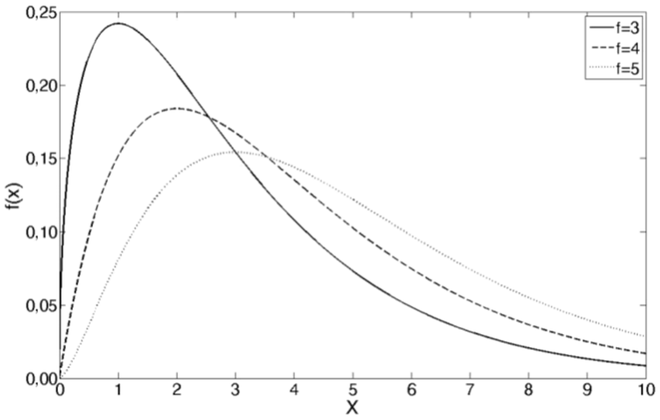


Institute of Geodesy and Geoinformation Science

Chair of Geodesy and Adjustment Theory

Adjustment Theory I

Chapter 5: The Gaussian or Normal Distribution

Prof. Dr.-Ing. Frank Neitzel

Last update: 17 November 2024

Contents

5. The Gaussian or Normal Distribution

- In previous sections we have discussed random variables with arbitrary distribution
- Now: we consider random variables with certain distribution functions
- Amongst all distribution functions the Normal Distribution plays a special role because random variables in geodesy (almost) follow this distribution
- For the derivation of normal distribution we apply „theory of elementary errors“.

5.1 Theory of Elementary Errors

Theory: Was developed by BESSEL and HAGEN in 1837

Random deviation ε of an observation L can be described by a sum of q very small „elementary errors" Δ_i.

$$\varepsilon = \Delta_1 + \Delta_2 + \Delta_3 + \ldots.. + \Delta_q \quad ; \qquad \varepsilon = \sum_{i=1}^{q} \Delta_i$$

Under the following conditions:

– All values Δ_i have the same absolute value $\Delta_i = |\delta| \quad \forall_i$

– All values Δ_i differ only in their sign $\varepsilon = \pm\delta \pm \delta \pm \delta + \ldots\ldots$

– Positive and negative signs have same probability

$$P(\Delta_i = +\delta) = P(\Delta_i = -\delta) = 0.5$$

5.1 Theory of Elementary Errors

For $q \rightarrow \infty \;\Rightarrow\; \delta \rightarrow 0$

(for ∞ number of elementary errors the magnitude of the errors is almost zero)

Expectation $E\left(\Delta_h\right) = 0$

$$\Rightarrow \quad E\left(\varepsilon\right) = \sum_{h=1}^{q} E\left(\Delta_h\right) = 0 \qquad\qquad \sigma^2 = E\left(\varepsilon^2\right) = \sum_{h=1}^{q} E\left(\Delta_h^2\right) = q \cdot \delta^2$$

Examples: $\varepsilon = \Delta_1 + \Delta_2 + \Delta_3 + ... + \Delta_q$

Probability: $P_i(\varepsilon) = \dfrac{k_i}{\sum k_i}$

$\Rightarrow q = 1 \quad : \quad \varepsilon = \Delta_1$

Case number	Δ_1	ε	Frequency (k)	Relative Frequency = Probability P
1	+	$+\delta$	1	½ = 0.5
2	-	$-\delta$	1	½ = 0.5

$\Rightarrow q = 2 \quad : \quad \varepsilon = \Delta_1 + \Delta_2$

Case number	Δ_1	Δ_2	ε	Frequency (k)	Relative Frequency = Probability P
1	+	+	$+2\delta$	1	¼ = 0.25
2	+	-	0	2	2/4 = 0.5
3	-	+	0		
4	-	-	-2δ	1	¼ = 0.25

$$\Rightarrow \; q = 4 \quad : \quad \varepsilon = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$$

Case number	Δ_1 Δ_2 Δ_3 Δ_4	ε	Frequency (k)	Relative Frequency = Probability P
1	+ + + +	4δ	1	1/16 = 0.0625
2	+ + + -	2δ	4	4/16 = 0.25
3	+ + - +			
4	+ - + +			
5	- + + +			
6	+ + - -	0	6	6/16 = 0.375
7	+ - - +			
8	- - + +			
9	- + + -			
10	- + - +			
11	+ - + -			

$$\Rightarrow \quad q = 4 \quad : \quad \varepsilon = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$$

Case number	Δ_1 Δ_2 Δ_3 Δ_4	ε	Frequency (k)	Relative Frequency = Probability P
12	- - - +	-2δ	4	4/16 = 0.25
13	- - + -			
14	- + - -			
15	+ - - -			
16	- - - -	-4δ	1	1/16 = 0.0625

$\longrightarrow$ **The larger the deviation (“error”) the smaller the probability**

Pascal Triangle

$\varepsilon = \sum\Delta_i$	-6δ	-5δ	-4δ	-3δ	-2δ	$-\delta$	0	$+\delta$	$+2\delta$	$+3\delta$	$+4\delta$	$+5\delta$	$+6\delta$	$\sum k$
$q=1$						1		1						2
$q=2$					1		2		1					4
$q=3$				1		3		3		1				8
$q=4$			1		4		6		4		1			16
$q=5$		1		5		10		10		5		1		32
$q=6$	1		6		15		20		15		6		1	64

Formula for all cases $\boxed{\varepsilon = (q - 2 \cdot w) \cdot \delta}$ with $w = 0, 1, 2, \ldots, q$

$q = 1$

$w = 0$	$\varepsilon = (1 - 2 \cdot 0) \cdot \delta$	$+\delta$
$w = 1$	$\varepsilon = (1 - 2 \cdot 1) \cdot \delta$	$-\delta$

$q = 2$

$w = 0$	$\varepsilon = (2 - 2 \cdot 0) \cdot \delta$	$+2\delta$
$w = 1$	$\varepsilon = (2 - 2 \cdot 1) \cdot \delta$	0
$w = 2$	$\varepsilon = (2 - 2 \cdot 2) \cdot \delta$	-2δ

5.1 Theory of Elementary Errors

General Law $$P(\varepsilon) = P\{\varepsilon = (q-2w)\delta\} = \frac{1}{2^q}\binom{q}{w}, \quad \text{with} \quad \binom{q}{w} = \frac{q!}{(q-w)!\,w!}$$

factorial

Please note: $$\sum_{w=0}^{q}\binom{q}{w} = 2^q \quad \text{and} \quad \sum_{w=0}^{q} P(\varepsilon) = \frac{1}{2^q}\sum_{w=0}^{q}\binom{q}{w} = 1$$

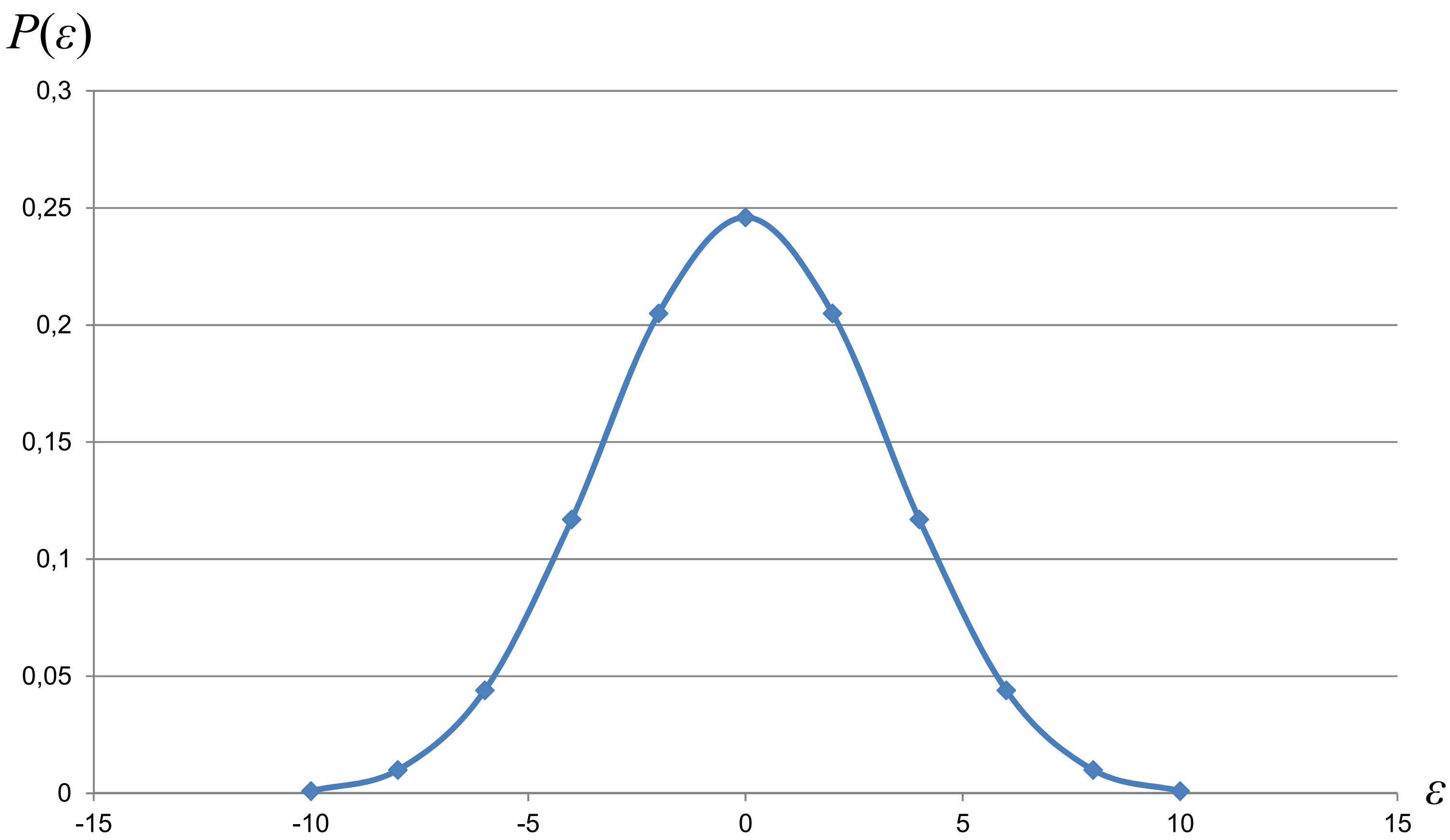
q = 10 We can now directly compute ε_i and $P(\varepsilon_i)$, w = 0, 1, 2, …, 10

w	0	1	2	3	4	5	6	7	8	9	10
ε	10δ	8δ	6δ	4δ	2δ	0	-2δ	-4δ	-6δ	-8δ	-10δ
$P(\varepsilon)$	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

0.377 0.246 0.377

all sum up to =1

Graphical representation for q = 10 .



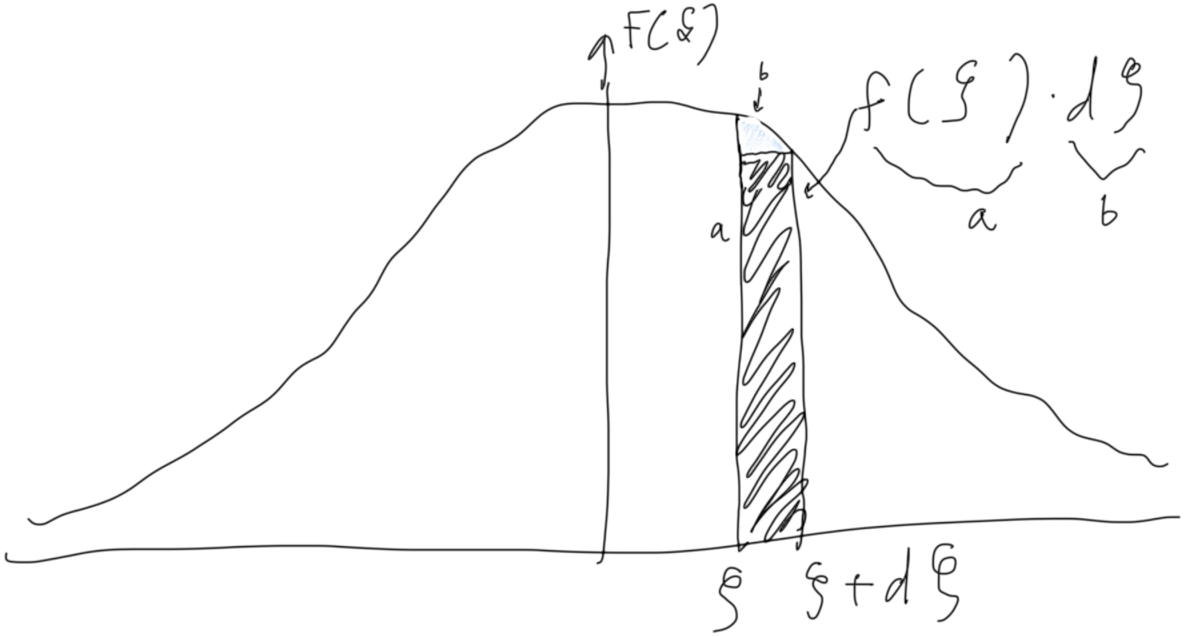
5.1 Theory of Elementary Errors

getting big ✓ getting small

Now: $q \to \infty \;\Rightarrow\; \delta \to 0$

✓ Continuous bell curve function

$$P\left(\xi \le \varepsilon < \xi + d\xi\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2\sigma^2}\right) d\xi = f\left(\xi\right) d\xi$$



Definition

A random variable is normally distributed if it can be described by a sum of a large amount of "elementary errors" with (more or less) same magnitude.

It is: $\varepsilon = \Delta_1 + \Delta_2 + \ldots + \Delta_n$

With: ε random deviation

Δ_i elementary components of the deviation

$E(\Delta_i) = 0 \quad \forall_i \leftarrow$ expectation

$n \to \infty$

<u>no</u> Δ_i is dominating

Comparison with BESSEL and HAGEN:

– All values Δ_i have the same absolute value $\Delta_i = |\delta| \quad \forall_i$

– All values Δ_i differ only in their sign $\varepsilon = \pm\delta \pm \delta \pm \delta \pm \ldots$

– Positive and negative signs have same probability $P(\Delta_i = +\delta) = P(\Delta_i = -\delta) = 0.5$

For $n \to \infty \;\Rightarrow\; \delta \to 0$

$\Rightarrow$ Then it holds:

$$\boxed{P(\xi \le \varepsilon < \xi + d\xi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2 \cdot \sigma^2}\right) d\xi = f(\xi)\, d\xi}$$

With σ = Constant value, to be defined (tbd)

$f(\xi)$ = Density of normal distribution

History of Central Limit Theorem (CLT)

de Moivre 1730
Laplace 1892
} First ideas, no proof

Gauss 1820 introduction into error theory (→ “Gauss Distribution”)

Lyapunov 1901 Proof!

In practice: If many “elementary errors” have (more or less) the same magnitude, we can ***assume*** a normal distribution.

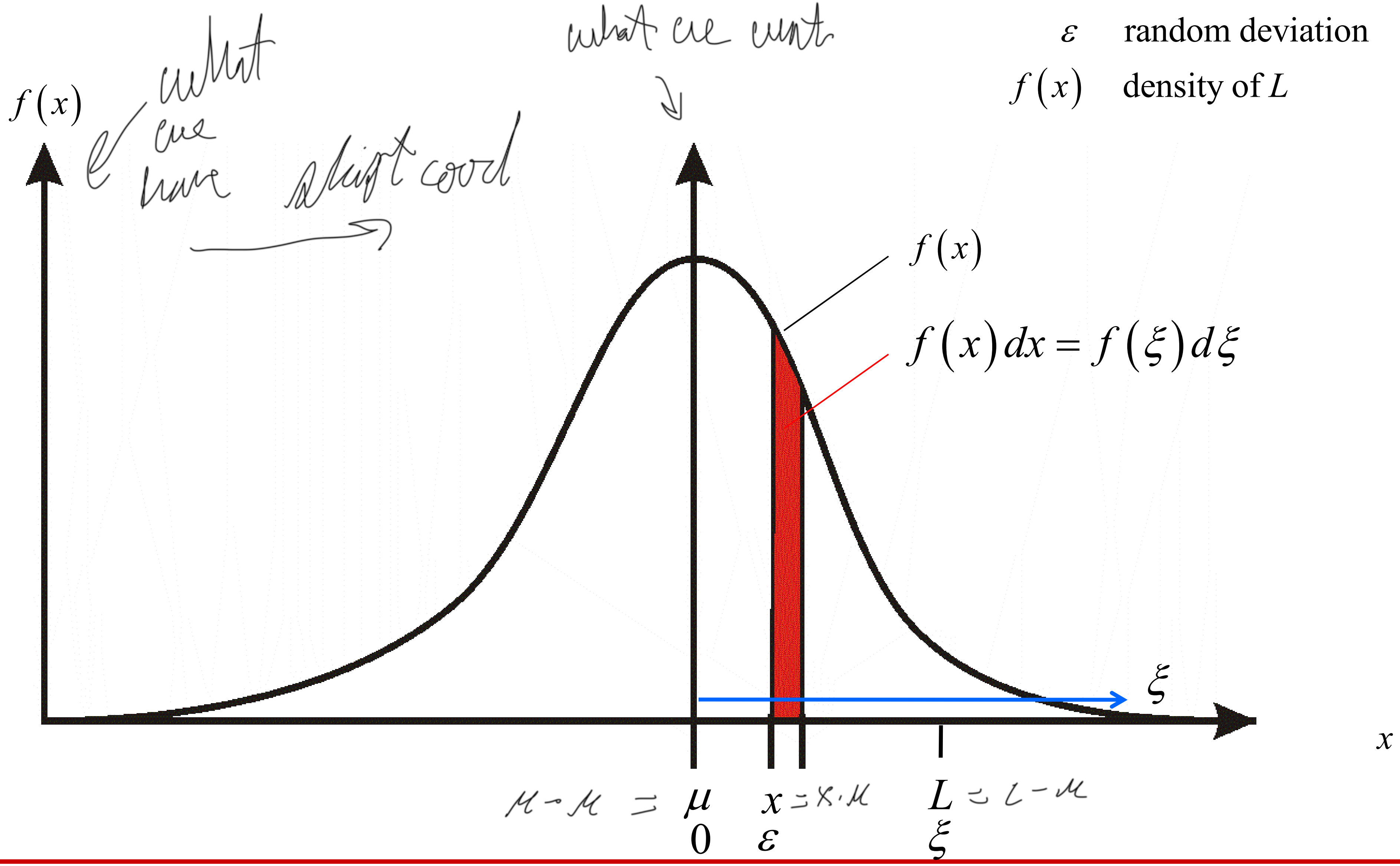
5.2 Expectation and Variance of a normally distributed random variable

<u>Given</u>: random variable $L = \mu + \varepsilon \rightarrow \varepsilon = L - \mu$

with: μ expectation $E(L)$

ε random deviation

$f(x)$ density of L



Shifting the coordinate system:

$$x = \mu + \xi \rightarrow \underbrace{\xi = x - \mu} \rightarrow d\xi = dx \rightarrow f(x) = f(\xi)$$

Substitution yields

$$P(\xi \leq \varepsilon < \xi + d\xi) = \underbrace{f(\xi)}_{f(x)} \underbrace{d\xi}_{dx}$$

$$P(x - \cancel{\mu} \leq L - \cancel{\mu} < x - \cancel{\mu} + dx)$$

$$P(x \leq L < x + dx) = f(x)\,dx$$

Normal distribution from Central Limit Theorem: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{\xi^2}{\sigma^2} \right)$

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

$$P(x \le L < x + dx) = \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right)}_{f(x)} dx$$

Probability density and thus the distribution function of a normally distributed random variable is completely defined by two parameters: Expectation μ and variance σ^2 !!

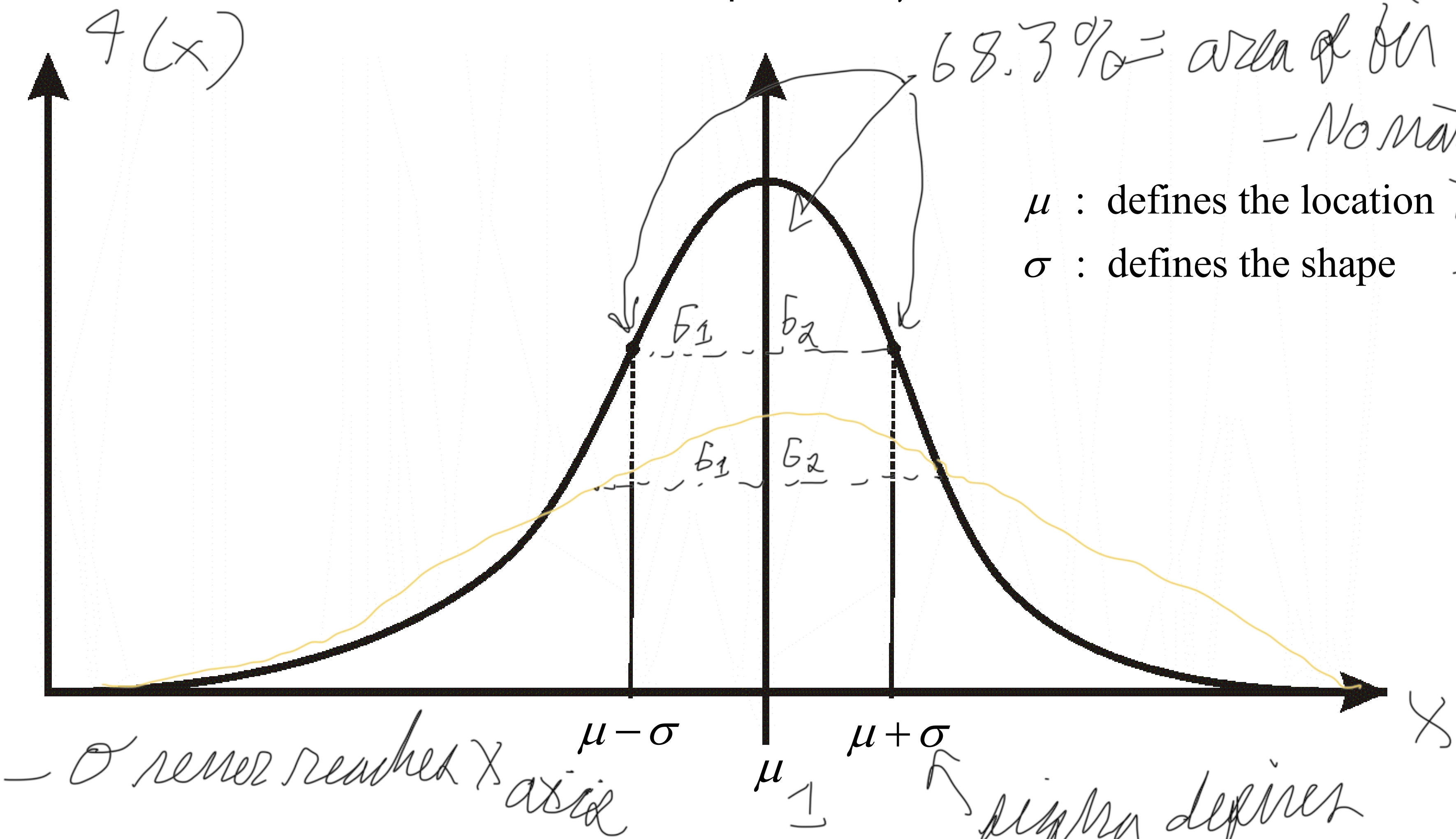
$$\boxed{L \sim N(\mu, \sigma^2)}$$

Variance

follows? Expectation

"L follows the normal distribution with expectation μ and variance σ^2 "

68.3% = area of bin = always!
- No matter the shape



μ : defines the location

σ : defines the shape

- σ never reaches x axis

sigma defines shape

3 sigma rule
if σ exceeds 3σ = 99.7% of area = its a error

Adjustment Theory I

Chapter 5: The Gaussian or Normal Distribution

Prof. Dr.-Ing. Frank Neitzel