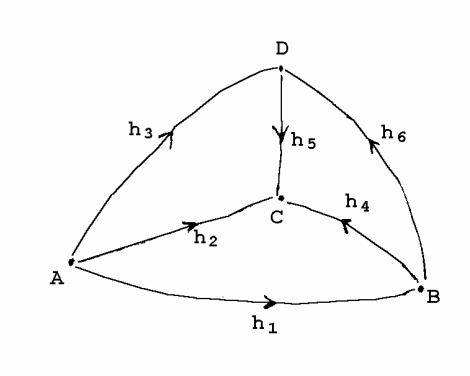
5.2.4. Cofactors of parameters which result from the adjustment

Analogous to the parametric adjustment, we get:

Table II: cofactor matrices

	1 l	2 w	3 k	4 v	5 $\bar{l}$
1 l	$Q_{l,l} = P^{-1}$	$Q_{l,w} = -P^{-1}B$	$Q_{l,k} = -P^{-1}B(B^TP^{-1}B)^{-1}$	$Q_{l,v} = -P^{-1}B(B^TP^{-1}B)^{-1} \cdot B^TP^{-1}$	$P^{-1}-P^{-1}B(B^TP^{-1}B)^{-1} \cdot B^TP^{-1}$
2 w	$Q_{w,l} = -B^TP^{-1}$	$(B^TP^{-1}B)$	E	B^TP^{-1}	0
3 k	$-(B^TP^{-1}B)^{-1}\ B^TP^{-1}$	E	$(B^TP^{-1}B)^{-1}$	$(B^TP^{-1}B)B^TP^{-1}$	0
4 v	$-P^{-1}B(B^TP^{-1}B)^{-1} \cdot B^TP^{-1}$	$P^{-1}B$	$P^{-1}B(B^TP^{-1}B)^{-1}$	$P^{-1}B(B^TP^{-1}B)^{-1} \cdot B^TP^{-1} = B_0P^{-1}$	0
5 $\bar{l}$	$P^{-1}-P^{-1}B(B^TP^{-1}B)^{-1}BP^{-1} = (E - B_0)P^{-1}$	0	0	0	$P^{-1} - Q_{v,v}$

Example



Observed height differences:

h_1 = 1.015 m;	s_1 = 6.25 km
h_2 = 12.570 m;	s_2 = 4.7 km
h_3 = 6.161 m;	s_3 = 7.15 km
h_4 = 11.563 m;	s_4 = 3.95 km
h_5 = 6.414 m;	s_5 = 4.25 km
h_6 = 5.139 m;	s_6 = 5.50 km

Main system: h_1, h_2, h_3

n = 6	number of observations
u = 3	number of unknowns
r = n-u = 3	number of conditions

Condition equations:

general: $f(\bar{l}) = f(l+v) = s$

1) $h_3 + v_3 + h_5 + v_5 - h_2 - v_2 = 0$
2) $h_2 + v_2 - h_4 - v_4 - h_1 - v_1 = 0$
3) $h_4 + v_4 - h_5 - v_5 - h_6 - v_6 = 0$

$B^T v = w$

$$\underbrace{\begin{pmatrix} 0 & -1 & 1 & 0 & 1 & 0 \\ -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{pmatrix}}_{B^T} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \underbrace{\begin{pmatrix} h_2 - h_5 - h_3 \\ h_1 + h_4 - h_2 \\ h_5 + h_6 - h_4 \end{pmatrix}}_{w}$$

Weighting: $P^{-1} \sim S$

P^{-1} = [6.25; 4.7; 7.15; 3.95; 4.25; 5.50]

$$B^T \underbrace{P^{-1}B \cdot k}_{v} = w \qquad \text{Normal equations}$$

$k = (B^T P^{-1} B)^{-1}\, w$

$v = P^{-1} B\, k$

$\bar{l} = l + v$

$v^T P v = w^T k$

In numbers:
$$B^T P^{-1} B = \begin{vmatrix} 16.1 & -4.7 & -4.25 \\ -4.7 & 14.9 & -3.95 \\ -4.25 & -3.95 & 13.7 \end{vmatrix}$$

$\det(B^T P^{-1} B) = 2306$

Determinant value:
$$16{,}1\ \overbrace{(14{,}9 * 13{,}7 - 3{,}95 * 3{,}95)}^{188{,}5} = 3035{,}3$$

$$+\ 4{,}7\ \overbrace{(-4{,}7 * 13{,}7 - 3{,}95 * 4{,}25)}^{-81{,}2} = -381{,}6$$

$$-\ 4{,}25\ \overbrace{(4{,}7 * 3{,}95 + 14{,}9 * 4{,}25)}^{-81{,}9} = \underline{-348{,}1}$$

$$\sum = 2306$$

$$(B^T P^{-1} B)^{-1} = \frac{1}{2306} \begin{vmatrix} 188 & 81{,}2 & 81{,}9 \\ & 202 & 83{,}6 \\ & & 218 \end{vmatrix}$$

$$w = \begin{vmatrix} -5 \\ +8 \\ -10 \end{vmatrix} \quad [\text{mm}] \qquad \text{misclosures}$$

$$k = \begin{vmatrix} -0{,}481 \\ 0{,}462 \\ -0{,}833 \end{vmatrix} \quad [\text{mm}]$$

$v^T = [\ -1{,}0 \quad +3{,}0 \quad -3{,}4 \quad -3{,}9 \quad +1{,}5 \quad +4{,}6]$ [mm]

Adjusted height differences

$\bar{h}_1 = 1{,}014$ m
$\bar{h}_2 = 12{,}573$ m
$\bar{h}_3 = 6{,}157_6$ m
$\bar{h}_4 = 11{,}599_1$ m
$\bar{h}_5 = 6{,}415_5$ m
$\bar{h}_6 = 5{,}143_6$ m

$v^T P v \overset{?}{=} w^T k$ (check)

$11{,}9 = 12{,}0 \quad v$

Precision values:

$$m_0^2 = \frac{v^T P v}{r} = \frac{12{,}0}{3} = 4$$

$$m_0 = \pm 2 \text{ mm/km}$$

$$Q_{v,v} = P^{-1}B(B^TP^{-1}B)^{-1}B^TP^{-1}$$

$$Q_{v,v}\cdot P = P^{-1}B(B^TP^{-1}B)^{-1}B^T = B_0$$

Diagonal of $Q_{v,v}P$ --> partial redundancy factor

$$P^{-1}B = \begin{vmatrix} 0 & -6{,}25 & 0 \\ -4{,}7 & 4{,}7 & 0 \\ 7{,}15 & 0 & 0 \\ 0 & -3{,}95 & 3{,}95 \\ 4{,}25 & 0 & -4{,}25 \\ 0 & 0 & -5{,}50 \end{vmatrix}$$

$$(B^TP^{-1}B)^{-1} = \begin{vmatrix} 0{,}55 & & & & & \\ & 0{,}46 & & & & \\ & & 0{,}58 & & & \\ & & & 0{,}43 & & \\ & & & & 0{,}45 & \\ & & & & & 0{,}43 \end{vmatrix}$$

$$B^T = \begin{vmatrix} 0 & -1 & 1 & 0 & 1 & 0 \\ -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{vmatrix}$$

$$\text{Tr}\,(Q_{v,v}P) = 2{,}9 \quad (\text{Shall} = 3{,}0)$$

$$Q_{\bar{l}\bar{l}} = P^{-1} - Q_{v,v} = (E - Q_{v,v}P)\cdot P^{-1}$$

$$\text{Diag}\,(Q_{\bar{l}\bar{l}}) = [\ 2{,}81 \quad 2{,}54 \quad 3{,}00 \quad 2{,}25 \quad 2{,}34 \quad 3{,}14\ ]$$

$$m_{\bar{h}_1} = m_0\sqrt{Q_{\bar{l}\bar{l}}} = \pm 2{,}0 \cdot \sqrt{2{,}81} = \pm 3{,}4 \text{mm}$$

$$m_{\bar{h}_2} = \pm 3{,}2 \text{mm}$$

$$m_{\bar{h}_3} = \pm 3{,}5 \text{mm}$$

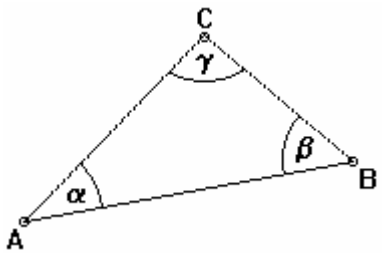
$$m_{\bar{h}_4} = \pm 3{,}0 \text{mm}$$

$$m_{\bar{h}_5} = \pm 3{,}1 \text{mm}$$

$$m_{\bar{h}_6} = \pm 3{,}5 \text{mm}$$

5.3. Parametric adjustment with conditions between the unknowns:

1. Example



measured: α, β, γ

$\alpha + v_\alpha = x$ $\qquad$ $x + y + z = 200$

$\beta + v_\beta = y$

$\gamma + v_\gamma = z$

2. Example

6-parameter-transformation

$x_i + v_{xi} = a\xi_i + b\eta_i + c$

$y_i + v_{yi} = d\xi_i + e\eta_i + f$

conditions:

$\left.\begin{aligned} a + e = 0 \\ b - d = 0 \end{aligned}\right\}$ $(4 - parameters)$

$a^2 + b^2 = 1$ (3-parameters)

4-parametric tansformation

$x_i + v_{xi} = a\xi_i + b\eta_i + c$
$y_i + v_{yi} = b\xi_i - a\eta_i + d$

3-parametric-transformation

$x_i + v_{xi} = \xi_i \cos\varphi + \eta_i \sin\varphi + c$
$y_i + v_{yi} = \xi_i \sin\varphi - \eta_i \cos\varphi + d$

Approach:

$$\left.\begin{aligned} v = f(x) - \tilde{l} \\ g(x) = \tilde{c} \end{aligned}\right\} \text{ nicht linear}$$

Linearization:

$$\begin{aligned} v &= A\,\Delta x - l \\ C^T \Delta x &= c \end{aligned}$$

Solution:

$$\Phi \equiv v^T P v + 2k^T (C^T \Delta x - c) \Rightarrow \min$$

resp..

$$\Phi \equiv (\Delta x^T A^T - l^T)\, P\, (A\Delta x - l) + 2k^T (C^T \Delta x - c) \Rightarrow \min$$

$$\frac{\partial \Phi}{\partial \Delta x} = 0 \quad , \quad \frac{\partial \Phi}{\partial k} = 0$$

$$\frac{\partial \Phi}{\partial \Delta x} = 0 \quad \rightarrow \quad 2\Delta x^T A^T P A + 2k^T C^T = 2l^T P A$$

$$\frac{\partial \Phi}{\partial k} = 0 \quad \rightarrow \quad C^T \Delta x - c = 0$$

$$\begin{aligned} A^T P A \Delta x + C^T k &= A^T P l \\ C^T \Delta x &= c \end{aligned}$$

Matrix of normal equations:

$$\begin{pmatrix} A^T P A & C \\ C^T & 0 \end{pmatrix}$$

Cases:

$A^T P A$ singular: -pivot choice
or - change column and row in advance
total system solved (inverted)

$A^T P A$ regular: $(A^T P A)^{-1}$ exists

Elimination von Δx: 1. multiply by $C^T(A^TPA)^{-1}$ from the left and subtract the second system of equations

$$C^T(A^TPA)^{-1}Ck = C^T(A^TPA)^{-1}A^TPl - c$$
$$k = (C^T(A^TPA)^{-1}C)^{-1}\,(C^T(A^TPA)^{-1}A^TPl - c)$$

k substituted:

$$\Delta x = (A^TPA)^{-1}\,(A^TPl - C\,k)$$

Calculation of cofactor matrices of k and x:

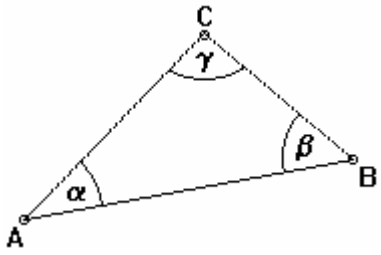
Q_{kk} and Q_{xx} are partial matrices out of $\begin{pmatrix} A^TPA & C \\ C^T & 0 \end{pmatrix}^{-1}$

$$(A^TPA)^{-1} = N^{-1}$$
$$k = (C^TN^{-1}C)^{-1}\,C^TN^{-1}A^TPl + const$$
$$Q_{kk} = (C^TN^{-1}C)^{-1}\,C^TN^{-1}A^TPP^{-1}PAN^{-1}C(C^TN^{-1}C)^{-1}$$
$$Q_{kk} = (C^TN^{-1}C)^{-1}$$
$$Q_{xx} = N^{-1} - N^{-1}C(C^TN^{-1}C)^{-1}C^TN^{-1}$$

Calculation of v^tPv:

$$v^TPv = v^TP(A\Delta x - l) = v^TPA\Delta x - v^TPl$$
$$v^TPl = l^TPv = l^TP(A\Delta x - l) = l^TPA\Delta x - l^TPl$$
$$v^TPv = l^TPl - l^TPA\Delta x + \Delta x^TA^TP(A\Delta x - l)$$
$$= l^TPl - l^TPA\Delta x + \Delta x^T \underbrace{A^TPA\Delta x}_{A^TPl - Ck} - \Delta x^TA^TPl$$
$$= l^TPl - l^TPA\Delta x + \cancel{\Delta x^TA^TPl} - \Delta x^TCk - \cancel{\Delta x^TA^TPl}$$
$$= l^TPl - l^TPA\Delta x - \Delta x^TCk$$
$$v^TPv = l^TPl - l^TPA\Delta x - c^Tk$$
$$(v^TPv = l^TPl - l^TPA\Delta x \qquad \text{valid for parametric adjustment})$$

Example:



observations:

α = 45,02 gon $\qquad v_\alpha = x - \alpha$

β = 55,03 gon $\qquad v_\beta = y - \beta$

γ = 100,01 gon $\qquad v_\gamma = z - \gamma$

$$x + y + z = 200 \text{ gon}$$

linearization: $x_0 = 45^g$, $y_0 = 55^g$, $z_0 = 100^g$

$$v_\alpha = \Delta x - 2^c$$
$$v_\beta = \Delta y - 3^c$$
$$v_\gamma = \Delta z - 1^c$$
$$\Delta x + \Delta y + \Delta z = 0$$

$$A = \begin{pmatrix} 1 \\ & 1 \\ & & 1 \end{pmatrix} \quad , \qquad C^T = [\,1 \quad 1 \quad 1\,] \qquad , \qquad P = E$$

$$A^T P A = E \ , \ (A^T P A)^{-1} = N^{-1} = E$$

$$3k = 6$$
$$k = 2$$

$$\Delta x = l - Ck = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$$

$$x = 45^g$$
$$y = 55{,}01^g$$
$$z = 99{,}99^g$$

$$v^T P v = 14 - 2 = 12$$

Check:

$$\left.\begin{aligned} v_\alpha &= -2 \\ v_\beta &= -2 \\ v_\gamma &= -2 \end{aligned}\right\} \quad v^T P v = 12$$

5.3.1. Condition equations with unknowns:

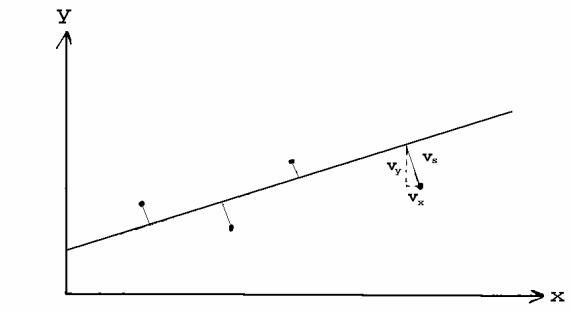
Approach:

$$f(\bar{l}, x) = u$$

(implicit)

$$\bar{l} = g(x)$$

(explicit)



Straight line

$$y = ax + b$$

$$v_s^2 = v_x^2 + v_y^2$$

$$\sum v_{xi}^2 + v_{yi}^2 \Rightarrow \min.$$

Assumption:

- x not random

$$y + v_y = x\underline{a} + \underline{b} \quad \text{(parametric)}$$

- y observed, random
- x,y both variable regarded as random

$$y + v_y = \underline{a}(x + v_x) + \underline{b} \quad \text{(condition adjustment with unknowns)}$$

Linearization of the condition equations at approximate positions l, x_0:

$$\frac{\partial f}{\partial l} v + \frac{\partial f}{\partial x} \Delta x = u - f(l, x_0)$$

$$B^T v + A\Delta x = w$$

Minimization of $v^T pv$:

$$\Phi = v^T pv - 2k^T (B^T v + A\Delta x - w) \Rightarrow \min.$$

$$\frac{\partial \Phi}{\partial v} = 0 \rightarrow 2v^T P - 2k^T B^T = 0 \rightarrow v = P^{-1} Bk \quad \text{(I)}$$

$$\frac{\partial \Phi}{\partial k} = 0 \rightarrow B^T v + A\Delta x - w = 0 \quad \text{(II)}$$

$$\frac{\partial \Phi}{\partial \Delta x} = 0 \rightarrow -2k^T A = 0 \quad \text{(III)}$$

(I) in (II) results into:

$$B^T P^{-1} B k + A\Delta x = w$$

$$A^T k = 0$$

$$\begin{bmatrix} B^T P^{-1} B & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} k \\ \Delta x \end{bmatrix} = \begin{bmatrix} w \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} k \\ \Delta x \end{bmatrix} = \begin{bmatrix} B^T P^{-1} B & A \\ A^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} w \\ 0 \end{bmatrix}$$

Elimination von k:

$$A^T (B^T P^{-1} B)^{-1} A\Delta x = A^T (B^T P^{-1} B)^{-1} w \quad \rightarrow \quad \Delta x$$

$$k = (B^T P^{-1} B)^{-1} (w - A\Delta x)$$

$$v = P^{-1} B k$$

5.3.2. Cofactors of the adjusted observations following the law of error propagation:

$$Q_{\Delta x \Delta x} \; ?$$

$$\text{mit } \Delta x = (A^T (B^T P^{-1} B)^{-1} A)^{-1} A^T (B^T P^{-1} B)^{-1} w$$

$$Q_{\Delta x \Delta x} = (A^T (B^T P^{-1} B)^{-1} A)^{-1} A^T (B^T P^{-1} B)^{-1} \cdot$$

$$Q_{ww} (B^T P^{-1} B)^{-1} A (A^T (B^T P^{-1} B)^{-1} A)^{-1}$$

$$w = u - f(l, x_0)$$

$$w = B^T \Delta l \rightarrow Q_{ww} = B^T Q_{ll} B$$

$$Q_{\Delta x \Delta x} = (A^T (B^T P^{-1} B)^{-1} A)^{-1}$$

5.3.3. Estimation of the variance of unit weight:

$$m_0 = \sqrt{\frac{v^T P v}{r}}$$

r = number of condition equations – number of unknowns

$v^T P v$ with $v = P^{-1} B k$:

$$v^T P v = k^T B^T P^{-1} P P^{-1} B k = k^T B^T P^{-1} B k$$

Due to

$$B^T P^{-1} B k = w - A\Delta x$$

It holds:

$$v^T P v = k^T w - kA\Delta x$$

And observing

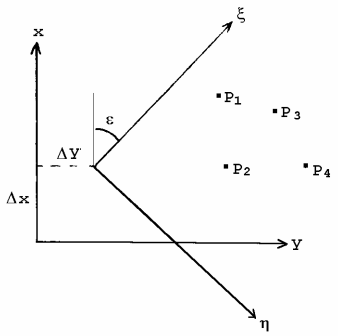
$$A^T k = 0$$

We get:

$$v^T P v = k^T w$$

5.3.4. Helmerttransformation

Determine the parameters of a linear transformation



Given: coordinates x_i,y_i, ξ_i, η_i in a local and global coordinate system

To be calculated: transformation parameters

Approaches for a solution:

$$x_i = \Delta x + m(\xi_i \cos\varepsilon - \eta_i \sin\varepsilon) \quad a = m\cos\varepsilon$$
$$y_i = \Delta y + m(\eta_i \cos\varepsilon + \xi_i \sin\varepsilon) \quad b = m\sin\varepsilon$$

1) x_i, y_i not random, ξ_i, $\eta_i$random:

$$x_i = \Delta x + m(\xi_i + v_{\xi_i})\cos\varepsilon - m(\eta_i + v_{\eta_i})\sin\varepsilon = \Delta x + a(\xi_i + v_{\xi_i}) - b(\eta_i + v_{\eta_i})$$
$$y_i = \Delta y + m(\eta_i + v_{\eta_i})\cos\varepsilon + m(\xi_i + v_{\xi_i})\sin\varepsilon = \Delta y + a(\eta_i + v_{\eta_i}) + b(\xi_i + v_{\xi_i})$$

Type: condition adjustment with unknowns

2) ξ_i, η_i not random, x_i,y_i random:

$$x_i + v_{x_i} = \Delta x + \xi_i a - \eta_i b$$
$$y_i + v_{y_i} = \Delta y + \eta_i a + \xi_i b$$

Type: parametric adjustment

3) x_i, y_i, ξ_i, η_i random

$$x_i + v_{x_i} - (\xi_i + v_{\xi_i})a + (\eta_i + v_{\eta_i})b - \Delta x = 0$$
$$y_i + v_{y_i} - (\eta_i + v_{\eta_i})a - (\xi_i + v_{\xi_i})b - \Delta y = 0$$

Type: condition adjustment with unknowns

Solution for parametric adjustment:

$$v = Ax - l$$

Linear error equations:

$$v = \begin{bmatrix} v_{x_i} \\ v_{y_i} \\ v_{x_2} \\ v_{y_2} \\ \vdots \\ v_{x_n} \\ v_{y_n} \end{bmatrix} \quad A = \left[\begin{array}{cccc} \xi_1 & -\eta_1 & +1 & 0 \\ \eta_1 & \xi_1 & 0 & +1 \\ \hline \xi_2 & -\eta_2 & +1 & 0 \\ \eta_2 & \xi_2 & 0 & +1 \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \xi_n & -\eta_n & +1 & 0 \\ \eta_n & \xi_n & 0 & +1 \end{array}\right] \quad x = \begin{bmatrix} a \\ b \\ \Delta x \\ \Delta y \end{bmatrix} \quad l = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \vdots \\ x_n \\ y_n \end{bmatrix}$$

Normal equations: P = E

$$A^T P A x = A^T P l$$

$$A^T P A = \left[\begin{array}{cc|cc} [\xi^2 + \eta^2] & 0 & [\xi] & [\eta] \\ & & (N_{12}) & \\ (N_{11}) & [\xi^2 + \eta^2] & -[\eta] & [\xi] \\ \hline & & n & 0 \\ & & (N_{22}) & \\ & & & n \end{array}\right]$$

$$A^T P l = \left[\begin{array}{c} [\xi_x + \eta_y] \\ (h_1) \\ [\xi_y - \eta_x] \\ \hline [x] \\ (h_2) \\ [y] \end{array}\right]$$

Subdivision of x into:

$$\begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix} \text{ with } \quad x_1 = \begin{bmatrix} a \\ b \end{bmatrix}, \quad x_2 = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

Then NGL-system:

$$N_{11}x_1 + N_{12}x_2 = h_1$$
$$N_{12}^T x_1 + N_{22}x_2 = h_2$$

Elimination of x_2:

$$(N_{11} - N_{12}N_{22}^{-1}N_{12}^T)x_1 = h_1 - N_{12}N_{22}^{-1}h_2$$

$$N_{22}^{-1} = \frac{1}{n}E \; ; \quad N_{12}N_{22}^{-1}N_{12}^T = \frac{1}{n}\begin{bmatrix} [\xi]^2 + [\eta]^2 & 0 \\ 0 & [\xi]^2 + [\eta]^2 \end{bmatrix}$$

Reduced NGL-system:

$$\{[\xi^2 + \eta^2] - \frac{1}{n}([\xi]^2 + [\eta]^2)\}E\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} [\xi_x + \eta_y] - \frac{1}{n}([\xi][x] + [\eta][y]) \\ [\xi_y - \eta_x] - \frac{1}{n}([\xi][y] - [\eta][x]) \end{bmatrix}$$

Then calculation of a,b:

$$x_2 = N_{22}^{-1}(h_2 - N_{12}^T x_1)$$
$$\Delta x = \frac{1}{n}([x] - [\xi]a + [\eta]b)$$
$$\Delta y = \frac{1}{n}([y] - [\eta]a - [\xi]b)$$

<u>Transition to gravity system:</u>

Taking into account that:

$$A^T P v = 0 \qquad \begin{bmatrix} [vx] \\ [vy] \end{bmatrix} = 0$$

Therefore approach:

$$x_{s_i} + v_{x_i} = \xi s_i a - \eta s_i b$$
$$y_{s_i} + v_{y_i} = \eta s_i a + \xi s_i b$$

$$x_{s_i} = x_i - \frac{[x]}{n} \qquad \xi_{s_i} = \xi_i - \frac{[\xi]}{n}$$
$$y_{s_i} = y_i - \frac{[y]}{n} \qquad \eta_{s_i} = \eta_i - \frac{[\eta]}{n}$$

$$[x_s] = 0 \ , \ [y_s] = 0 \ , \ [\xi_s] = 0 \ , \ [\eta_s] = 0$$

Then it holds:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{[\xi_s^2 + \eta_s^2]} \cdot \begin{bmatrix} [\xi_s x_s + \eta_s y_s] \\ [\xi_s y_s - \eta_s x_s] \end{bmatrix}$$

$$\Delta x = x_s - \xi_s a - \eta_s b$$
$$\Delta y = y_s - \eta_s a - \xi_s b$$

Redundancy: 2n - 4

Controls:

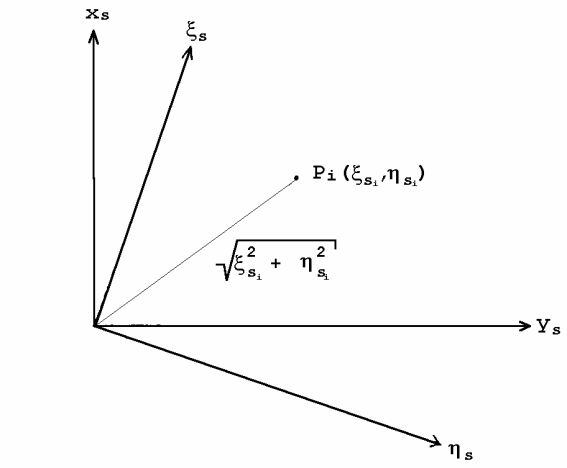
$$[v_x] = 0 \ , \ [v_y] = 0$$
$$[v_x \xi_s] + [v_y \eta_s] = 0$$
$$[v_x \eta_s] + [v_y \xi_s] = 0$$

$$m_0^2 = \frac{v^T P v}{2n - 4} \ ; \ v^T P v = l^T P l - x^T A^T P l$$
$$v^T P v = [x^2] + [y^2] - a([\xi_s x_s] + [\eta_s y_s]) - b([\xi_s y_s] - [\eta_s x_s])$$

Accuracy:

$$m_a^2 = m_0^2 \frac{1}{[\xi_s^2 + \eta_s^2]} = m_b^2$$

$$m_a^2 = m_0^2 \frac{1}{[s_i^2]} = m_b^2 \quad s_i : \text{distance from center of gravity}$$



distance from centre of gravity

$$d(\Delta x) = -\xi_s da + \eta_s db$$
$$d(\Delta y) = -\eta_s da - \xi_s db$$

$$Q_{\Delta x \Delta x} = (\xi_s^2 + \eta_s^2) Q_{aa} \;\; ; \;\; m_{\Delta x}^2 = m_0^2 Q_{\Delta x \Delta x}$$
$$Q_{\Delta y \Delta y} = (x_s^2 + h_s^2) Q_{aa} \;\; ; \;\; m_{\Delta x}^2 = m_0^2 Q_{\Delta x \Delta x}$$
$$Q_{\Delta x \Delta y} = 0$$
$$Q_{vv} = Q_{ll} - A Q_{xx} A^T$$
$$Q_{vv_{ii}} = 1 - \frac{s_i^2}{[s_i^2]}$$

6. Binomial distribution and normal distribution

6.1. Binomial distribution

$$r = \frac{N_r}{N} \; ; \; w = \frac{N_w}{N}$$

Question: What is the probability that, based on n draws and placing back, x times the ball will be red and (n-x)-times the ball will be white?

Preassumption: Every ball has the same chance to be drawn!

$N = N_r + N_w$ = Total number of balls
N_r = Number of red balls
N_w = Number of white balls

Probability to draw a red ball: $r = \frac{N_r}{N}$

Probability to draw a white ball: $w = 1 - r$

Random variable: X (discrete, $0 \le x \le n$)

One possibility to get x times red with n trials:

$$\underbrace{RRR\ldots R}_{x} \quad \underbrace{WWW\ldots W}_{n-x}$$

The probability for this possibility - with the assumption of stochastically independent draws:

$$\underbrace{r \cdot r \cdot r \ldots r}_{r^x} \cdot \underbrace{w \cdot w \cdot w \ldots w}_{(1-r)^{n-x}} = r^x \cdot (1-r)^{n-x}$$

Probabilistic density of random variable X:

$$f(x) = A \cdot r^x \cdot (1-r)^{n-x}$$

A = Number of all possibilities, where x-times R and (n-x) times W may occur.

= Number of permutations of n elements if $\alpha_1 = x$ elements and $\alpha_2 = (n-x)$ elements are equal.

General formula: $A = \frac{n!}{\alpha_1 ! \cdot \alpha_2 !}$; here $A = \frac{n!}{x! \cdot (n-x)!} = \binom{n}{x}$

Probabilistic density of random variable X (Binomial distribution):

$$f(x) = \binom{n}{x} \cdot r^x \cdot (1-r)^{n-x}$$

Recursion formula:

$$\begin{aligned} f(x+1) &= \binom{n}{x+1} \cdot r^{x+1} \cdot (1-r)^{n-x-1} \\ &= \frac{n!}{(x+1)!\cdot(n-x-1)!} \cdot r^{x+1} \cdot (1-r)^{n-x-1} \\ &= \frac{n!\cdot(n-x)}{x!\cdot(n-x)!\cdot(x+1)} \cdot r^x \cdot (1-r)^{n-x} \cdot \frac{r}{1-r} \\ &= \binom{n}{x} \cdot r^x \cdot (1-r)^{n-x} \cdot \frac{n-x}{x+1} \cdot \frac{r}{1-r} \end{aligned}$$

$$f(x+1) = f(x) \cdot \frac{n-x}{x+1} \cdot \frac{r}{1-r}$$

Expectation value of random variable X:

$$\xi = \sum x f(x)$$

From recursion formula we get:

$$(1-r)\cdot(x+1)\cdot f(x+1) = rn\cdot f(x) - rx\cdot f(x)$$

Summing up, in order to take all possibilities under consideration (i.e. from -1 till n):

$$(1-r)\underbrace{\sum (x+1) f(x+1)}_{\xi} = rn\underbrace{\sum f(x)}_{1} - \underbrace{\sum x f(x)}_{\xi}$$

$$\begin{aligned} (1-r)\cdot\xi + r\cdot\xi &= rn \\ \xi &= rn \end{aligned} \qquad \text{(Expectation value)}$$

Variance:

General $\quad \sigma^2 = E(X^2) - \{E(X)\}^2$

Here $\quad \sigma_x^2 = \sum x^2 f(x) - \xi^2$

Calculation of $\sum x^2 f(x)$ applying recursion formula:

$$(1-r)\cdot(x+1)\cdot f(x+1) = rn\cdot f(x) - rx\cdot f(x)$$

Multiply by (x+1)

$$(1-r)\cdot(x+1)^2\cdot f(x+1) = (x+1)\cdot\{rn\cdot f(x) - rx\cdot f(x)\}$$
$$= \underbrace{rn}_{\xi}\cdot x\cdot f(x) + \underbrace{rn}_{\xi}\cdot f(x) - r\cdot x^2\cdot f(x) - r\cdot x\cdot f(x)$$

Summing up, in order to consider all possibilities:

$$(1-r)\underbrace{\sum(x+1)^2 f(x+1)}_{\sum x^2 f(x)} = \xi\underbrace{\sum xf(x)}_{\xi} + \xi\underbrace{\sum f(x)}_{1} - r\sum x^2 f(x) - r\underbrace{\sum xf(x)}_{\xi}$$

$$(1-r)\sum x^2 f(x) = \xi^2 + \xi - r\xi - r\sum x^2 f(x)$$
$$\sum x^2 f(x) = \xi^2 + \xi - r\xi$$

$$\sigma_x^2 = \sum x^2 f(x) - \xi^2$$
$$= \xi^2 + \xi - r\xi - \xi^2$$

$$\sigma_x^2 = \xi\cdot(1-r) = nr\cdot(1-r) \quad \text{(Variance of random variable X)}$$

Binomial distribution can be favourable for evaluation the sign of a resulting sample.

Example: Sign of independent differences of observations.

Model	Model of observational differences
Red ball	positive difference
White ball	negative difference
n = number of drawings	n = number of differences
x = number of red balls	x = number of positive differences
r = probability of getting a red ball	r = probability of getting a positive difference
(1-r) = probability of getting a white ball	(1-r) = probability of getting a negative difference

Assumption: The probability of getting a positive difference is equal to the probability of getting a negative difference

r = p (+) = 0,5 ; (1 - r) = p (-) = 0,5

a) n = 5 observational differences

$$\xi = 0{,}5 \cdot 5 = 2{,}5$$
$$\sigma_x^2 = 5 \cdot 0{,}5 \cdot 0{,}5 = 1{,}25$$
$$\sigma_x = \pm 1{,}12$$

$$P(x = 0) = f(0) = \binom{5}{0} \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^5 = 0{,}0312_5 = P(x = 5)$$
$$P(x = 1) = f(1) = \binom{5}{1} \cdot \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^4 = 0{,}1562_5 = P(x = 4)$$
$$P(x = 2) = f(2) = \binom{5}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^3 = 0{,}3125 = P(x = 3)$$

Applying the recursion formula:

$$P(x = 1) = f(1) = \frac{5}{1} \cdot \frac{0{,}5}{0{,}5} \cdot f(0) = 0{,}1562_5$$
$$P(x = 2) = f(2) = \frac{4}{2} \cdot \frac{0{,}5}{0{,}5} \cdot f(1) = 0{,}3125$$

$$P(x \le 1) = P(x = 0) + P(x = 1) = 0{,}1875$$

b) n = 35

$$\xi = 0{,}5 \cdot 25 = 12{,}5$$
$$\sigma_x^2 = 25 \cdot 0{,}5 \cdot 0{,}5 = 6{,}25$$
$$\sigma_x = \pm 2{,}5$$

$P(x = 0) = 0{,}0000 \quad P(x = 3) = 0{,}0001$

$P(x = 1) = 0{,}0000 \quad P(x = 4) = 0{,}0004$

$P(x = 2) = 0{,}0000 \quad P(x = 5) = 0{,}0016$

$$P(x \le 5) = \sum_{i=0}^{5} P(x = i) = 0{,}0021$$

6.2. Normal distribution

6.2.1. General problem

The probability density function of independent observations l_i is searched for as well as its distributive function.

Random variables of the observations l_i:

$$L \text{ with } E(L) = \lambda$$

Or random variable of the ε_i-values:

$$\varepsilon = \lambda - L \text{ with } E(\varepsilon) = 0$$

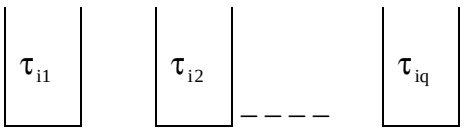
Deviations from the expectation value; "true " error

1. Model assumption

The deviation ε_i of the i-th observation l_i is seen as sum of different independent random variables τ_{iv} :

$$\varepsilon_i = \sum_{v=1}^{q} \tau_{iv}$$

Visualisation of assumption of the 1. Model



From each of the pots a partial value τ_{iv} is randomly drawn, ε_i is the sum of the individual partial values
The distributions of the values τ_{iv} in the in the pots are unknown.
Example: angular measurements

τ_{i1} ... reading error
τ_{i2} ... targeting error
⋮
τ_{iq} ... unequal scale

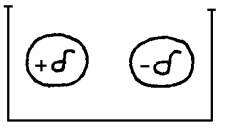
} of the i-th observation

6.2.2. Hagen's hypothesis of elementary errors

2. Model assumption

Hagen's hypothesis (1837): All $\tau_{i\nu}$ consist of elementary errors of equal size with positive and negative sign $+\delta$ and $-\delta$, where both signs are of equal probability. The probabilistic density function $\varphi(\varepsilon)$ can be derived for the boundary case when the number of elementary errors to be summed up is infinite $n \to \infty$ and the size of the elementary errors approaches 0, $|\delta| \to 0$.

Visualisation of the 2. model assumption (Hagen's hypothesis)



$p(+\delta) = 0{,}5$
$p(-\delta) = 0{,}5$

Out of the pot values are drawn n – times with replacement: ($n \to \infty$)
ε_i is the sum of $+\delta$ and $-\delta$-values of all n-drawings in order to get for the i-th measurement.

$$\varepsilon_i = x_i(+\delta) + (n - x_i) \cdot (-\delta) = 2\delta x_i - n\delta$$

x_i = number of drawings where $+\delta$ resulted $(\delta \to 0)$.

6.2.3. Normal distribution as boundary case of the binomial distribution

The model assumption of Hagen corresponds to the binomial distribution:

Red ball $\to +\delta$; white ball $\to -\delta$

$r = 0{,}5$ $w = 0{,}5$

Random variable X = number of draws, where a red ball resulted $\to +\delta$.

Expectation value and variance of the random variable X:

$$\xi = rn = \frac{n}{2}$$

$$\sigma_x^2 = \xi \cdot (1 - r) = \frac{n}{4}$$

The random variable X is binomially distributed with the density function:

$$f(x) = \binom{n}{x} r^x (1 - r)^{n-x}$$

The random variable ε is a linear function of the discrete random X

$$\varepsilon = 2\delta X - n\delta$$

Expectation value and variance of random variable ε

$$E(\varepsilon) = 2\delta\xi - n\delta = 2\delta\frac{n}{2} - n\delta = 0$$

$$\sigma_\varepsilon^2 = \sigma^2 = 4\delta^2\sigma_x^2 \quad \text{(according to error propagation)}$$

$$\sigma_x^2 = \frac{n}{4}$$

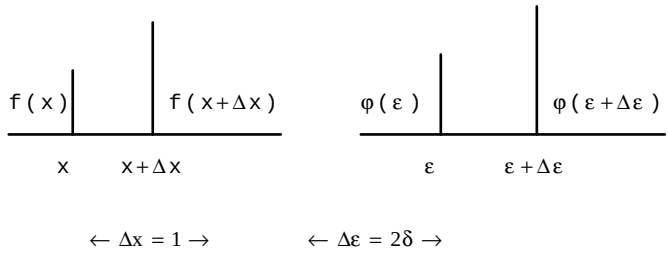
$$\sigma^2 = n\delta^2$$

Transition from density function f(x) to density function $\varphi(\varepsilon)$:

$$f(x)\cdot\Delta x = \varphi(\varepsilon)\cdot\Delta\varepsilon$$

This equation holds for monotone functions fulfilling the requirement of equal differential probability dP.

Graphical visualisation of probabilistic densities:



Change of x by $\Delta x = +1$, changes ε by $+2\delta$.

$$f(x)\cdot\underbrace{\Delta x}_{1} = \varphi(\varepsilon)\cdot\underbrace{\Delta\varepsilon}_{2\delta}$$

$$f(x) = \varphi(\varepsilon)\cdot 2\delta$$

$$\varphi(\varepsilon) = \frac{f(x)}{2\delta}$$

$$\varphi(\varepsilon+\Delta\varepsilon) = \frac{f(x+1)}{2\delta}$$

f(x+1) will be removed applying the recursion formula:

$$f(x+1) = \frac{n-x}{x+1}\cdot\frac{r}{\underbrace{1-r}_{=1}}\cdot f(x)$$

$$\varphi(\varepsilon+\Delta\varepsilon) - \varphi(\varepsilon) = \left(\frac{n-x}{x+1} - 1\right)\cdot\underbrace{\frac{f(x)}{2\delta}}_{\varphi(\varepsilon)}$$

$$\varphi(\varepsilon+\Delta\varepsilon) - \varphi(\varepsilon) = \frac{n-2x-1}{x+1}\cdot\varphi(\varepsilon)$$

x will be replaced by the functional dependency of ε, δ and n replaced:

$$\varepsilon = 2\delta x - n\delta \;\rightarrow\; x = \frac{\varepsilon}{2\delta} + \frac{n}{2}$$

$$\varphi(\varepsilon + \Delta\varepsilon) - \varphi(\varepsilon) = \frac{n - \frac{\varepsilon}{\delta} - n - 1}{\frac{\varepsilon}{2\delta} + \frac{n}{2} + 1} \cdot \varphi(\varepsilon)$$

$$= \frac{-2\varepsilon - 2\delta}{\varepsilon + n\delta + 2\delta} \cdot \varphi(\varepsilon)$$

Division by $\Delta\varepsilon = 2\delta$

$$\frac{\varphi(\varepsilon + \Delta\varepsilon) - \varphi(\varepsilon)}{\Delta\varepsilon} = \frac{-2\varepsilon - 2\delta}{2\varepsilon\delta + \underbrace{2n\delta^2}_{2\sigma^2} + 4\delta^2} \cdot \varphi(\varepsilon)$$

Transition from the discrete to the smooth functional distribution $\varphi(\varepsilon)$
The boundary transition $n \rightarrow \infty;\ \delta \rightarrow 0$ corresponds to the transition from the difference equation to the differential equation:

$$\frac{d\varphi(\varepsilon)}{d\varepsilon} = -\frac{\varepsilon}{\sigma^2} \cdot \varphi(\varepsilon)$$

Remark: σ^2 may be first regarded as boundary of the variance of the binomial distribution.

$$\frac{d\varphi(\varepsilon)}{\varphi(\varepsilon)} = -\frac{1}{\sigma^2}\varepsilon \cdot d\varepsilon$$

Integration:

$$\ln\varphi(\varepsilon) = -\frac{1}{\sigma^2} \cdot \frac{\varepsilon^2}{2} + c_1$$

$$\varphi(\varepsilon) = c_2 \cdot e^{-\frac{\varepsilon^2}{2\sigma^2}}$$

Integration constant c_2 is to be determined in such a way, that:

$$\int_{-\infty}^{+\infty} \varphi(\varepsilon)d\varepsilon = 1 = C_2 \int_{-\infty}^{+\infty} e^{\frac{-\varepsilon^2}{2\sigma^2}} d\varepsilon$$

Generally:

$$\int_{-\infty}^{+\infty} e^{-ax^2} dx = 2\int_{-\infty}^{+\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}} \;;\; a > 0$$

Here:

$$a = \frac{1}{2\sigma^2} \quad ; \quad \sqrt{\frac{\pi}{a}} = \sqrt{2\pi\sigma^2}$$

$$1 = C_2 \cdot \sqrt{2\pi} \cdot \sigma \quad \rightarrow \quad C_2 = \frac{1}{\sigma\sqrt{2\pi}}$$

Probabilistic density function of the random variable ε
(Gauß's law of error; Normal distribution N (0, σ)):

$$\varphi(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-\varepsilon^2}{2\sigma^2}}$$

From the definition of the variance it follows, that variance σ_ε^2 of the random variable ε equals the boundary value σ^2 for the normal distribution.

$$\sigma_\varepsilon^2 = E(\varepsilon^2) - \{E(\varepsilon)\}^2 = \int_{-\infty}^{+\infty} \varepsilon^2 \varphi(\varepsilon) \cdot d\varepsilon = \sigma^2 \; ; \; \{E(\varepsilon)\}^2 = 0$$

Probabilistic density function of the continual random variable L (normal distribution N (λ, σ)):

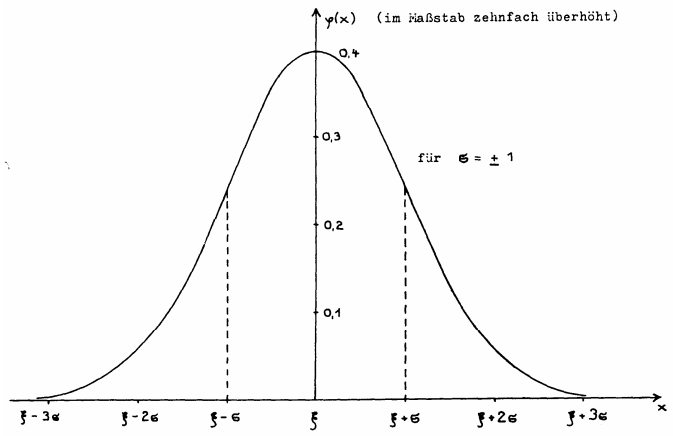
$$\varphi(l) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\lambda - l)^2}{2\sigma^2}}$$

Probabilistic density function of any random variable X, which fulfils the model assumptions (normal distribution N (ξ, σ_x):

$$\varphi(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{\frac{-(\xi - x)^2}{2\sigma_x^2}}$$

Remark: In statistics literature $(x - \xi)^2$ will often be used instead of $(\xi - x)^2$. The density function of a normally distributed random variable only depends on the two parameters ξ and σ.
ξ, the positional parameter does not influence the shape of the density function.

The probabilistic density function can be visualised as a symmetrical bell shaped graph.
Maximum $x = \xi$



Turning point of the curve:

$$\varphi(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\varepsilon^2}{2\sigma^2}}$$

$$\varphi'(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\varepsilon^2}{2\sigma^2}} \cdot \left(-\frac{\varepsilon}{\sigma^2}\right) = -\frac{\varepsilon}{\sigma^3\sqrt{2\pi}} e^{\frac{-\varepsilon^2}{2\sigma^2}}$$

$$\varphi''(\varepsilon) = -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{\frac{-\varepsilon^2}{2\sigma^2}} \cdot 1 + \varepsilon \cdot e^{\frac{-\varepsilon^2}{2\sigma^2}}\left(-\frac{\varepsilon}{\sigma^2}\right)\right]$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}} e^{\frac{-\varepsilon^2}{2\sigma^2}}\left(1 - \frac{\varepsilon^2}{\sigma^2}\right)$$

Turning points: curvature = 0 ; i.e. $\varphi''(\varepsilon_w) = 0$

$$1 - \frac{\varepsilon_w^2}{\sigma^2} = 0$$

$$\varepsilon_w = \pm\sigma$$

The coordinates ε_w of the turning points for $N(\xi, \sigma)$ are symmetrically to ξ at $\pm\sigma$.

6.2.4. The Normalized Normal Distribution N(0,1)

6.2.4.1. Normalization of a random variable for any distribution

$$u = \frac{x - \xi}{\sigma}$$ (normalized random variable)

$$E(u) = 0$$

$$\sigma_u = \pm 1$$

Every normalised distribution has the expectation value 0 and the variance 1.

Relationships between the probabilistic density functions f(x) and f(u) (based on equal differential probability dP):

$$f(u) \cdot du = f(x) \cdot dx = dP$$

$$du = \frac{1}{\sigma} \cdot dx$$

$$f(u) = \sigma \cdot f(x)$$

$$f(x) = \frac{1}{\sigma} \cdot f(u)$$

6.2.4.2. Normalized Normal Distribution N(0,1)

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\xi - x)^2}{2\sigma^2}} \quad \text{für } N(\xi, \sigma)$$

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} \quad \text{für } N(0,1)$$

The density function $\varphi(u)$ of the normalised normal distribution is available in tabular form.

Example for calculating $\varphi(x)$ using $\varphi(u)$:

$$\xi = 125{,}723 \text{ m} \; ; \; x = 125{,}728 \text{ m} \; ; \; \sigma = \pm 2{,}0 \text{ mm}$$

$$u = \frac{125728 - 125723}{2{,}0} = 2{,}5$$

$$\varphi(u) = 0{,}0175$$

$$\varphi(x) = \frac{1}{\sigma}\varphi(u) = \frac{1}{2} \cdot \varphi(u) = 0{,}0088$$

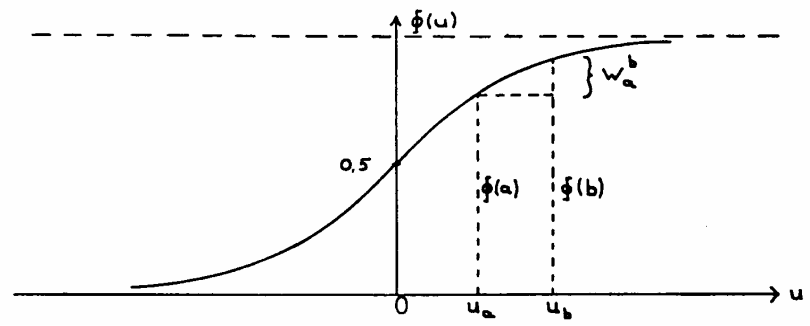
6.2.5. The Distributive Function and the Probability Function of Errors.

Distributive function:

$$\text{für} N(\xi,\sigma) \qquad \Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(\xi-t)^2}{2\sigma^2}} \cdot dt$$

$$\text{für } N(0,\sigma) \qquad \Phi(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\varepsilon} e^{\frac{-t^2}{2\sigma^2}} \cdot dt$$

$$\text{für } N(0,1) \qquad \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} \cdot dt$$



The values $\Phi(u)$ of the normalised normal distribution are available in tabular form:

The function of the probability of errors W_a^b is defined as probability that the error is within the interval a – b.

$$W_a^b = \Phi(b) - \Phi(a)$$

Examples:

1.) If events show the normal distribution $N(\xi,\sigma)$, then theoretically 15,87 % are to be expected to show up to be smaller than $\xi - \sigma$.

2.) $W_{-\sigma}^{+\sigma} = \Phi(+\sigma) - \Phi(-\sigma) = \Phi(u=+1) - \Phi(u=-1) = 0{,}8413 - 0{,}1587 = 0{,}6826$

In the range $\xi - \sigma \ldots\ldots \xi + \sigma$ theoretically 68,3 % of all events are to be expected.

3.) $W_{-3\sigma}^{+3\sigma} = 0{,}9987 - 0{,}0013 = 0{,}9974$

Only less then 0,3 % of all events will be theoretically outside the range $\xi - 3\sigma \ldots\ldots \xi + 3\sigma$ ("maximal error").

All conclusions stated above are only valid if the events are normal distributed $N(\xi,\sigma)$ and σ is known in advance.

6.2.6. The Distribution of Functions of Normal Distributed Random Variables.

6.2.6.1. Linear functions

Let x_ν be independent, $N(\xi_\nu, \sigma_\nu)$ distributed random variables.

Then the linear function

$$X = k_0 + k_1 x_1 + k_2 X_2 + k_3 X_3 + \ldots\ldots + k_n X_n = k_0 + \sum_{\nu=1}^{n} k_\nu \cdot X_\nu$$

(For real constant values k_ν)

is also normal distributed with

$$\begin{aligned}\xi &= E(X) = k_0 + k_1 E(X_1) + k_2 E(X_2) + \ldots + k_n E(X_n) \\ &= k_0 + \sum_{\nu=1}^{n} k_\nu E(X_\nu) = k_0 + \sum_{\nu=1}^{n} k_\nu \cdot \xi_\nu \\ \sigma_x^2 &= k_1^2 \sigma_1^2 + k_2^2 \sigma_2^2 + \ldots + k_n^2 \sigma_n^2 = \sum_{\nu=1}^{n} k_\nu^2 \sigma_\nu^2\end{aligned}$$

(Proof of this addition theorem of the normal distribution in statistic literature.)

Example:

$$X = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \ldots + \frac{1}{n} X_n \quad \text{(arithmetic mean)}$$

$$\begin{aligned}\xi_1 &= \xi_2 = \ldots = \xi_n \\ \sigma_1 &= \sigma_2 = \ldots = \sigma_n \\ E(X) &= \sum_{i=1}^{n} \frac{1}{n} \xi_i = \xi \\ \sigma_x^2 &= \sum_{i=1}^{n} \frac{1}{n^2} \sigma_i^2 = \frac{\sigma^2}{n}\end{aligned}$$

The arithmetic mean of n independent $N(\xi, \sigma)$ distributed observations has the distribution $N(\xi, \frac{\sigma}{\sqrt{n}})$.

6.2.6.2. Nonlinear Functions

For nonlinear functions of normal distributed random variables the specific distributions have to be determined.

Functions of the kind:

$$F = f(l_1 + \varepsilon_1, l_2 + \varepsilon_2, \ldots, l_n + \varepsilon_n)$$

With $E(F) = \Phi$ may in most cases be linearised with sufficient accuracy.

$$F_l = \Phi + \left(\frac{\delta f}{\delta l_1}\right)_0 \varepsilon_1 + \left(\frac{\delta f}{\delta l_2}\right)_0 \varepsilon_2 + \ldots + \left(\frac{\delta f}{\delta l_n}\right)_0 \varepsilon_n$$

F_l therefore is approximately normal distributed.

Functions of the kind:

$$F = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2$$

can not be normal distributed as $F_{min} = 0$.

Examples:

$$F_1 = m^2 = \frac{[\varepsilon\varepsilon]}{n}$$

$$F_2 = \frac{[pvv]}{n-u}$$

$$F_3 = \frac{x}{m_x} = \frac{\frac{[l]}{n}}{\sqrt{\frac{[vv]}{n(n-1)}}}$$

Some of the most essential distributions of nonlinear functions of normal distributed random variables are t-distribution; χ^2-distribution; F-distribution.

Central boundary statement:

Under certain pre-assumptions which may be very weak, any sum of independent variables τ_{iv} of any distribution

$\varepsilon_i = \sum_{v=1}^{n} \tau_{iv}$ for $n \to \infty$ becomes normal distributed.

Pre-assumptions:

a) None of the variables τ shall contribute too much to the sum compared to the others.
b) The distributive function of the τ must approach 0 or 1 respectively sufficiently fast for $-\infty$ or $+\infty$.

Let x_ν be independent random variables with expectation value ξ_ν and variance σ_ν^2. Then the following holds:

a) the sum of x_ν within the range of $n \to \infty$
will be normal distributed with the expectation value $(\xi_1 + \xi_2 + ... + \xi_\nu)$ and
the variance $(\sigma_1^2 + \sigma_2^2 + ... + \sigma_\nu^2)$

b) the normalised sum of the random variables x_ν

$$\frac{\sum_{\nu=1}^{n}(x_\nu - \xi_\nu)}{\sqrt{\sum_{\nu=1}^{n}\sigma_\nu^2}}$$

in the range of $n \to \infty$ will be normal distributed with expectation value 0 and variance 1.

In practise it can be shown that a sum of a limited number n of random variables will be normal distributed with sufficient accuracy. As rule of thumb $n = 30$ is valid. For symmetric distributions of $x_\nu, n \geq 5$ will be sufficient.

From the fact that the sum of random variables or the arithmetic mean of n random variables show normal distribution it can not be concluded that the individual variables are normal distributed.

6.2.7. Principle of Maximum Likelihood Estimation

(1. Gauß's argument for the method of least squares)

$\varepsilon_1 = \xi - l_1$	Probabilistic density:	$\varphi_1(\varepsilon_1)$
$\varepsilon_2 = \xi - l_2$		$\varphi_2(\varepsilon_2)$
.		.
.		.
.		.
$\varepsilon_n = \xi - l_n$		$\varphi_n(\varepsilon_n)$

Probability for the simultaneous existence of the independent values $\varepsilon_1, \varepsilon_2 \ldots \varepsilon_n$:

$$d\Phi = \varphi_1(\varepsilon_1)d\varepsilon \cdot \varphi_2(\varepsilon_2)d\varepsilon \cdot ... \cdot \varphi_n(\varepsilon_n)d\varepsilon$$
$$d\Phi = \varphi_1(\varepsilon_1) \cdot \varphi_2(\varepsilon_2) \cdot ... \cdot \varphi_n(\varepsilon_n) \cdot (d\varepsilon)^n$$

Assumptions:
All ε_i date from the same normal distributed sample, the parameters ξ and σ however are unknown. For ξ a value should be estimated in such a way that $d\Phi$ will become maximal.

$$L = \varphi(\varepsilon_1) \cdot \varphi(\varepsilon_2) \cdot ... \cdot \varphi(\varepsilon_n) \qquad \text{(Likelihood-Function)}$$

$$L = \left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right)^n \cdot e^{-\frac{\varepsilon_1^2}{2\sigma^2}} \cdot e^{-\frac{\varepsilon_2^2}{2\sigma^2}} \cdot \ldots \cdot e^{-\frac{\varepsilon_n^2}{2\sigma^2}}$$

$$L = \left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right)^n \cdot e^{-\left[\frac{\varepsilon_i^2}{2\sigma^2}\right]} = \left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\left[(\xi - l_i)^2\right]}$$

For $L \to L_{max}$ $\xi \to x$ holds:

$$L_{max} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\left[(x - l_i)^2\right]}$$

$$\text{i.e...} \left[(x - l_i)^2\right] = [vv] = \text{Minimum}$$

This is the 1.Gauß's argument in favour of the method of least squares. It is based on normal distributed observations.

R. A. Fischer generalised the estimation:

$$L = f(l_1; \theta_1, \theta_2 \ldots) \cdot f(l_2; \theta_1, \theta_2 \ldots) \cdot \ldots \cdot f(l_n; \theta_1, \theta_2 \ldots)$$

The unknown parameters $\theta_1, \theta_2 \ldots$ of the density functions $f(l_i; \theta_1, \theta_2 \ldots)$ should be determined in such a way that the Likelihood-Function L results in maximal value L_{max}.

$$\left.\begin{aligned} \frac{\delta L}{\delta \theta_1} &= 0 \\ \frac{\delta L}{\delta \theta_2} &= 0 \\ &\vdots \end{aligned}\right\}$$

from these equations the values of the parameters $\xi - u_s \cdot \sigma_x \le x \le \xi + u_s \cdot \sigma_x$ are to be calculated

The following logarithmic function holds:

$$\ln L = \ln f(l_1; \theta_1, \theta_2 \ldots) + \ldots + \ln f(l_n; \theta_1, \theta_2 \ldots)$$

$$\frac{\delta \ln L}{\delta \theta_i} = 0$$

Likelihood-estimations do not have to be unbiased, however they have to be asymptotically normal distributed. For the case of normal distribution the estimation of the variance σ^2 results in:

$$\overline{m}^2 = \frac{[vv]}{n} = \frac{n-1}{n} \cdot \frac{[vv]}{n-1}$$

$$E(\overline{m}^2) = \frac{n-1}{n} \cdot E(m^2) = (1 - \frac{1}{n}) \cdot E(m^2)$$

This estimation is not unbiased, but for $n \to \infty$ $\overline{m}^2 \to \sigma^2$ holds...

6.2.8. Other Derivations of the Normal Distribution

6.2.8.1. Gauß's law of errors

$$L = \varphi(\varepsilon_1) \cdot \varphi(\varepsilon_2) \cdot \ldots \cdot \varphi(\varepsilon_n)$$

ξ is unknown, the estimation x should be used for which the Likelihood-Function will become maximal.

$$\begin{aligned} v_1 &= x - l_1 \\ v_2 &= x - l_2 \\ &\vdots \\ v_n &= x - l_n \end{aligned}$$

$$\varphi(v_1) \cdot \varphi(v_2) \cdot \ldots \cdot \varphi(v_n) = \text{Maximum}$$

$$\ln\varphi(v_1) + \ln\varphi(v_2) + \ldots + \ln\varphi(v_n) = \text{Maximum}$$

$$\frac{d\ln\varphi(v_1)}{dv_1} \cdot \underbrace{\frac{dv_1}{dx}}_{1} + \frac{d\ln\varphi(v_2)}{dv_2} \cdot \underbrace{\frac{dv_2}{dx}}_{1} + \ldots = 0$$

$$\frac{d\ln\varphi(v_1)}{dv_1} + \frac{d\ln\varphi(v_2)}{dv_2} + \ldots = 0 = \sum \frac{d\ln\varphi(v_i)}{dv_i} = \sum \frac{v_i \cdot d\ln\varphi(v_i)}{v_i \cdot dv_i}$$

(Multiplying numerator and denominator by v_i).

Gauß's assumption: The arithmetic mean is the most probable value:

$$x = \frac{l_1 + l_2 + \ldots + l_n}{n}$$

This assumption corresponds to:

$$\sum v = 0$$

In addition it holds:

$$\sum v \cdot \frac{d\ln\varphi(v)}{v \cdot dv} = 0$$

Both equations can be valid without contradiction.

$$\frac{d\ln\varphi(v_i)}{v_i \cdot dv_i} = \text{const}$$

$$\frac{d\ln\varphi(v_i)}{dv_i} = k \cdot v_i$$

Integration:

$$\ln \varphi(v) = \frac{1}{2} k \cdot v^2 + c$$

$$\varphi(v) = e^{\frac{1}{2}k \cdot v^2 + c} = e^c \cdot e^{\frac{1}{2}k \cdot v^2}$$

For the assumption of a very large number of observations it holds:

$x \to \xi$ and $v \to \varepsilon$:

$$\varphi(\varepsilon) = e^c \cdot e^{\frac{1}{2}k \cdot \varepsilon^2}$$

As the probability of a growing ε should decrease:

$$\frac{1}{2} k = -h^2 \; ; \text{ in addition } \quad e^c = A$$

$$\varphi(\varepsilon) = A \cdot e^{-h^2 \varepsilon^2}$$

Two conditions are needed in order to determine the integration constants A and h:

$$1) \int_{-\infty}^{+\infty} \varphi(\varepsilon) d\varepsilon = A \int_{-\infty}^{+\infty} e^{-h^2 \varepsilon^2} d\varepsilon = 1$$

According to chapter 1.2.3: $A = \frac{h}{\sqrt{\pi}}$

2) The precision value h of Gauß can be expressed by the variance σ^2:

$$\sigma^2 = \int_{-\infty}^{+\infty} \varepsilon^2 \varphi(\varepsilon) d\varepsilon = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \varepsilon^2 e^{-h^2 \varepsilon^2} d\varepsilon$$

In general:

$$\int_{-\infty}^{+\infty} x^2 e^{-(ax^2 + bx + c)} dx = \frac{a + 2b^2}{2a^2} \sqrt{\frac{\pi}{a}} \cdot e^{\frac{b^2 - ac}{a}} \quad ; \; a > 0$$

Here: $a = h^2$; $b = c = 0$

$$\sigma^2 = \frac{h}{\sqrt{\pi}} \cdot \frac{1}{2h^2} \cdot \sqrt{\frac{\pi}{h^2}} = \frac{1}{2h^2} \to h = \frac{1}{\sigma \cdot \sqrt{2}}$$

Gauß's law of errors density function for $N(0, \sigma)$):

$$\varphi(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon^2} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{\varepsilon^2}{2\sigma^2}}$$

6.2.9. Different Error Values and their Relationship to Normal Distributed Observations

Definition of the variance:

$$\sigma^2 = E\left\{(\xi - l_\lambda)^2\right\} = E(\varepsilon^2) = \int_{-\infty}^{+\infty} \varepsilon^2 \varphi(\varepsilon) d\varepsilon = 2\int_{0}^{+\infty} \varepsilon^2 \varphi(\varepsilon) d\varepsilon$$

Definition of the average error:

$$\tau = E|\varepsilon| = \int_{-\infty}^{+\infty} |\varepsilon| \varphi(\varepsilon) d\varepsilon = 2\int_{0}^{+\infty} |\varepsilon| \varphi(\varepsilon) d\varepsilon$$

Definition of the most probable error ρ:

$$\int_{-\rho}^{+\rho} \varphi(\varepsilon) d\varepsilon = 0{,}5 = \Phi(+\rho) - \Phi(-\rho) = \Phi(+\rho) - \{1 - \Phi(+\rho)\} = 2\Phi(\rho) - 1$$

i.e. $\Phi(+\rho) = \frac{3}{4} \rightarrow \rho$

For independent normal distributed observations the integrations result in:

$$\sigma^2 = \frac{1}{2h^2}$$

h = precision value

$$\tau = \frac{1}{\sqrt{\pi} \cdot h} = \frac{\sqrt{2} \cdot \sigma}{\sqrt{\pi}}$$

The following relationships result:

$$\tau = 0{,}80\sigma = 1{,}18\rho$$
$$\sigma = 1{,}25\tau = 1{,}47\rho$$
$$\rho = 0{,}85\tau = 0{,}67\sigma$$

$$\tau \approx \frac{4}{5}\sigma \;\; ; \;\; \rho \approx \frac{2}{3}\sigma$$

6.3. Statistical Properties of Normal Distributed Random Variables

6.3.1. Distribution of the arithmetic mean

If the variables **l** are normal distributed following $N(\xi, \sigma)$, then also the sample mean $x = \frac{[l]}{n}$ will strictly be $N(\xi, \frac{\sigma}{\sqrt{n}})$ normal distributed.

Based on the central boundary statement it holds: For any distribution with expectation value ξ and standard deviation σ, the sample mean will approximately be normal distributed $N(\xi, \frac{\sigma}{\sqrt{n}})$; the approach will be better for growing n and the more the individual distribution resembles the normal distribution.

6.3.2. Confidence Interval for the Expectation Value ξ

6.3.2.1. Confidence Interval for ξ when σ_x is given

Assumption: All observations come from the same normal distributed total sample and are independent of each other.

$$P(-u_s \le \frac{\varepsilon_x}{\sigma_x} \le +u_s) = S = \int_{-u_s}^{+u_s} \varphi(t)dt$$

$$u = \frac{x-\xi}{\sigma_x} = \frac{\varepsilon_x}{\sigma_x} \quad \text{mit} \quad x = \frac{[l]}{n} \quad \text{und} \quad \sigma_x = \frac{\sigma}{\sqrt{n}}$$

u ... normalized random variable

$$P(-u_s \le \frac{x-\xi}{\sigma_x} \le +u_s) = S$$

By re-arrangement of the relation, we get:

$$\begin{aligned} -u_s \cdot \sigma_x \le x - \xi \le u_s \cdot \sigma_x \\ -x - u_s \cdot \sigma_x \le -\xi \le -x + u_s \cdot \sigma_x \\ x - u_s \cdot \sigma_x \le \xi \le x + u_s \cdot \sigma_x \end{aligned}$$

and:

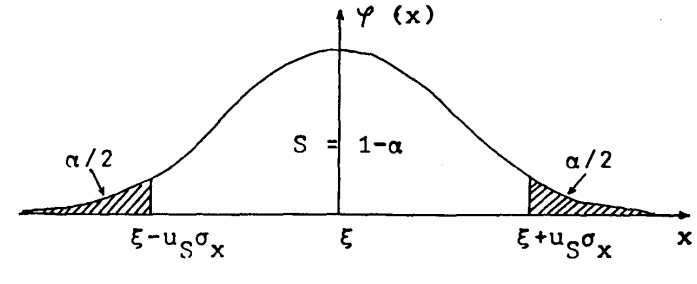
$$\xi - u_s \cdot \sigma_x \le x \le \xi + u_s \cdot \sigma_x$$

Therefore:

$$P(x - u_s \cdot \sigma_x \le \xi \le x + u_s \cdot \sigma_x) = S$$
$$P(\xi - u_s \cdot \sigma_x \le x \le \xi + u_s \cdot \sigma_x) = S$$

S = statistical safety
The choice of S depends on the problem.

S%	u_s
68,3	1,00
95,0	1,96
99,0	2,58
99,9	3,29



Confidence interval:

$x - u_s\sigma_x \ldots x + u_s\sigma_x$ (two sides).

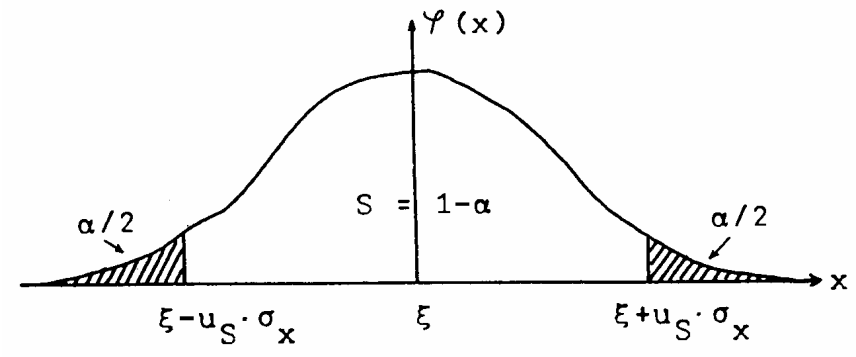
(Random interval of constant length, which holds the expectation value ξ with the probability S).

Confidence boundaries: $\pm u_s\sigma_x$

The statements are valid if σ_x is known.

Necessary size of the sample

How large the size of the sample must be that the deviation $|x - \xi|$ of the expectation value ξ from the estimation x does not exceed a given amount C_{Max} in S% of all cases?



x is $N(\xi, \sigma_x)$ distributed

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$x = \frac{[l]}{n}$$

$$C_{Max} = u_s \cdot \sigma_x = u_s \cdot \frac{\sigma}{\sqrt{n_{Min}}}$$

$$C_{Max}^2 = u_s^2 \cdot \sigma_x^2 = u_s^2 \cdot \frac{\sigma^2}{n_{Min}}$$

$$n_{Min} = \left(\frac{u_s \cdot \sigma}{C_{Max}}\right)^2$$

Example:

Measuring the height differences at the fundament of a tall building two levelling instruments could be used. The standard deviations of a single measurement for both instruments are known.

1) Engineering leveling instrument: $\sigma_1 = \pm 0{,}6\text{mm}$
2) Precise instrument: $\sigma_2 = \pm 0{,}2\text{mm}$

a)How often the height difference has to be determined with instrument 1 or 2 if the maximal difference would be smaller than $C_{Max} = |x - \xi| < 0{,}3\text{mm}$ based on a statistical safety factor of S = 95%?

$S = 95\% \; ; \; u_s = 1{,}96$

1. Instrument:

$$n_{Min} = \left(\frac{1{,}96 \cdot 0{,}6}{0{,}3} \right)^2 = 15{,}36 \approx 16$$

2. Instrument:

$$n_{Min} = \left(\frac{1{,}96 \cdot 0{,}2}{0{,}3} \right)^2 = 1{,}7 \approx 2$$

b) It is requested that the following condition holds:
$C_{Max} < 0{,}1\text{mm}$ with S = 99 %.
Could this possibly be achieved with both instruments?

$S = 99\% \; ; \; u_s = 2{,}58$

1. Instrument:

$$n_{Min} = \left(\frac{2{,}58 \cdot 0{,}6}{0{,}1} \right)^2 = 239{,}63 \approx 240$$ completely unrealistic

2. Instrument:

$$n_{Min} = \left(\frac{2{,}58 \cdot 0{,}2}{0{,}1} \right)^2 = 26{,}62 \approx 27$$

c) How large the standard deviation σ of an instrument should be if a total number of n = 5 measurements would be allowed to fulfil the request of b)?

$$5 = \left(\frac{2{,}58 \cdot \sigma}{0{,}1} \right)^2$$

$$\sigma = \frac{0{,}1}{2{,}58} \cdot \sqrt{5} = 0{,}087\text{mm}$$

The t-Distribution (Student-Distribution)

The t-distribution was derived by W.S. Gosset (Pseudonym: Student) (1908).

l_i ..... observations of $N(\xi, \sigma)$ normal distributed random variables X
x ..... arithmetic mean (random variable)
m .....estimation for σ (random variable)
m_x ... estimation for the standard deviation of the arithmetic mean x (random variable)

m_x and x must be independent from each other.

Transition to the random variable:

$$T = \frac{X - \xi}{m_x}$$

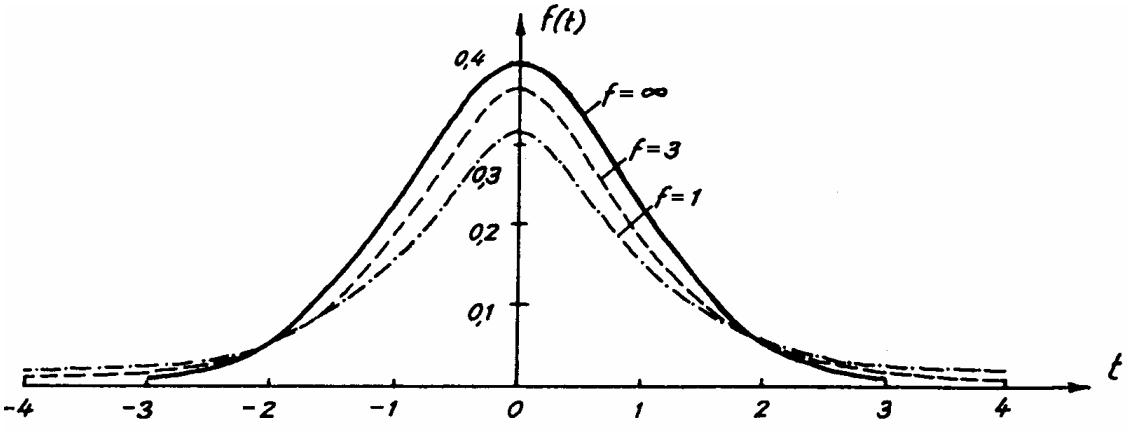
Probabilitstic density functions of the random variable T:

$$f(t, f) = C(f) \cdot \left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}}$$

$$C(f) = \frac{1}{\sqrt{f\pi}} \cdot \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)} \; ; \; \Gamma(\lambda + 1) = \int_0^\infty x^\lambda e^{-x} dx \text{ (Gamma - function)}$$

f = degree of freedom = number of redundant observations, here f = n - 1

Graphical view:
The curve belonging to f(t) is also symmetrical to ξ, resp. t = 0, due to the dependency of t^2, however it is more flat than $\varphi(\varepsilon)$.



The t-distribution approaches normal distribution N(0,1) for $f \to \infty$.

Table of values t_s as function of the probability

$$S = \int_{-t_s}^{+t_s} f(t)dt \text{ and the degree of freedom f.}$$

(f = n - 1 for one unknown parameter, resp. f = n - u for n unknown parameters)

f	S = 68,3%	S = 95,0%	S = 99,0%
1	1,82	12,71	63,66
2	1,32	4,30	9,92
3	1,20	3,18	5,84
4	1,14	2,78	4,60
5	1,11	2,57	4,03
10	1,05	2,23	3,17
20	1,025	2,09	2,85
100	1,005	1,98	2,63
∞	1,000	1,96	2,58

Instead of the table above, nomograms can be used. If f is larger than 30 $\mathbf{u_s}$ can be used , instead of **t**. However there are significant differences for small values f.

6.3.2.2. Confidence Interval for ξ with given m_x

Pre-assumption: The observations l_i date from a whole set of $N(\xi, \sigma)$ distributed independent random variables.

$$P(-t_s \le \frac{x-\xi}{m_x} \le +t_s) = S = \int_{-t_s}^{+t_s} f(t)dt$$

Corresponding to:

$$P(-u_s \le \frac{x-\xi}{\sigma_x} \le +u_s) = S = \int_{-u_s}^{+u_s} \varphi(t)dt$$

Main difference:

σ_x = fixed constant number	m_x = Estimation for σ_x (random variable)
$\varphi(u)$ density function of the random variables	f(t) density function of the random variables
$u = \frac{x-\xi}{\sigma_x}$	$t = \frac{x-\xi}{m_x}$

$$P(x - t_s \cdot m_x \le \xi \le x + t_s \cdot m_x) = S$$

The transition corresponds to 6.3.2.1 using t_s instead of u_s.

(Random interval which contains the expectation value ξ with the probability of S. The size of the interval depends from the random variable m_x).

Under the assumption of normal distributed observations, and for f=3 and S=95%, $t_s = 3{,}18$ results. In about 5% of all cases the expectation value ξ will be outside the boundary values of $\pm 3{,}18 \cdot m_x$ if m_x only contains 3 degrees of freedom. For a given standard deviation σ_x instead of m_x only 0,1 % of all measurements would be outside the range of confidence $\pm 3{,}29 \cdot \sigma_x$

Example: angular measurements:

Assumption: normal distributed observations

γ	v	v^2
3,7364	0	0
61	+3	9
69	-5	25
62	+2	4
X = 3,73640		38

$$m^2 = \frac{38}{3} = 12{,}7$$

$$m_x^2 = \frac{m^2}{4} = 3{,}17 \qquad m_x = 1{,}8^{cc}$$

$S = 95\%$ $f = n - 1 = 3$ $t_s = 3{,}18$

Confidence boundaries: $\pm t_s \cdot m_x = \pm 3{,}18 \cdot 1{,}8^{cc} = \pm 5{,}7^{cc} \approx 6^{cc}$

Comparing to: $\pm u_s \sigma_x = \pm 1{,}96 \cdot 1{,}8^{cc} = \pm 3{,}5^{cc} \approx 4^{cc}$
$(with\ \sigma_x = m_x)$

Confidence interval in which ξ will be with a probability of 95 %:

$x - t_s m_x$ $\qquad$ $x + t_s m_x$

$3^g{,}7364 - 6^{cc}$ ..... $3^g{,}7364 + 6^{cc}$
$3^g{,}7358$ ..... $3^g{,}7370$

(Comparing to $x - u_s m_x$ .... $x + u_s m_x$

3,7360 .... 3,7368)

6.3.3. Statistical Properties of the Estimation of the Variance, Calculated from True Random Errors ε_i via $m^2 = [\varepsilon^2] / f$ (E(x) given)

6.3.3.1. Relationship between the Random Variables m^2 and χ^2_f

Pre-assumption:
The independent observations l_i are $N(\xi, \sigma)$ distributed. ξ is a known parameter.

Estimation for σ^2 (random variable):

$$m^2 = \frac{\left[\varepsilon^2\right]_l^f}{f}$$

with f = degree of freedom (here f = n = number of observations).

Relationship between ε_i from $N(\xi, \sigma)$ and the normalized variables u_i from N(0,1):

$$u_i = \frac{\varepsilon_i}{\sigma}$$

$$\varepsilon_i = \sigma \cdot u_i$$

$$\left[\varepsilon_i^2\right]_l^f = \sigma^2 \left[u_i^2\right]_l^f$$

$$m_f^2 = \frac{\sigma^2}{f} \cdot \left[u_i^2\right]_l^f$$

Usual notation:

$$\left[u_i^2\right]_l^f = \chi_f^2$$

χ_f^2-Distribution: Distribution of the sum of squares of f independent N (0,1) distributed values u_i.

$$m_f^2 = \frac{\sigma^2}{f} \cdot \chi_f^2$$

6.3.3.2. χ^2-Distribution

Helmert (1876), Pearson (1900).
Distributive function $VF(\chi_f^2)$ of the random variables χ_f^2:

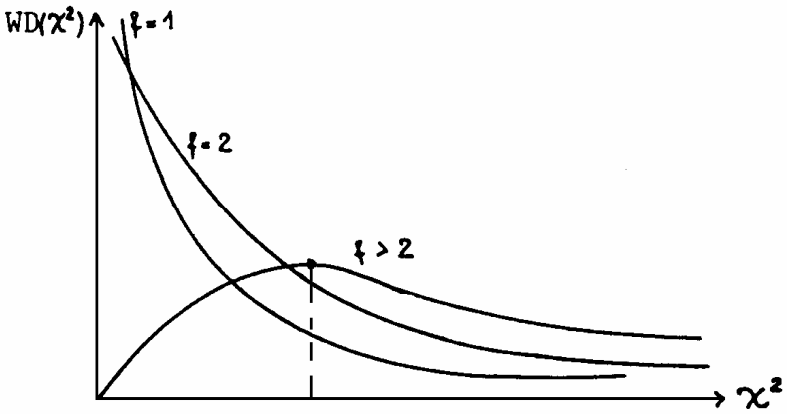
$$P(\chi_f^2 = \left[u_i^2\right]_l^f < z) = \left\{\Gamma\left(\frac{f}{2}\right)\right\}^{-1} 2^{-\frac{f}{2}} \cdot \int_0^z y^{\left(\frac{f}{2}-1\right)} \cdot e^{-\frac{y}{2}} \cdot dy$$

Probabilistic density function $WD(\chi_f^2)$ of the random variables χ_f^2:

$$WD(\chi_f^2 = z) = \left\{\Gamma\left(\frac{f}{2}\right)\right\}^{-1} 2^{-\frac{f}{2}} \cdot z^{\left(\frac{f}{2}-1\right)} \cdot e^{-\frac{z}{2}}$$

f = number of the summed up independent N (0,1) distributed u_i^2

Graph:
Maximum of $\chi_f^2 = f - 2$ for f > 2



Addition theorem of χ^2-distribution:
The sum of two independent χ^2-distributed random variables again is χ^2 - distributed.

$$\chi_1^2 = y_1^2 + y_2^2 + \ldots + y_{f1}^2 \qquad \text{Degree of freedom } f_1$$
$$\chi_2^2 = z_1^2 + z_2^2 + \ldots + z_{f2}^2 \qquad \text{Degree of freedom } f_2$$
$$\chi^2 = \chi_1^2 + \chi_2^2 = \left[y_i^2\right]_1^{f_1} + \left[z_j^2\right]_1^{f_2} \qquad \text{Degree of freedom } f_1 + f_2$$

The variables y_i and z_j are independent from each other N(0,1) distributed.

Expectation value $E(\chi^2)$ of the random variables χ_f^2:

$$\chi_f^2 = u_1^2 + u_2^2 + \ldots + u_f^2$$
$$E(\chi^2) = E(u_1^2) + E(u_2^2) + \ldots + E(u_f^2) = f \underbrace{E(u^2)}_{1} = f$$

Variance $VAR(\chi_f^2)$ of the random variables χ_f^2:

$$VAR(\chi_f^2) = E\left\{(\chi_f^2)^2\right\} - \left\{E(\chi_f^2)\right\}^2 = E(\chi_f^4) - f^2$$

$$E(\chi_f^4) = E\left\{(u_1^2 + u_2^2 + \ldots + u_f^2)^2\right\}$$
$$= E(u_1^4 + u_1^2u_2^2 + \ldots + u_1^2u_f^2$$
$$+ u_2^2u_1^2 + u_2^4 + \ldots + u_2^2u_f^2$$
$$\vdots \qquad \vdots$$
$$+ u_f^2u_1^2 + u_f^2u_2^2 + \ldots + u_f^4)$$

$$E(u^4) = \int_{-\infty}^{+\infty} u^4 \varphi(u)du \qquad \text{(Special case: u is N(0,1) distributed)}$$
$$= \int_{-\infty}^{+\infty} \frac{u^4}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} du$$

In general it holds (from Integral formula collection):

$$I = \int_0^{\infty} x^{2n} \cdot e^{-ax^2} dx = \frac{1 \cdot 3 \ldots (2n-1)}{2^{n+1} \cdot a^n} \sqrt{\frac{\pi}{a}} \; ; \; n = 0,1,2\ldots$$

Here: $x \to u;\ n \to 2;\ a \to +\frac{1}{2}$; therefore it is:

$$E(u^4) = \frac{1}{\sqrt{2\pi}} \cdot 2I = \frac{2}{\sqrt{2\pi}} \cdot \frac{1 \cdot 3}{2^3} \cdot 2^2 \cdot \sqrt{2\pi} = 3$$

In addition it holds:

$$E(u_i^2 u_k^2) = E(u_i^2) \cdot E(u_k^2) = 1 \cdot 1 = 1$$
$$E(\chi_f^4) = f \cdot \underbrace{E(u^4)}_{3} + (f^2 - f) \cdot \underbrace{E(u_i^2)}_{1} \cdot \underbrace{E(u_k^2)}_{1} = 2f + f^2$$
$$VAR(\chi_f^2) = \underbrace{E(\chi_f^2)}_{2f+f^2} - f^2$$
$$VAR(\chi_f^2) = 2f$$

Standard deviation $SA(\chi_f^2)$ of the random variable χ_f^2:

$$SA(\chi_f^2) = \sqrt{VAR(\chi_f^2)} = \sqrt{2f}$$

6.3.3.3. Probabilistic Density Function and Estimated Variance m^2 for f Degrees of Freedom

Density function $WD(m^2)$ of m^2:

$$m^2 = [\varepsilon\varepsilon] / f :$$

$$m^2 = \frac{\sigma^2}{f} \cdot \chi_f^2 \quad \chi_f^2 = \frac{f}{\sigma^2} m^2$$

As $WD(m^2)$ and $WD(\chi_f^2)$ are monotone within a definite range, it holds:

$$WD(m^2) = \frac{d(\chi^2)}{d(m^2)} \cdot WD(\chi_f^2)$$

$$\frac{d(\chi^2)}{d(m^2)} = \frac{f}{\sigma^2}$$

$$WD(m^2) = \frac{f}{\sigma^2} WD(\chi_f^2)$$

Remark: The probability function WD(m) of the estimated standard deviation m can not be derived from that.

$$\sigma^2 = E(\varepsilon^2) = \frac{[\varepsilon\varepsilon]}{n} \qquad \sigma = \sqrt{E(\varepsilon^2)} = \sqrt{\frac{[\varepsilon\varepsilon]}{n}}$$

$n \to \infty$ unequal

$$E(m) = \frac{m_1 + ... + m_\nu}{\nu}$$

$\nu \to \infty$

For practical applications only special values of the distributive function $VF(\chi_f^2)$ are required.

Expectation value $E(m^2)$ **and Variance** $VAR(m^2)$ **of** m^2**:**

$$m^2 = \frac{\sigma^2}{f} \cdot \chi_f^2$$

$$E(m^2) = \frac{\sigma^2}{f} \cdot E(\chi_f^2)$$

$$VAR(m^2) = (\frac{\sigma^2}{f})^2 \cdot VAR(\chi_f^2) = \sigma^4 \cdot \frac{2}{f} \qquad VAR(\chi_f^2) = 2f$$

Standard deviation $SA(m^2)$ of the estimation m²:

$$SA(m^2) = \sqrt{VAR(m^2)} = \sigma^2 \cdot \sqrt{\frac{2}{f}}$$

For f independent normal distributed observations the relationship are strictly valid.
Variance VAR(m) of m.

Based on very problematic pre-assumptions:

$dm << m$

It holds with $y = m^2$:

dy = 2m . dm + terms of higher order.

$$VAR(y) = (2m)^2 \cdot VAR(m) \quad VAR(y) = VAR(m^2)$$

$$VAR(m) = \frac{1}{(2m)^2} \cdot VAR(m^2) = \frac{1}{4m^2} \cdot \sigma^4 \frac{2}{f} = (\frac{\sigma^4}{m^2}) \cdot \frac{1}{2f}$$

Standard deviation SA(m) of m:

$$SA(m) = \sqrt{VAR(m)} = \frac{\sigma^2}{m}\sqrt{\frac{1}{2f}}$$

For all formula the knowledge of σ is needed. This case is rarely given (i.e. from experience).

6.3.3.4. Confidence Interval for σ

Instead of the random variable

$$\chi^2 = \frac{[\varepsilon^2]}{\sigma^2} \quad \text{with} \quad E(\chi^2) = f \quad \text{(here f = n, in general f = n - u)}$$

The random variable

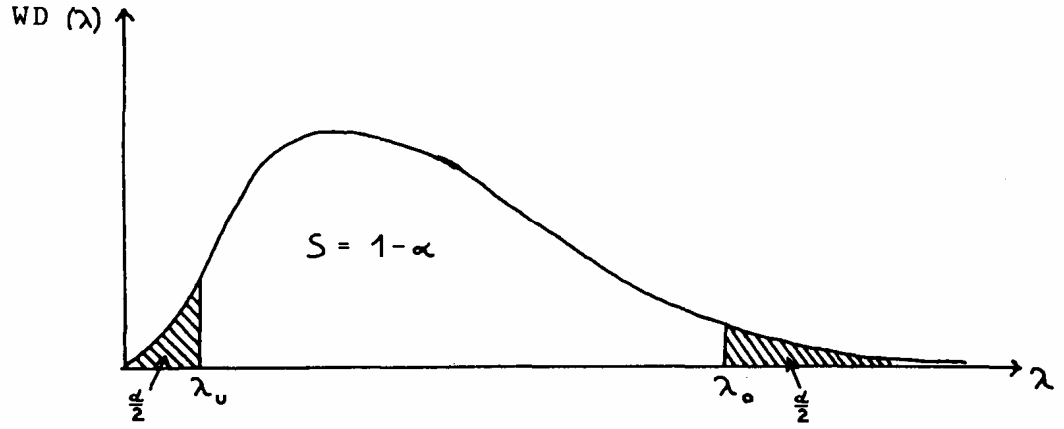
$$\lambda^2 = \frac{\chi^2}{f} = \frac{[\varepsilon^2]}{f\sigma^2} = \frac{m^2}{\sigma^2} \quad ; \quad \lambda = \frac{m}{\sigma}$$

will be used.

$$E(\lambda^2) = \frac{E(\chi^2)}{f} = 1$$

$$VAR(\lambda^2) = \frac{1}{f^2} VAR(\chi^2) = \frac{2f}{f^2} = \frac{2}{f}$$

Two-sided statistical safety:



$$P(\lambda_u \le \lambda \le \lambda_o) = S$$

Two-Sided Confidence Interval of the Random Variable λ

$$\lambda_u \ldots \lambda_o , \qquad \lambda = \frac{m}{\sigma}$$

$$P(\lambda \le \lambda_u) = P(\lambda \ge \lambda_o) = \frac{\alpha}{2}$$

$$P(\frac{m}{\sigma} \le \lambda_u) = P(\frac{m}{\sigma} \ge \lambda_o) = \frac{\alpha}{2}$$

$$P(m \le \sigma \cdot \lambda_u) = P(m \ge \sigma \cdot \lambda_o) = \frac{\alpha}{2}$$

$$P(\lambda_u \le \frac{m}{\sigma} \le \lambda_o) = 1 - \alpha = S$$

$$P(\sigma \cdot \lambda_u < m < \sigma \cdot \lambda_o) = 1 - \alpha = S$$

$\sigma \cdot \lambda_u \ldots \sigma \cdot \lambda_o$ Interval of constant length, in which m (as estimation for σ) will be with probability S :

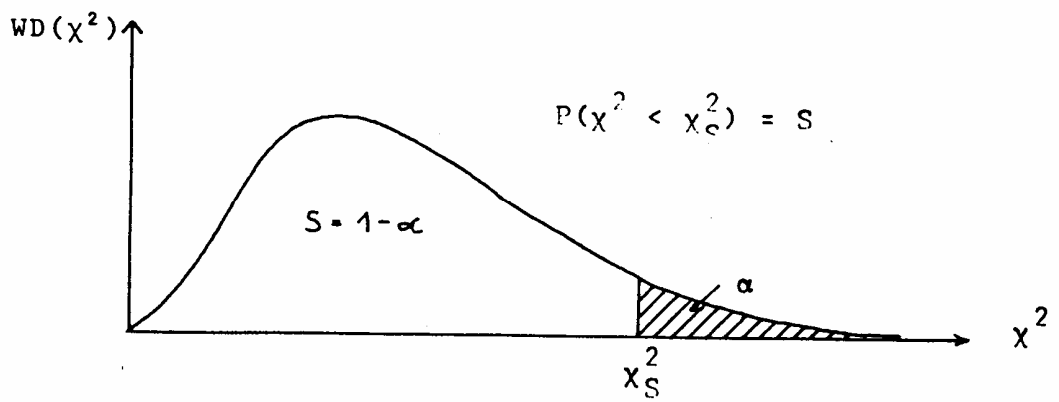
$$P(m / \lambda_o < \sigma < m / \lambda_u) = 1 - \alpha = S$$

$m / \lambda_o \ldots m / \lambda_u$ random interval which contains σ with probability S.

The values λ_o and λ_u can be taken out of a specific graph or table.

One-sided statistical safety of the χ^2 -distribution

$$P(\chi^2 < \chi_S^2) = S$$



Examples: Given n independent $N(0, \sigma_d)$ distributed observations.

1) The standard deviation σ_d of the entire set is known as $\sigma_d = \pm 3^{cc}$

In which range the estimation m_d is to be expected for S=95%:

$$m_d = \sqrt{\frac{[d^2]}{n}}$$

And f = n = 4 degrees of freedom?

$$\sigma_d \cdot \lambda_u < m_d < \sigma_d \cdot \lambda_o$$

For f = 4 and S = 95 % it holds: $\lambda_u = 0{,}35$ $\lambda_o = 1{,}66$

(lower boundary of interval) ; (upper boundary of interval)

$$3 \cdot 0.35 = 1{,}0^{cc} \qquad 3 \cdot 1{,}66 = 5{,}0^{cc}$$

2) m_d derived from observational differences resulted in:

$$m_d = \sqrt{\frac{[d^2]}{n}} = 3^{cc}$$

How large is the confidence interval of σ_d for S = 95 %?

For f = n = 4	For f = n = 30
$\lambda_u = 0{,}35;\ \lambda_o = 1{,}66$	$\lambda_u = 0{,}75;\ \lambda_o = 1{,}25$
$3 / 1{,}66 < \sigma_d < 3 / 0{,}35$	$3 / 1{,}25 < \sigma_d < 3 / 0{,}75$
$1{,}8^{cc} < \sigma_d < 8{,}6^{cc}$	$2{,}4^{cc} < \sigma_d < 4{,}0^{cc}$

$$m_d / \lambda_o < \sigma_d < m_d / \lambda_u$$

6.3.4. Confidence Intervals for the Expectation Values ξ_i

Pre-assumption: independent normal distributed observations and validity of the functional and stochastical model

$$E(x_i) = \xi_i$$

$$P(-t_s \le \frac{x_i - \xi_i}{m_{x_i}} \le +t_s) = S$$

$$P(x_i - t_s m_{x_i} \le \xi_i \le x_i + t_s m_{x_i}) = S$$

with $m_{x_i} = m_0 \sqrt{Q_{ii}} = \sqrt{\frac{[vv]}{n-u}} \cdot \sqrt{Q_{ii}}$;

Degree of freedom to determine t_s : $f = n - u$

Example: calibration of a scale of 1m - length

Comparative measurements l_i at different temperatures:

$$l_i + v_i = x_i + x_2(t_i - 20^o) \quad \text{in } [\mu m] \ (i = 1...7)$$

Calculated deviation at 20^o C : $x_1 = +3{,}2\ [\mu m]$

Coefficient of extension : $x_2 = +1{,}3\ [\mu$ m/ ^{o} C $]$

$$Q_{xx} = \begin{pmatrix} +0{,}1020 & +0{,}0102 \\ +0{,}0102 & +0{,}0216 \end{pmatrix}$$

Cofactor matrix of unknowns

$[vv] = 45{,}13\ [(\mu m)^2]$

Degree of freedom : f = n – u = 7 – 2 = 5

Estimation of variance of unit weight: $m_0 = \sqrt{\frac{45{,}13}{5}} = \pm 3{,}0\ [\mu m]$

Estimated variances of unknowns:

$$m_{x_1} = 3{,}0 \cdot \sqrt{0{,}1020} = \pm 0{,}96\ [\mu m]$$
$$m_{x_2} = 3{,}0 \cdot \sqrt{0{,}0216} = \pm 0{,}44\ [\mu m/^o C]$$

Estimation of intervals:

Confidence interval of the unknowns (S = 95%)

For S = 95% based on $f = 5$: $t_s = 2{,}57$

$$x_i - t_s \cdot m_{x_i} \le \xi_i \le x_i + t_s \cdot m_{x_i}$$
$$3{,}2 - 2{,}57 \cdot 0{,}96 \le \xi_1 \le 3{,}2 + 2{,}57 \cdot 0{,}96$$
$$+0{,}7 \le \xi_1 \le +5{,}7\ [\mu m]$$

Theoretically the corresponding interval will contain the expectation value ξ_1 in 95% of all cases.

$$1{,}3 - 2{,}57 \cdot 0{,}44 \le \xi_2 \le 1{,}3 + 2{,}57 \cdot 0{,}44$$
$$0{,}2 \le \xi_2 \le 2{,}4\ [\mu m/^o C]$$

Confidence interval for the expectation value Φ of a function F

F = Deviation of the scale from the correct value at t = 17^{o}C

$$F = x_1 + x_2 (t - 20^o)$$
$$F = x_1 + 3x_2$$
$$= 3{,}2 - 3{,}9 = 0{,}7\ [\mu m] \qquad \text{(point estimation)}$$

$$Q_{FF} = Q_{11} - 6Q_{12} + 9Q_{22}$$
$$= 0{,}1020 - 0{,}0612 + 0{,}1944 = 0{,}2352$$
$$m_F = 3{,}0 \cdot \sqrt{0{,}2352} = \pm 1{,}46 \,[\mu m]$$

(interval estimation)

$$-0{,}7 - 2{,}57 \cdot 1{,}46 \le \Phi \le -0{,}7 + 2{,}57 \cdot 1{,}46$$
$$-4{,}4 \le \Phi \le 3{,}0 \,[\mu m]$$

6.3.5. Confidence Intervals for the Standard Deviations σ_0, σ_x, σ_F

In order to get χ^2, λ_o, λ_u the degree of freedom $f = n - u$ has to be used.

$$P(\sigma_0 \cdot \lambda_u \le m_0 \le \sigma_0 \cdot \lambda_o) = S$$
$$P(m_0 / \lambda_o \le |\sigma_0| \le m_0 / \lambda_u) = S$$

Example above:

For S = 95% and $f = 5$ we get : $\lambda_o = 1{,}60$; $\lambda_u = 0{,}405$

$m_0 = \pm 3{,}0 \,[\mu m]$ (point estimation)

$$3{,}0 / 1{,}60 \le \sigma_o \le 3{,}0 / 0{,}405$$
$$1{,}9 \le |\sigma_o| \le 7{,}4 \,[\mu m] \quad \text{(Interval estimation)}$$

For $m_{x_1} = m_0 \cdot \sqrt{Q_{11}} = \pm 0{,}95 \,[\mu m]$ based on $S = 95\%$ and $f = 5$:

$$0{,}95 / 1{,}60 \le \sigma_{x_1} \le 0{,}95 / 0{,}405$$
$$0{,}6 \le |\sigma_{x_1}| \le 2{,}3 [\mu m]$$

For $m_{x_2} = m_0 \cdot \sqrt{Q_{22}} = \pm 0{,}44 \,[\mu m/^{\circ}C]$ based on $S = 95\%$ and $f = 5$:

$$0{,}44 / 1{,}60 \le \sigma_{x_2} \le 0{,}44 / 0{,}405$$
$$0{,}3 \le |\sigma_{x_2}| \le 1{,}1 \,[\mu m/^{\circ}C]$$

For $m_F = m_0 \cdot \sqrt{Q_{FF}} = \pm 1{,}46 \,[\mu m]$ based on S = 95% and $f = 5$:

$$1{,}46 / 1{,}60 \le \sigma_F \le 1{,}46 / 0{,}405$$
$$0{,}9 \le |\sigma_F| \le 3{,}6 \,[\mu m]$$

6.3.6. Comparison of Two Estimations of Standard Deviations

6.3.6.1. F-Distribution

It is assumed that two independent series of measurements are given where the observations are independent of each other. They might have their individual expectation values however they share a common variance σ^2.

$$m_1^2 = estimation \text{ for } \sigma^2 \text{ from the 1. serie}$$

$$m_1^2 = \frac{[vv]_1}{n_1 - 1} = \frac{\chi_1^2 \cdot \sigma^2}{f_1}$$

$$m_2^2 = estimation \text{ for } \sigma^2 \text{ from the 2. serie}$$

$$m_2^2 = \frac{[vv]_2}{n_2 - 1} = \frac{\chi_2^2 \cdot \sigma^2}{f_2}$$

$$F - test - \text{variable}:$$

$$F(f_1, f_2) = \frac{m_1^2}{m_2^2} = \frac{f_2 \cdot \chi_1^2}{f_1 \cdot \chi_2^2}$$

The F-test-variable follows an F-distribution. The graph of the density function resembles the χ^2 - distribution. For a given safety factor S, F_S can be calculated based on f_1 and f_2 according to the following formula. (Table and nomograms could be used).

$$\frac{1}{F_{(1-\alpha)}(f_1, f_2)} = \frac{f_1 \cdot \chi_2^2}{f_2 \cdot \chi_1^2} = F_\alpha(f_2, f_1)$$

It is sufficient to choose $m_1^2 > m_2^2$ and to regard the range $F \geq 1$ only.

F is suited for a test of the hypothesis $\sigma_1^2 = \sigma_2^2$ (F-Test).

One-sided F-Test

In most cases the need for a one-sided F-test exists.

The numeration is chosen that $m_1 > m_2$.

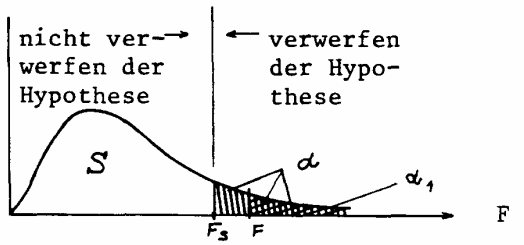
Based on experience the possibility exists that σ_1 will be larger or in the best case equal to σ_2.

The possibility $\sigma_1 < \sigma_2$ will be excluded in advance.

Hypothesis: $\sigma_1 = \sigma_1;\ E(m_1^2) = E(m_2^2)$

Alternative: $\sigma_1 > \sigma_1;\ E(m_1^2) > E(m_2^2)$

Error of first kind: α_1



$P(F < F_S) = S = 1 - \alpha$

The error of first kind is given if the hypothesis will be rejected and the alternative $\alpha_1 \leq \alpha$ will be accepted.

For the boundary significance α = 5%, F_s has to calculated using f_1 and f_2 at S=95%.

Example:

ε_I	ε_{II}
+ 3 mm	0
- 5	- 1
- 2	0
+ 2	+ 7
0	- 1
- 1	+ 5
- 2	- 1
- 3	0
+ 1	- 4
0	0

$$m_I = \sqrt{\frac{[\varepsilon_I^2]}{n}} = \pm 2{,}4\text{mm} \; ; \; m_{II} = \sqrt{\frac{[\varepsilon_{II}^2]}{n}} = \pm 3{,}1\text{mm} \; ; \; f = n = 10 = f_1 = f_2$$

Hypothesis: $\sigma_I = \sigma_{II}$

Alternative: $\sigma_{II} > \sigma_I$

$m_I \rightarrow m_2 \quad m_{II} \rightarrow m_1$

Test parameter: $F = (\frac{m_{II}}{m_I})^2 = 1{,}66$

$F_{95\%}(10{,}10) = 3{,}0 \quad \alpha_1 = 5\%; \; F < F_S$.

The test does not give any indication to reject the hypothesis and to accept the alternative also the felt impression is different.

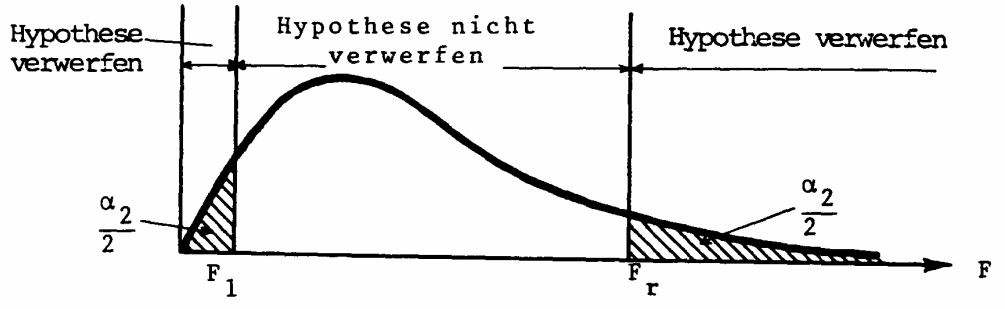
6.3.6.2. Two-sided F-Test

Based on experience there is no indication that σ_1 could be larger or smaller than σ_2 .
Numeration of 1 and 2 is first optional.

Hypothesis: $\sigma_1 = \sigma_2$

Alternative: $\sigma_1 \neq \sigma_2$

Error of first kind: α_2



$$P(F < F_1) = P(F > F_r) = \frac{\alpha_2}{2}$$

$$P(\frac{1}{F} > \frac{1}{F_1}) = P(F > F_r) = \frac{\alpha_2}{2}$$

The numeration will be chosen that $m_1 > m_2$ and the test variable F will be calculated:

$$F(f_1, f_2) = \frac{m_1^2}{m_2^2}$$

For $F(f_1, f_2) > F_r(f_1, f_2)$, the hypothesis will be rejected.

Observe that for F_r it holds: $P(F < F_r) = S = 1 - \frac{\alpha_2}{2}$

The error of first kind is $\alpha_2 = 2\alpha$ if the hypothesis will be rejected and the alternative will be accepted. For S = 95 % the error of first kind will be α_2 = 10 %. For a given probability of error of 5 % , F-distribution has to be used at 97,5 %.

Observer 1

$$m_S = \sqrt{\frac{[v_S^2]}{n-1}} = \pm 3{,}50^{cc}$$

Observer 2

$$m_M = \sqrt{\frac{[v_M^2]}{n-1}} = \pm 4{,}49^{cc}$$

Degrees of freedom: $f_S = f_M = 14$

Hypothesis: $\sigma_S = \sigma_M$

Alternative: $\sigma_S \neq \sigma_M$ $\quad m_M \rightarrow m_1$; $m_S \rightarrow m_2$

Test variable: $$F = \left(\frac{m_1}{m_2}\right)^2 = 1{,}66$$

$$F_{95\%}(14{,}14) = 2{,}5 \qquad \alpha_2 = 2\alpha = 10\% \ ;$$

$$F_{97{,}5\%}(14{,}14) = 3{,}0 \qquad \alpha_2 = 2\alpha = 5\% \ ; \qquad F < F_{90\%}$$

The test does not give any reason for rejecting the hypothesis that the measurements of both series have the same accuracy.