

Messunsicherheiten

1 Vorbemerkungen

Misst man eine (physikalische) Größe x mehrmals unter vermeintlich gleichen Bedingungen, so erhält man im Allgemeinen verschiedene Messwerte

$$x_1, x_2, \dots, x_n, \dots, x_N, \quad (1)$$

die vom unbekanntem wahren Wert x_w mehr oder weniger stark abweichen und somit eine Messunsicherheit nach sich ziehen.¹⁾ Mit der sogenannten *Fehlerrechnung* bestimmt man aus der Messreihe (1) ein Intervall, in welchem x_w mit einer gewissen Wahrscheinlichkeit liegt, man nennt dies *Intervallschätzung*.

Im Folgenden wird zwischen drei Arten von Messabweichungen bzw. –unsicherheiten unterschieden: *systematische*, *grobe* und *zufällige*. Sie sind im Allgemeinen in eben dieser Reihenfolge zu behandeln, was hier in kompakter Form dargestellt wird. Die beschriebenen Verfahren haben tiefergehende theoretische Begründungen, wozu es umfangreiche Literatur gibt ([1, 2, 3] u.v.a.m). Darüber hinaus wurden in jüngerer Zeit durch Einrichtungen wie das Deutsche Institut für Normung e.V. (DIN) Leitfäden zur Behandlung von Messunsicherheiten entwickelt.²⁾

¹⁾ Hier wird angenommen, dass ein einzelner wahrer Wert existiert und prinzipiell beliebig genau gemessen werden kann. Dies ist jedoch idealisierend. So kann man beispielsweise die Länge eines Stabes nicht als eine wohl bestimmte Größe ansehen, wenn die Oberfläche rau ist. Darüber hinaus gibt es Messwerte, die prinzipiell keinen einzelnen wahren Wert haben, wie zum Beispiel die Lebensdauer eines Atoms in einem stofflich homogenen radioaktiven Präparat. Hier würde nur die mittlere Lebensdauer vieler Atome einen wohl bestimmten Wert haben.

²⁾ Die Normierungs-Bemühungen des DIN bez. der Behandlung von Messunsicherheiten sind eingebettet in jene verschiedener internationaler Institutionen, wie z. B. die *Internationale Organisation für Normung* (ISO). Eine Gruppe solcher Institutionen hat u.a. den frei zugänglichen GUM-Leitfaden [4] herausgegeben. Darin werden eine Vielzahl von Empfehlungen für Bezeichnungen und Verfahren aufgeführt. Nach dem DIN sagt man beispielsweise *Messunsicherheit* statt *-fehler*, oder kurz nur *Unsicherheit*. Solche werden in die zwei Typen A und B unterteilt, wobei allein Typ-A-Unsicherheiten durch eine statistische Analyse berechnet werden. Dies entspricht also hier den zufälligen Unsicherheiten. Die Leitfäden werden immer wieder kritisch gesichtet und gegebenenfalls überarbeitet, sie gehen weit über die hier behandelten Verfahren hinaus. Ziel ist es letztlich, Messungen quantitativ zu bewerten, mit einem einheitlichen Regelwerk und somit allgemein nachvollziehbar. Allerdings werden diese Vorschläge insbesondere von Physikern nur recht zögerlich aufgegriffen. Auch diese kurze Abhandlung folgt nicht allen GUM-Empfehlungen, insbesondere nicht jener zu systematischen Abweichungen, für welche die Umrechnung der Messgeräte-Präzision in Werte einer Standardabweichung vorgeschlagen wird.

2 Systematische Abweichungen

Systematische Messabweichungen treten auf, wenn eine Größe regelmäßig entweder zu klein oder zu groß gemessen wird. Das ist beispielsweise der Fall, wenn der Nullpunkt eines Maßstabes um den Wert Δ_{sys} verschoben ist (Fehljustierung). Zur Korrektur muss dieser Wert von einem jeden Messwert abgezogen werden, und mit den transformierten Werten

$$y_n = x_n - \Delta_{\text{sys}}$$

weitergerechnet werden. Andere systematische Abweichungen treten durch Verzerrungen der Messskale auf (Fehlkalibrierung). Beispielsweise müssen bei einer um den Faktor a gestauchten Skale alle Messwerte mit dem inversen Faktor a^{-1} multipliziert werden,³⁾

$$y_n = a^{-1} \cdot x_n.$$

Eine Fehljustierung wie auch lineare Skalen-Verzerrungen können auch erst nach den groben und zufälligen Abweichungen korrigiert werden, indem entsprechende Korrekturen des Mittelwertes (Verschiebung und Entzerrung) und der Standardabweichung (nur Entzerrung) vorgenommen werden (s.u.). Streuen die Messwerte jedoch im relevanten Messbereich nichtlinear verzerrt, so müssen systematische Abweichungen zunächst so gut wie möglich eliminiert werden, wozu die (nichtlineare) Kalibrierkurve bekannt sein muss. Sind die Verzerrungen nicht allzu groß, so gibt der Hersteller eines Messgerätes mit der Präzision die größtmögliche systematische Messabweichung im gesamten Messbereich an, bei einem rein mechanischen Messschieber ist dies üblicherweise 0,05 mm. Dabei bleibt jedoch unklar, ob ein bestimmter Wert systematisch zu groß oder zu klein gemessen wird.

Kann man eine Größe nur indirekt messen, so muss man nach einer bestimmten Formel aus dem Messwert die gesuchte Größe berechnen. Diese Formel entstammt einer mehr oder weniger komplizierten theoretischen Überlegung, bei der man (immer) gewisse Näherungen machen muss (Modellbildung). Darüber hinaus können Konstanten eine Rolle spielen. Würde man beispielsweise auf der Erdoberfläche die Masse m eines Körpers aus dem gemessenen Betrag G seines Gewichtes bestimmen, so muss der Betrag g der Erdbeschleunigung am Messort bekannt sein, $m = G/g$. Dabei hängt der effektive g -Wert von der Verteilung der Erdmasse ab, wegen der Fliehkräfte in Folge der Erdrotation aber auch vom Breitengrad. Die Unsicherheit von g trägt dann zur Messunsicherheit von m bei. Verwendet man für alle Messwerte G_n den gleichen Wert g , so wird $m_n = G_n/g$ systematisch zu groß angegeben, wenn g die wahre Beschleunigung unterschätzt.

³⁾ Eine verzerrte Längenskale entsteht, wenn man auf einem PC ein Lineal zeichnet und dieses mit einer falschen Einstellung druckt, beispielsweise 620 statt 600 Pixel pro 2,54 cm (600dpi). Dann würde der Maßstab um den Faktor $a = 600/620$ gestauch sein.

3 Verteilung der Messdaten

Die unten beschriebenen Verfahren hängen wesentlich von der (unbekannten) Wahrscheinlichkeitsverteilung der Messdaten ab. Durch die Konstruktion der (empirischen) Verteilungsfunktion $F^*(x)$ erhält man hiervon eine bildliche Vorstellung. Für die Daten in Tab. 1 ist dies exemplarisch in Abb. 1 ausgeführt. F^* ist eine stückweise konstante, monoton wachsende Funktion (Treppenfunktion), mit Sprüngen der Höhe $1/N$ an den Stellen der Messwerte. Es gelten $F^*(x) = 0$ für $x < x_{\min}$ und $F^*(x) = 1$ für $x \geq x_{\max}$. Somit gibt $F^*(x)$ näherungsweise die Wahrscheinlichkeit an, dass ein Messwert kleiner oder gleich x ist.⁴⁾ Stammen die Messwerte aus einem Kontinuum, so werden die Sprunghöhen mit wachsender Anzahl kleiner, asymptotisch ($N \rightarrow \infty$) würde man eine stetige Funktion $F(x)$ erhalten. Praktisch ist dieser Übergang jedoch nicht möglich, man hat hier mit F^* nur eine Schätzung von F .⁵⁾ Dies liegt auch daran, dass man Messwerte immer nur als ganzzahlige Vielfache der Präzision angibt, beispielsweise für die Daten in Tab. 1 als Vielfache von 0,05 mm. Somit kann es bei einem beschränktem Messwertebereich immer nur endlich viele verschiedene Messwerte geben.⁶⁾

Eine noch übersichtlichere Veranschaulichung der Messwertverteilung gelingt mit dem *Boxplot* (Abb. 1, rechts). Dazu werden zunächst α -Quantile q_α eingeführt, für die $\alpha \approx F^*(q_\alpha)$ gilt. Um Quantile einer (endlichen) Messreihe (1) für beliebige reelle Werte α , mit $0 \leq \alpha \leq 1$, eindeutig zu definieren, ordnet man die Werte nach ihrer Größe,

$$\tilde{x}_1 < \tilde{x}_2 < \dots < \tilde{x}_n < \dots < \tilde{x}_N . \quad (2)$$

⁴⁾ Hiermit wird die Funktion F^* rechtsseitig stetig definiert, andere Autoren definieren sie auch linksseitig stetig.

⁵⁾ Die theoretische Statistik stellt verschiedene Verfahren zur Prüfung bereit, ob eine empirische Verteilung mit einer theoretischen verträglich ist (Anpassungstests).

⁶⁾ Gleiches trifft zu, wenn man Messwerte zur Weiterverarbeitung mit einem Digitalrechner quantisiert. Den kleinsten Digitalisierungsschritt nennt man dann *Quantisierungsrauschen* (engl.: *observational noise*).

Tab. 1: Messreihe aus 10 Längenmessungen mit einem Messschieber der Präzision 0,05 mm. Der kleinste Messwert $x_{\min} = x_4$ hat den Rang $r_4 = 1$, der größte $x_{\max} = x_7$ hat Rang $r_7 = 10 = N$. Die beiden Quartile sind $q_u = x_8$ und $q_o = x_3$. Die Daten stammen aus einer Computersimulation normalverteilter Werte, mit dem Mittelwert $\mu = 50$ mm und der Standardabweichung $\sigma = 0,5$ mm, und einer Rundung auf ganzzahlige Vielfache der Präzision. Der Ausreißer x_7 wurde willkürlich eingefügt.

n	1	2	3	4	5
r_n	4	2	8	1	5
x_n/mm	49,75	49,60	50,65	49,20	49,80
n	6	7	8	9	10
r_n	9	10	3	7	6
x_n/mm	50,75	52,30	49,65	50,60	50,30

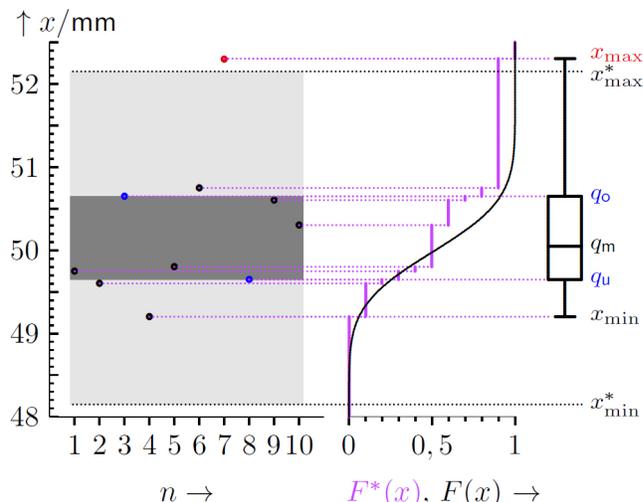


Abb. 1: Konstruktion der empirischen Verteilungsfunktion F^* (Mitte, violett: Treppenfunktion) und des Boxplots (rechts) zu 10 verschiedenen Messwerten x_n (links: Werte in Tab. 1). F^* approximiert die wahre Verteilungsfunktion F (Mitte, schwarz: Normalverteilung mit Parametern $\mu = 50$ mm und $\sigma = 0,5$ mm, s. Gln. (11, 16)). Die Daten stammen aus einer Computersimulation, weshalb F im Unterschied zur Messpraxis bekannt ist. Der Maximalwert x_7 ist ein Ausreißer und sollte verworfen werden, wohingegen der Minimalwert x_4 im Toleranzbereich $[x_{\min}^*, x_{\max}^*]$ (hellgrau) liegt. Ohne Ausreißer läge F^* dichter an F , weil die Stufen dann die Höhe $1/9$ statt $1/10$ hätten.

Somit sind der kleinste und der größte Wert $x_{\min} \equiv \tilde{x}_1$ bzw. $x_{\max} \equiv \tilde{x}_N$. Dann setzt man⁷⁾

$$q_\alpha = \begin{cases} (\tilde{x}_{\alpha N} + \tilde{x}_{\alpha N + 1})/2 & : \alpha N \text{ ganzzahlig und } < N \\ \tilde{x}_{\lceil \alpha N \rceil} & : \text{andernfalls} . \end{cases} \quad (3)$$

Sind einige Messwerte gleich groß, so modifiziert man die Definition sinngemäß. Spezielle Quantile sind der *Median*, $q_m \equiv q_{0,5}$, sowie das *untere* und *obere Quartil* $q_u \equiv q_{0,25}$ bzw. $q_o \equiv q_{0,75}$ (s. Beispiel in Abb. 1, Tab. 1).

Man nennt nun das Intervall $[q_u, q_o]$ *Box*, hier liegt etwa die Hälfte aller Messwerte, und die Intervalle $[x_{\min}, q_u]$ und $[q_o, x_{\max}]$ heißen *untere* bzw. *obere Antenne*, die jeweils etwa ein Viertel der Werte umfassen.

⁷⁾ Mit $\lceil \dots \rceil$ wird die Aufrundungsfunktion (Gauss-Klammer) bezeichnet, $\lceil \alpha N \rceil \equiv \min\{k \in \mathbb{Z} | k \geq \alpha N\}$. Beispielsweise gelten $\lceil 4,3 \rceil = \lceil 4,7 \rceil = \lceil 5,0 \rceil = 5$. In der Literatur findet man auch leicht abweichende Quantil-Definitionen.

4 Grobe Abweichungen

Durch Unachtsamkeiten des Experimentators oder andere außergewöhnliche Einflüsse können in einer Messreihe Werte auftreten, die sich extrem stark von der überwiegenden Mehrzahl der anderen Werte unterscheiden. Solche Ausreißer sollten eliminiert werden, bevor man die zufälligen Messabweichungen analysiert. Damit dies nach möglichst objektiven Kriterien erfolgt, konstruiert man den Boxplot, wie oben beschrieben.

Nun fragt man, ob die Länge der Antennen plausibel sind. Dazu muss man eine Annahme über die (unbekannte) wahre Wahrscheinlichkeitsverteilung F der Messwerte machen. Man geht häufig davon aus, dass sich die zufällige Abweichung eines Messwertes vom wahren Wert x_w aus einer Vielzahl von (ebenso unbekanntem) einzelnen Zufälligkeiten ergeben, welche sich additiv überlagern. Dann sind die Messwerte näherungsweise *normalverteilt*, mit dem Mittelwert $\mu = x_w$ (s. Anhang). Damit kann man die Wahrscheinlichkeit dafür schätzen, dass eine Antenne eine bestimmte Länge hat.

Man legt nun fest, dass x_{\max} ein *Ausreißer* ist, wenn die obere Antenne länger als das 1,5-fache der Boxbreite $q_o - q_u$ ist. Dies ist gleichbedeutend mit,

$$x_{\max} > (5q_o - 3q_u)/2 \equiv x_{\max}^* .$$

Entsprechend ist x_{\min} ein Ausreißer, falls

$$x_{\min} < (5q_u - 3q_o)/2 \equiv x_{\min}^* .$$

Die Toleranzbreite $x_{\max}^* - x_{\min}^*$ beträgt somit das 4-fache der Boxbreite $q_o - q_u$. Die Quartile der Normalverteilung sind $\approx \mu \pm 0,675\sigma$, und ein einzelner Wert wird mit der recht geringen Wahrscheinlichkeit von nur

$$F(x_{\min}^*) + (1 - F(x_{\max}^*)) = 2 \cdot F(x_{\min}^*) \approx 0,007$$

irrtümlich als Ausreißer klassifiziert. Das Verfahren ist geeignet, falls $10 \lesssim N \lesssim 100$ gilt. Für sehr viele Messwerte ($N \gtrsim 100$) würde man die Toleranzbreite sinngemäß erhöhen.

Hat man einen Ausreißer erkannt, sollte man versuchen, Ursachen für sein Auftreten zu ergründen. Erst die Aufdeckung dieser Gründe berechtigt letztlich, ihn gegebenenfalls zu verwerfen. Generell ist es selbstverständlich unzulässig, Messwerte willkürlich zu verwerfen und damit das Endergebnis in eine vorab gewünschte Richtung zu verfälschen.

5 Zufällige Unsicherheiten

Die Messwerte in (1) seien voneinander unabhängig und unter gleichen Bedingungen erhalten worden. Werte mit groben Messabweichungen seien eliminiert und systematische Abweichungen vernachlässigbar. Mit der folgenden Rechnung wird versucht, die Parameter Mittelwert μ und Standardabweichung σ der unbekanntem wahren Verteilungsfunktion F möglichst gut aus den Messwerten zu schätzen. Aus dem Schätzwert für σ kann dann die Qualität der μ -Schätzung beurteilt werden. Letztlich wird aus den Messwerten in folgenden Schritten ein Intervall berechnet, in welchem der unbekanntem wahre Wert x_w mit einer gewissen Wahrscheinlichkeit p liegt:

1. Mittelwert der Einzelmessung:

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (4)$$

2. Standardabweichung der Einzelmessung:⁸⁾

$$s_x = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (\bar{x} - x_n)^2} \quad (5)$$

3. Standardabweichung des Mittelwertes:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{N}} \quad (6)$$

4. Erweiterter Vertrauensbereich zur Konfidenz p :

$$\bar{x} \pm k_{p,N} \cdot s_{\bar{x}} \quad (7)$$

Die Werte \bar{x} und s_x sind Schätzungen des Mittelwertes und der Standardabweichung der (unbekanntem wahren) Verteilung F , für normalverteilte Messwerte sind dies μ und σ in (12) bzw. (13).⁹⁾ Der *Erweiterungsfaktor* $k_{p,N}$ hängt

⁸⁾ Die Größe s_x schätzt die unbekanntem wahre Standardabweichung σ . Würde man immer wieder eine gleich lange Messreihe auswerten, wäre s_x im Allgemeinen nicht immer gleich groß sein, also selbst zufällig. Die s_x -Werte hätten dann ebenso einen Mittelwert (Erwartungswert). Um zu erreichen, dass dieser gleich σ ist, muss man unter der Wurzel nicht durch N sondern $N-1$ dividieren (Bessel-Korrektur). Damit wird die Schätzung *erwartungstreu* (engl.: unbiased). Praktisch ist dies jedoch nur bei wenigen Messwerten ($N \lesssim 30$) von Bedeutung. Würde man in der Formel jedoch statt des Schätzwertes \bar{x} den (unbekanntem wahren) Mittelwert μ einsetzen können, wäre die Erwartungstreue bei Division durch N erreicht.

⁹⁾ Gibt es keinen einzelnen wahren Wert, so schätzt \bar{x} den Mittelwert der Verteilungsfunktion. Beispielsweise würde beim radioaktiven Zerfall die Anzahl A der nach der Zeit t zerfallenen Atome eines radioaktiven Präparates durch $A(t) = A_0 \cdot e^{-\tau t}$ gegeben sein, wobei der Wert \bar{x} die mittlere Lebenszeit $1/\tau$ schätzt. Einen einzelnen wahren Wert für die Lebensdauer eines Atoms gibt es hier jedoch nicht.

von der Anzahl N der Messwerte und der gewählten Wahrscheinlichkeit (*Konfidenz*, Zuverlässigkeit) p ab. Darüber hinaus ist die Verteilung der einzelnen Messwerte von Bedeutung. In Tab. 2 sind einige Faktoren für normal- und für gleichverteilte Messwerte angegeben. Man kann dann Folgendes aussagen:

Der Mittelwert \bar{x} liegt (etwa) mit der Wahrscheinlichkeit p im Intervall

$$[x_w - k_{p,N} \cdot s_{\bar{x}}, \quad x_w + k_{p,N} \cdot s_{\bar{x}}] .$$

Tab. 2: Erweiterungsfaktoren $k_{p,N}$ (Genauigkeit $\pm 0,01$) für normal- und gleichverteilte Messwerte, zum Einsetzen in (7) (N : Anzahl der Messwerte, p : Konfidenz). Die Werte berechnet man für normalverteilte Messwerte aus den Quantilen der sogenannten t -Verteilung zum Freiheitsgrad $N - 1$, nach der die Größe $(\bar{x} - \mu)/s_{\bar{x}}$ verteilt ist (s. Anhang B). Für $N \gtrsim 30$ ist dies in guter Näherung die Standardnormalverteilung F^* . Für gleichverteilte Messwerte (rechteckige Verteilungsdichte) stammen die Faktoren aus einer Monte-Carlo-Simulation ($3 \cdot 10^7$ Simulationen für jeden N -Wert). Für $N \rightarrow \infty$ werden die Unterschiede der Faktoren in den beiden Tabellen bedeutungslos, umso eher je kleiner p ist. Entsprechende Faktoren, die sich um mehr als 0,1 unterscheiden, sind farblich hervorgehoben.

normalverteilte Messwerte					
N	p				
	0,683	0,90	0,95	0,98	0,99
4	1,20	2,35	3,18	4,54	5,84
5	1,15	2,13	2,78	3,75	4,60
6	1,11	2,02	2,57	3,36	4,03
7	1,09	1,94	2,45	3,14	3,71
8	1,08	1,89	2,36	3,00	3,50
9	1,07	1,86	2,31	2,90	3,36
10	1,06	1,83	2,26	2,82	3,25
15	1,04	1,76	2,14	2,62	2,98
20	1,03	1,73	2,09	2,54	2,86
30	1,02	1,70	2,05	2,46	2,76
50	1,01	1,68	2,01	2,40	2,68
100	1,00	1,66	1,98	2,36	2,63

gleichverteilte Messwerte					
N	p				
	0,683	0,90	0,95	0,98	0,99
4	1,15	2,63	3,85	5,98	8,09
5	1,10	2,26	3,15	4,60	5,95
6	1,07	2,07	2,79	3,92	4,93
7	1,06	1,97	2,59	3,52	4,35
8	1,05	1,91	2,46	3,27	3,97
9	1,04	1,87	2,38	3,10	3,71
10	1,04	1,84	2,32	2,98	3,53
15	1,02	1,76	2,17	2,70	3,10
20	1,02	1,73	2,11	2,59	2,94
30	1,01	1,70	2,05	2,49	2,80
50	1,01	1,68	2,01	2,42	2,70
100	1,00	1,66	1,99	2,37	2,64

Daraus folgt, dass auch umgekehrt x_w mit der Wahrscheinlichkeit p im Intervall $\bar{x} \pm k_{p,N} \cdot s_{\bar{x}}$ liegt. Erstere Schreibweise entspricht jedoch eher den Verhältnissen, denn x_w wird als eine feste Größe angesehen, wohingegen bei Wiederholung des gesamten Messprozesses \bar{x} zufällig schwankt. Zur Vereinfachung der Schreibweise setzt man $x \equiv x_w$ und schreibt auch kürzer,

$$x = \bar{x} \pm k_{p,N} \cdot s_{\bar{x}} \quad (p \cdot 100\%) . \quad (8)$$

Den (erweiterten) Vertrauensbereich (7) nennt man auch *Konfidenzintervall*.

Mit größerer Konfidenz p wächst im Allgemeinen auch die Abweichung $k_{p,N} \cdot s_{\bar{x}}$. Die Aussage wird somit zuverlässiger, allerdings auch weniger spezifisch. Häufig wählt man $p = 0,95 = 95\%$, mit steigender Tragweite einer Fehlansage jedoch auch größer.

Für die normalverteilten Messwerte in Tab. 1 (ohne Ausreißer x_7) erhält man,

$$\begin{aligned} \bar{x} &= 50,033.. \text{ mm} \\ s_x &= 0,553.. \text{ mm} \\ s_{\bar{x}} &= 0,184.. \text{ mm} \\ k_{p,N} &= 2,31 \quad \text{für } N = 9 \text{ und } p = 0,95 \\ k_{p,N} \cdot s_{\bar{x}} &= 0,426.. \text{ mm} \end{aligned}$$

Die zufällige Messabweichung $k_{p,N} \cdot s_{\bar{x}}$ wird nun auf zwei signifikante Stellen (immer) aufgerundet und der Mittelwert auf ebensoviele Stellen (normal) gerundet. Schließlich schreibt man,¹⁰⁾

$$x = (50,03 \pm 0,43) \text{ mm} \quad (95\%) . \quad (9)$$

Erhöht man die Anzahl N von Messwerten, dann nähert sich mit hoher Wahrscheinlichkeit die empirische Standardabweichung s_x dem (unbekannten) wahren Wert σ . Somit ist s_x ein nahezu konstanter Wert, wohingegen die Standardabweichung des Mittelwertes (6) und folglich auch die Breite des Konfidenzintervalls (bei fest gewählter Konfidenz p) gegen Null streben, wenn auch nicht allzu schnell — für jede Halbierung der Breite des Konfidenzintervalls muss N immerhin vervierfacht werden. Dieser wachsende Messaufwand lohnt allerdings nur soweit, bis das Konfidenzintervall etwa die Breite der (doppelten) Präzision des Messinstrumentes hat. Jede weitere Verringerung der Intervallbreite würde eine Scheingenauigkeit vortäuschen. Beispielsweise haben die Messwerte in Tab. 1 die Präzision $\delta = 0,05 \text{ mm}$. Die Standardabweichung der Einzelmessung, geschätzt aus $N = 9$ Messwerten, ist $s_x \approx 0,55 \text{ mm}$. Folglich beträgt die maximal sinn-

¹⁰⁾ Eine kompaktere moderne Schreibweise ist,
 $x = 50,03(43) \text{ mm} \quad (95\%)$.

volle Anzahl von Messwerten $(s_x/\delta)^2 \approx 100$. Unterscheiden sich jedoch schon die einzelnen Messwerte x_n um weniger als die Messpräzision, so kann selbstverständlich auf die hier beschriebene Rechnung verzichtet werden. Dann würde die Präzision die maximale Messabweichung angeben.

Weil sowohl \bar{x} wie auch s_x (nur) Schätzwerte sind, sollte man die Konfidenz p nicht allzu genau angeben, also nicht überinterpretieren. Dies ist auch deshalb geraten, weil schon die Behandlung grober Messabweichungen eine gewisse Willkür hat. Wählt man beispielsweise die Toleranzbreite für die Daten in Tab. 1 nur geringfügig größer, so dass x_7 nicht als Ausreißer verworfen wird, dann erhält man das Ergebnis,

$$x = (50,26 \pm 0,65) \text{ mm} \quad (95\%) \quad , \quad (10)$$

welches sich recht deutlich von (9) unterscheidet. Der wahre Wert $x_w = 50 \text{ mm}$ liegt jedoch in einem jeden Fall im angegebenen Intervall mit einer Konfidenz $p > 95\%$.¹¹⁾

Literatur

- [1] Gelbrich, Götz (1998): *Statistik für Anwender* (Shaker Verlag, Aachen)
- [2] Ross, Sheldon M. (2006): *Statistik für Ingenieure und Naturwissenschaftler*, 3. Auflage (Elsevier, München)
- [3] Barlow, Roger (2002): *Statistics: a guide to the use of statistical methods in the physical sciences* (Wiley, München)
- [4] GUM–Leitfaden (1995, 2008): *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (frei zugängliches pdf-File; 1,9 MB)

¹¹⁾ Der wahre Wert $x_w = \mu = 50 \text{ mm}$ ist hier bekannt, weil die Messreihe aus einer Computersimulation stammt (s. Tab. 1), mit der Standardabweichung $\sigma = 0,5 \text{ mm}$. Folglich hat der (erwartungstreue) Mittelwertschätzer \bar{x} den Mittelwert μ und die Standardabweichung σ/\sqrt{N} . Somit kann die tatsächliche Konfidenz für die Intervalle in (9) und (10) mit 98,89..% bzw. 95,91..% berechnet werden.

A Normalverteilung

Die zufälligen Schwankungen von Messwerten sind oftmals *normalverteilt*. Das bedeutet, ihre *Wahrscheinlichkeitsdichte* ist durch die *gaußsche Glockenkurve* gegeben (Abb. 2),

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (11)$$

Die beiden Parameter sind der *Mittelwert*,

$$\mu = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad (12)$$

und die *Standardabweichung*

$$\sigma = +\sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x) dx} \quad (13)$$

Die Größen μ und σ haben dieselbe Maßeinheit wie x , wohingegen die Dichte in der inversen Einheit anzugeben ist. Die Dichte ist symmetrisch um den Mittelwert, $f(\mu+x) = f(\mu-x)$. Wegen dieser Symmetrie sind Mittelwert und Median gleich, was jedoch bei asymmetrischen (sogenannten schiefen) Verteilungen nicht der Fall ist.

Die Wahrscheinlichkeit, dass ein Messwert x_n im Bereich I liegt, berechnet man aus,

$$W(I) = \int_I f(x) dx \quad (14)$$

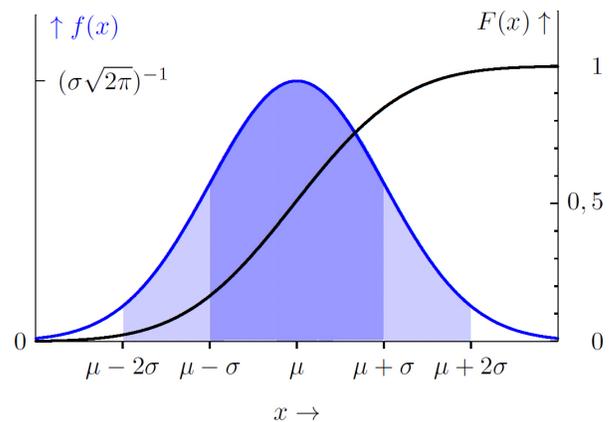


Abb. 2: Wahrscheinlichkeitsdichte f , Gl. (11), und zugehörige Verteilungsfunktion F , Gl. (16), der Normalverteilung, mit dem Mittelwert μ und der Standardabweichung σ . Die Dichte ist an der Stelle $x = \mu$ maximal, und bei $x = \mu \pm \sigma$ liegen Wendepunkte vor. Hingegen ist die Verteilungsfunktion hier streng wachsend, von Null bei $x = -\infty$ zu Eins bei $x = +\infty$.

Für $I =] - \infty, +\infty[$ erhält man $W(I) = 1$ (sicheres Ereignis). Wählt man ein Intervall symmetrisch um den Mittelwert, $I = \mu \pm k \cdot \sigma$, so gilt,

$$W(I) \approx \begin{cases} 0,6826 & : k = 1 \\ 0,9546 & : k = 2 \\ 0,9974 & : k = 3 \end{cases} \quad (15)$$

Das Integral $W(I_x)$ über das Intervall $I_x =] - \infty, x]$ hängt von der oberen Intervallgrenze x ab. Diese Funktion

$$F(x) = \int_{-\infty}^x f(y) dy \quad (16)$$

nennt man *Verteilungsfunktion*, Abb. 2. Folglich ist $F(x)$ die Wahrscheinlichkeit, dass ein Messwert kleiner als x ist.

Mit der Transformation $z = (x - \mu)/\sigma$ erhält man eine standardisierte Größe, die den Mittelwert null und die Standardabweichung eins hat. Ihre Verteilungsfunktion $f^*(z)$ heißt *gaußsches Fehlerintegral*.¹²⁾ Die entsprechende Dichte ist

$$f^*(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (17)$$

Für $\sigma > 0$ wird jedem Intervall von Maßzahlen eine positive Wahrscheinlichkeit zugeordnet, also zum Beispiel auch einem Intervall im negativen Bereich. Offenbar ist dies für viele Messgrößen, wie zum Beispiel Längen oder Massen, die nur positive Maßzahlen haben, bedeutungslos. Dies verdeutlicht, dass die Normalverteilung nur eine Approximation der wahren Verteilung sein kann.

Dennoch ist die Normalverteilung für die Fehlerrechnung von grundlegender Bedeutung. Nach dem zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung beschreibt sie die zufälligen Schwankungen einer Messgröße immer dann recht gut, wenn sie sich aus der additiven Überlagerung vieler untereinander unabhängiger¹³⁾ zufälliger Störgrößen ergeben, was in der Messpraxis oftmals der Fall ist. Dabei ist es erstaunlicherweise nahezu gleichgültig, wie die einzelnen Störungen verteilt sind.¹⁴⁾ Sind die einzelnen Messwerte nicht normalverteilt, so ist wiederum nach dem

¹²⁾ Das Fehlerintegral kann nicht für alle z -Werte analytisch berechnet werden. In der einschlägigen Literatur wird es deshalb tabelliert angegeben. Häufig werden Werte der Funktion $\Phi(x) = F^*(z) - 1/2$ für $x \geq 0$ aufgeführt. Wegen der Symmetrie $f^*(x) = f^*(-x)$ gilt $F^*(x) = 1 - F^*(-x)$ für $x \leq 0$.

¹³⁾ Die Unabhängigkeit der Messwerte bedeutet, vereinfacht ausgedrückt, dass ein Messwert nicht vom Ergebnis vorheriger Messungen beeinflusst wird. Das bedeutet, dass vergangene Messwerte keine Information über zukünftige Werte liefern. Formal drückt man es so aus: Die Verbundwahrscheinlichkeit für eine Messreihe ist das Produkt der Wahrscheinlichkeiten der einzelnen Messwerte. Für die Messpraxis bedeutet dies, dass alle relevanten Einstellungen des Messgerätes, also der Messablauf zur Gewinnung eines jeden Messwertes, immer wieder auf gleiche Weise erfolgen muss.

¹⁴⁾ Der Grenzwertsatz ist gültig, wenn sowohl der Mittelwert wie auch die Standardabweichung einer jeden Störgröße endlich sind.

Grenzwertsatz der Mittelwertschätzer (4) für eine wachsende Messwertzahl N näherungsweise normalverteilt, mit der Standardabweichung (6).

B t-Verteilung

Bei der Analyse der zufälligen Messunsicherheiten berechnet man den (empirischen) Mittelwert nach (4) und die (empirische) Standardabweichung des Mittelwertes nach (6). Werden systematische Messabweichungen ausgeschlossen, so ist \bar{x} eine Schätzung des (unbekannten) wahren Wertes $x_w = \mu$. Hierbei ist μ der Mittelwert der (unbekannten) Verteilung des Messwertes. Würde man immer wieder neue Messreihen gleicher Länge gewinnen, so erhält man aus einer jeden dieser Reihen gewisse Werte für \bar{x} wie auch für $s_{\bar{x}}$. Man fragt nun nach der Wahrscheinlichkeitsverteilung der auf $s_{\bar{x}}$ bezogene Abweichung $\bar{x} - \mu$, des sogenannten *t-Wertes*

$$t \equiv \frac{\bar{x} - \mu}{s_{\bar{x}}},$$

wenn \bar{x} wie auch $s_{\bar{x}}$ aus der selben Messreihe berechnet werden.

Unter den recht allgemeinen Voraussetzungen des zentralen Grenzwertsatzes sind die einzelnen Messwerte normalverteilt. Man kann dann die Verteilungsdichte des *t-Wertes* berechnen,¹⁵⁾

$$f_{N-1}(t) = \frac{1}{\sqrt{\pi(N-1)}} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \left(1 + \frac{t^2}{N-1}\right)^{-N/2}.$$

Darin bezeichnet Γ die Gamma-Funktion, welche für jede ganze Zahl $n > 0$ wie folgt definiert ist,

$$\Gamma(n+1) = n! \quad , \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{n! 4^n} \sqrt{\pi}.$$

Diese Dichte hängt von der Anzahl N der Messwerte ab, $N - 1$ nennt man Freiheitsgrad, Abb. 3. Für jeden Freiheitsgrad sind sie symmetrisch um den Mittelwert null, $f_{N-1}(t) = f_{N-1}(-t)$. Die Standardabweichung beträgt,

$$\sigma_{N-1} = \sqrt{\frac{N-1}{N-3}}, \quad \text{falls } N > 3.$$

Für endlich viele Störgrößen ist die Messgröße nahezu normalverteilt, wenn keine einzelne dieser Störgrößen dominant ist, wenn also jede einzelne Standardabweichung klein gegenüber der Summe aller Standardabweichungen ist, s.z.B. [4], S. 71.

¹⁵⁾ Die *t*-Verteilung wird auch *Student-Verteilung* genannt, nach dem englischen Statistiker W. S. GOSSET, der sie als erster auf empirischem Weg fand und 1901 unter dem Pseudonym *Student* publizierte.

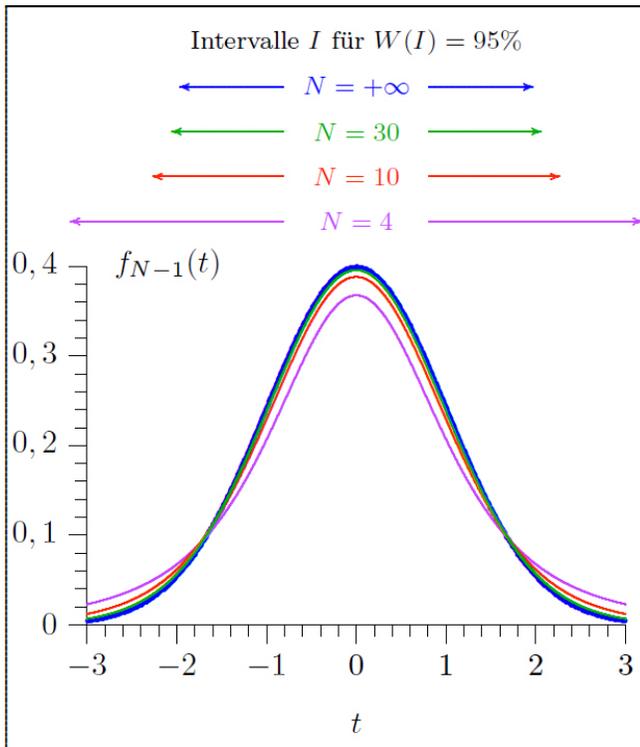


Abb. 3: Wahrscheinlichkeitsdichte $f_{N-1}(t)$ der t -Verteilung für verschiedene Freiheitsgrade $N - 1$. Für kleinere Werte N ist die Dichte flacher und das mittlere Intervall I mit der Wahrscheinlichkeit $W(I) = 95\%$ entsprechend breiter. Für $N = 30$ (rote Kurve) und alle noch größeren Freiheitsgrade sind die Unterschiede zwischen t - und Standardnormalverteilung (blaue Kurve) oftmals vernachlässigbar.

Für $N < +\infty$ gilt somit $\sigma_{N-1} > 1$. Für viele Messwerte N geht $f_{N-1}(t)$ in die Dichte der Standardnormalverteilung $f^*(t)$ über, Gl. (17), und folglich gilt $\sigma_\infty = 1$.¹⁶⁾ Für niedrige Freiheitsgrade fällt die Dichte der t -Verteilung für kleine bzw. große t -Werte deutlich langsamer als die Normalverteilung ab.

Die *Erweiterungsfaktoren* $k_{p,N}$ in Tab. 2 (normalverteilte Messwerte, oben) erfüllen die Gleichung,

$$p = \int_{\{t : -k_{p,N} < t < k_{p,N}\}} f_{N-1}(t) dt .$$

Folglich ist $k_{p,N}$ das $(\frac{1+p}{2})$ -Quantil der t -Verteilung zum Freiheitsgrad $N - 1$. Liegt nun t mit der Wahrscheinlichkeit p im Intervall $[-k_{p,N}, k_{p,N}]$, so folgt, dass \bar{x} mit eben

¹⁶⁾ Die Größe $\bar{x} - \mu$ ist für eine beliebige Anzahl N von Messwerten normalverteilt, sofern die einzelnen Messwerte normalverteilt sind. Der t -Wert $(\bar{x} - \mu)/s_{\bar{x}}$ ist jedoch nicht normalverteilt, weil der Divisor $s_{\bar{x}}$ keine feste Größe ist. Er wird aus der gleichen Messreihe berechnet wie \bar{x} und unterliegt somit auch von Messreihe zu Messreihe zufälligen Schwankungen, allerdings nicht unabhängig vom Dividend $(\bar{x} - \mu)$. Dies führt dazu, dass der t -Wert für $N < +\infty$ nicht normal- sondern t -verteilt ist.

dieser Wahrscheinlichkeit im Intervall

$$[\mu - k_{p,N} \cdot s_{\bar{x}}, \mu + k_{p,N} \cdot s_{\bar{x}}]$$

liegt. Steigt die Anzahl der Messwerte, also auch der Freiheitsgrad, so wird die t -Verteilung der Normalverteilung immer ähnlicher, Abb.3. Für $k_{p,N} = 1; 2$ und 3 erhält man dann die Wahrscheinlichkeiten in (15). Praktisch kann man häufig ab $N \approx 30$ mit der Normalverteilung rechnen.