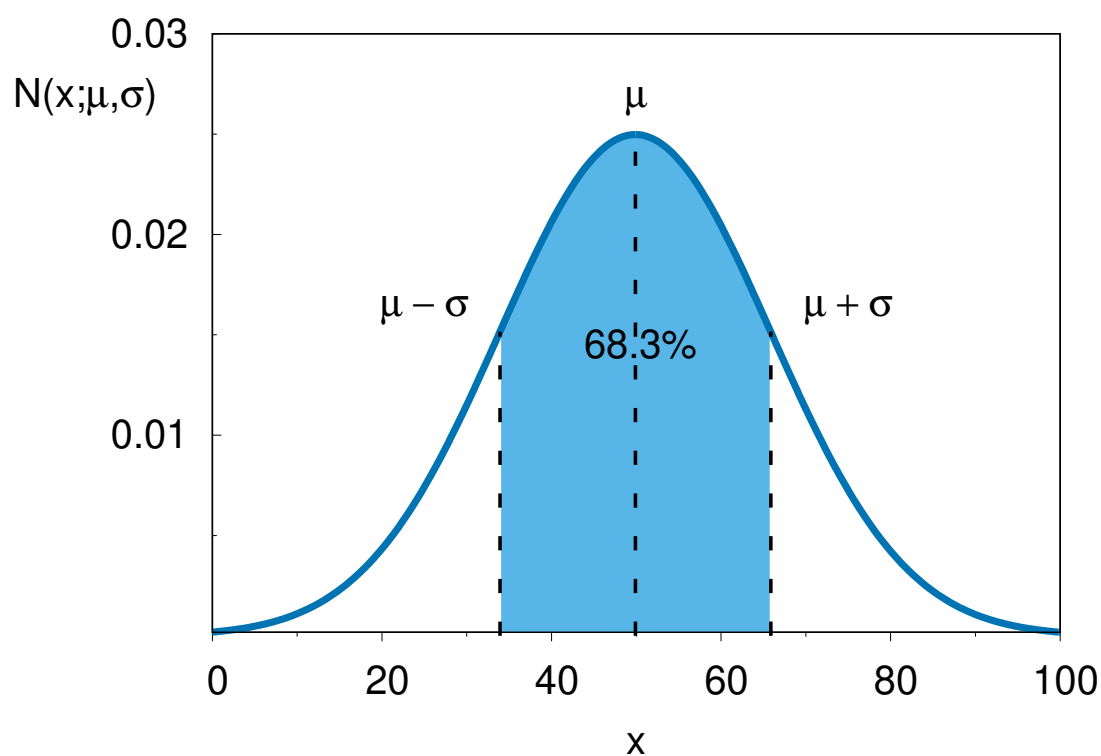

Einführung in die Messdatenanalyse für das Physikalische Grundpraktikum

Thomas Blochowicz
Jonny Birkhan
Franco Laeri
13. Oktober 2023



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Inhaltsverzeichnis

1. Datenanalyse und Messunsicherheiten	3
2. Grundlagen der konventionellen Statistik	7
2.1. Charakterisierung eines Datensatzes: Mittelwert und Streuung	7
2.1.1. Schätzung der Qualität des Mittelwertes	8
2.2. Häufigkeitsverteilungen	9
2.2.1. Allgemeines über Zufallsgrößen und deren Verteilung	9
2.2.2. Die Binomialverteilung $B_n(p)$	11
2.2.3. Die Poissonverteilung $P_\mu(k)$	12
2.2.4. Das Histogramm	13
2.2.5. Die Normalverteilung $\mathcal{N}(x; \mu, \sigma)$	15
2.2.6. Der Zentrale Grenzwertsatz (Central Limit Theorem)	16
2.3. Überdeckungsintervall und Erweiterungsfaktor	19
2.3.1. Die Student'sche t -Verteilung	20
3. Angabe von Messunsicherheiten nach GUM	22
3.1. Typ A Messunsicherheiten	22
3.2. Typ B Messunsicherheiten	22
3.3. Das Prinzip der maximalen Informationsentropie (PME)	23
3.4. Die Fortpflanzung von Messunsicherheiten	27
3.5. Angabe des vollständigen Messergebnisses mit Unsicherheit	30
4. Beurteilende Datenanalyse	31
4.1. Hypothesentests in der konventionellen Statistik	31
4.1.1. Hypothesentest in einer einfachen Situation	31
Beispiel: Das Stromnormal	32
4.2. Der χ^2 -Test: Vergleiche Theorie und Experiment	33
4.2.1. Vergleiche Datensatz mit hypothetischer Verteilung $\varphi(x)$	34
Beispiel: Mendels Erbsen	35
4.2.2. Vergleich einer Messreihe $y_i(x_i)$ mit theoretischer Voraussage $\eta(x)$	36
Beispiel: Der Zerfall von Bierschaum	37
4.3. Hypothesentests in der Bayes'schen Statistik	39
Beispiel: Das Stromnormal – Fortsetzung	41
A. Die p-Quantile der χ^2-Verteilung	46

1. Datenanalyse und Messunsicherheiten

In den Naturwissenschaften bilden Messverfahren und Messdaten die Grundlage, um quantitative Aussagen über physikalische Größen zu ermöglichen. Dabei soll der Erfahrungsprozess dem subjektiven Einfluss des beobachtenden Individuums soweit es geht entzogen, und die Erfahrung so objektiviert werden. Messdaten sind daher die Säulen, auf denen die modernen Wissenschaften stehen. Dementsprechend hoch sind die Ansprüche, die wir als Wissenschaftler an die Qualität unserer Messdaten stellen müssen.

Auf der anderen Seite wissen wir aber auch, dass moderne Messmethoden äußerst aufwändig sind. Messdaten müssen im Allgemeinen mit einem großen Einsatz von Arbeit und Kosten gesammelt werden. Von daher sind Messdaten eine kostbare Sache und es liegt in der Verantwortung jedes Wissenschaftlers, überlegt und sorgfältig mit den Daten umzugehen.

Messdaten sind dabei grundsätzlich immer mit Unsicherheiten behaftet, d. h. unsere Information über eine Messgröße ist immer in gewissem Maße unvollständig. Im alltäglichen Leben sind wir es dabei meist gewohnt implizit anzunehmen, dass auftretende Messunsicherheiten klein genug sind, um einen sinnvollen Umgang mit den jeweiligen Messgrößen zu ermöglichen (z. B. Waage beim Metzger) und in vielen Fällen wird dies auch z. B. durch Vorschriften des gesetzlichen Eichwesens sichergestellt. Im Gegensatz dazu ist jedoch eine explizite Angabe von Messunsicherheiten in den Naturwissenschaften stets *zwingend erforderlich*:

Die Angabe des Messwertes einer physikalischen Größe ist ohne Angabe der zugehörigen Unsicherheit wissenschaftlich wertlos. Die Unsicherheit beschreibt dabei die Qualität der Information, die im Messergebnis enthalten ist.

Die Unsicherheitsangabe erfolgt dabei gemäß:

$$\text{Messwert} \pm \text{Unsicherheit} \quad (1)$$

Darüber hinaus ist es wichtig, dass Unsicherheiten international auf einheitliche Weise ermittelt und angegeben werden. Nur so ist es beispielsweise möglich, Messdaten verschiedener Arbeitsgruppen untereinander zu vergleichen oder die Resultate einer neuen Theorie anhand des Pools weltweit gesammelter Daten zu einer physikalischen Größe zu überprüfen.

Der derzeit *de facto* weltweit akzeptierte Standard zur Angabe von Messunsicherheiten wird dabei durch den *Guide to the Expression of Uncertainty in Measurement* (kurz: GUM [1]) festgelegt. Er wird mittlerweile vom *Joint Committee for Guides in Metrology (JCGM)* herausgegeben, dem sowohl die *Internationale Organisation für Normung (ISO)* in Genf als auch das *Bureau International des Poids et Mesures (BIPM)* in Paris angehören.

Gemäß GUM lautet die international vereinbarte Definition des Begriffes Messunsicherheit [1]:

Messunsicherheit ist ein Parameter, der dem Messergebnis beigeordnet ist und der die Streuung derjenigen Schätzwerte kennzeichnet, die der Messgröße auf der Basis vorliegender Information vernünftiger Weise zugewiesen werden können.

Wie genau diese Zuweisung quantitativ zu erfolgen hat, ist im Folgenden zu klären. Interessant an dieser Stelle ist, dass es sich bei der Unsicherheit um eine Größe handelt, die Information charakterisiert: Sie beschreibt unsere *reale*, nach dem Stand der vorliegenden Informationen *unvollständige Kenntnis der Messgröße*.

Im Gegensatz dazu steht der Begriff der Messabweichung (früher: Messfehler). Er basiert auf der Annahme, dass einer Messgröße zumindest hypothetisch ein eindeutiger "wahrer Wert" zugeordnet werden kann.

Messabweichung bezeichnet dann die Differenz des Messwertes zu dem "wahren Wert" der Messgröße.

Messabweichungen sind deshalb ebenso wie der wahre Wert einer Messgröße grundsätzlich nie genau bekannt und können höchstens abgeschätzt werden. Als Ursachen für die Messabweichungen unterscheidet man üblicher Weise *systematische* und *zufällige* Einflüsse.

Zufällige Abweichungen zeigen sich darin, dass bei wiederholten Messungen unter den selben Bedingungen im Allgemeinen nicht der gleiche Messwert resultiert sondern zufällige Schwankungen um einen Mittelwert auftreten. Manchmal ist die Ursache dieser Abweichungen bekannt. Immer aber gilt, dass sie prinzipiell nicht eliminiert werden können. Beispiele wären Lagerreibung, Erschütterungen, elektronisches Rauschen o. ä. Man kann sogar sagen, dass an einem Messdatensatz, der keine oder zu geringe Schwankungen aufweist, etwas „faul“ sein muss, wie wir an einem Beispiel im Zusammenhang mit dem χ^2 -Test sehen werden. Durch häufiges Wiederholen der Messung und entsprechende Mittelwertbildung kann der Einfluss der zufälligen Abweichungen aber verringert werden.

Systematische Abweichungen fließen auf systematische Weise in die Messung ein, d. h. sie beeinflussen alle Messwerte in gleicher Weise. Einige Einflüsse dieser Art können durch sorgfältigeres Arbeiten ganz vermieden werden (sog. *grobe Messabweichungen*). Andere sind bekannt oder können durch Referenzmessungen bestimmt werden, so dass der Messwert korrigiert werden kann. Wieder andere Abweichungen lassen sich nur in gewissen Grenzen abschätzen und wiederum andere bleiben u. U. vollständig unerkannt und erweisen sich erst nachdem z. B. die gleiche Größe mit unterschiedlichen experimentellen Methoden untersucht wurde. Die Ursache systematischer Messabweichungen kann in der Konstruktion der Messapparatur verborgen sein, einer fehlerhaften Kalibrierung, einer unbemerkten Veränderung der Messbedingungen (z. B. Temperatur) und ähnlichen Überraschungen. Typisch ist dabei, dass sich der Einfluss systematischer Messabweichungen durch häufiges Wiederholen der Messungen und Mittelwertbildung *nicht* reduzieren lässt.

Die vorangegangene Unterscheidung zufälliger und systematischer Messabweichungen ist jedoch nicht immer eindeutig: So können systematische Messabweichungen auch statistischer Natur sein. Beispielsweise stellt die Hintergrund-Zählrate bei einem kernphysikalischen Experiment eine (korrigierbare) systematische Messabweichung dar, die hervorgerufen wird durch natürliche Hintergrundstrahlung. Möchte man in einem Experiment z. B. die Aktivität eines Präparates mit einem Geigerzähler bestimmen, so wird dabei die Zählrate *systematisch* um den Wert der Hintergrund-Zählrate zu groß ausfallen. Der physikalische Prozess, der dieser Abweichung zugrunde liegt, ist seiner Natur nach jedoch ein statistischer Prozess, und wie groß die Abweichung in einem bestimmten Zeitintervall ausfällt ist *zufällig*. Um Zweideutigkeiten solcher Art zu vermeiden, wird daher im GUM

eine etwas andere Unterscheidung getroffen. Diese bezieht sich dabei nicht auf die (unbekannte) Messabweichung sondern auf die Messunsicherheit, d. h. den Informationsgehalt der gewonnenen Daten:

Typ A Unsicherheiten sind Messunsicherheiten, die durch eine statistische Auswertung wiederholter Messungen bestimmt wurden

Typ B Unsicherheiten sind Messunsicherheiten, die auf eine andere Weise bestimmt wurden.

In diesem Sinne wären im obigen Beispiel sowohl die Unsicherheit bei der Zählrate des radioaktiven Präparates, als auch die Unsicherheit, die von der Messung der Hintergrundstrahlung herrührt, Unsicherheiten vom Typ A.

Als Grundlage zur Bestimmung von Typ B Unsicherheiten kommen unterschiedlichste Möglichkeiten in Frage, z. B.:

- Informationen aus Vormessungen, Angaben in der Literatur
- Erfahrung oder allgemeines Wissen über das Verhalten bestimmter Messgeräte
- Herstellerangaben, z. B. über Genauigkeitsklassen von Bauteilen oder Messgeräten
- Referenzdaten aus Handbüchern
- ...

Auch wenn es grundsätzlich vorzuziehen ist, Unsicherheiten nach Methode A zu bestimmen, wird man als Experimentator generell häufig darauf angewiesen sein, eine Unsicherheitsabschätzung aus nicht-statistischen Informationen nach Typ B vorzunehmen. Insbesondere Unsicherheiten aufgrund systematischer Messabweichungen lassen sich häufig nur nach Typ B einschätzen.

Wichtig ist nun, dass nur Unsicherheiten, die auf statistischen Informationen beruhen, also solche, die aus häufigen Wiederholmessungen bestimmt wurden, mit den Mitteln der konventionellen mathematischen Statistik erfassbar sind. Im Rahmen dieses Ansatzes müssten deshalb Unsicherheitsbeiträge, die anders als statistisch abgeschätzt wurden, von den statistischen Beiträgen *getrennt* angegeben und weiterverarbeitet werden. Eine einfache Angabe gemäß Glg. (1) wäre dann nicht mehr ausreichend. Insbesondere auch die *Fortpflanzung von Unsicherheiten* (s. Abschnitt 3.4) in dem Fall, dass eine Größe nur indirekt messbar ist, müsste für die statistischen und nicht-statistischen Anteile *getrennt* vorgenommen werden [2].

Im Gegensatz dazu zeigt es sich, dass, wenn man vom Wahrscheinlichkeitsbegriff der konventionellen Statistik zu dem der sog. Bayes'schen Statistik übergeht, beide Arten von Unsicherheiten auf einer gemeinsamen Basis behandelt werden können, ja sogar, dass kein wirklicher Unterschied mehr zwischen beiden Arten der Unsicherheitsermittlung besteht [1, 3, 4]. Während der Wahrscheinlichkeitsbegriff in der konventionellen *frequentistischen Statistik* auf relativen Häufigkeiten basiert,¹ charakterisiert er in der Bayes'schen Herangehensweise die *unvollständige Kenntnis* über das Eintreten von Ereignissen auf der Basis der vorliegenden Information. Diese Information kann, anders als in der konventionellen Statistik, auch nicht-statistischer Natur sein. Dies ist in Abb. 1 verdeutlicht. Eine derart gewonnene Unsicherheitsangabe kann dann, ungeachtet der Tatsache ob statistische, nicht-statistische oder beiderlei Information in ihre Bestimmung eingegangen sind,

¹Wahrscheinlichkeit eines Ereignisses ist die relative Häufigkeit des Ereignisses, ermittelt auf der Grundlage unendlich vieler Messungen unter gleichen Bedingungen.

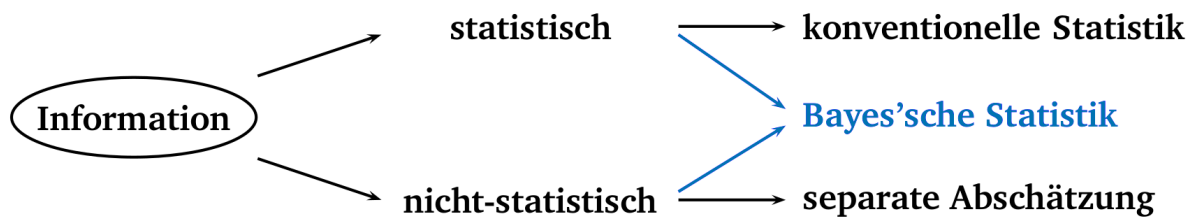


Abbildung 1: Bei der Bestimmung von Messunsicherheiten spielen meist sowohl statistische als auch nicht-statistische Informationen eine Rolle. Die konventionelle frequentistische Statistik kann dabei nur die statistischen Informationen verwerten, alle anderen Informationen müssen getrennt davon verarbeitet werden. Die Bayes'sche Statistik hingegen verwertet beide Arten von Information auf die gleiche Weise.

einheitlich gemäß Glg. (1) angegeben werden. Auch die Fortpflanzung von Unsicherheiten lässt sich dann auf eine einfache und einheitliche Weise realisieren. Aus diesem Grunde stützt sich der GUM entscheidend auf die Konzepte der Bayes'schen Statistik, insbesondere zur Ermittlung von Typ B Unsicherheiten [1]. Eine weiterführende Analyse [3] und neuere Ergänzungen zum GUM [6] zeigen dann, dass im Rahmen der Bayes'schen Statistik eine einheitliche Behandlung aller Unsicherheiten möglich ist, so dass die Unterscheidung der Unsicherheiten nach Typ A/B letztendlich hinfällig wird.

Um den Rahmen des vorliegenden Skriptes, das vorwiegend zum Gebrauch im Physikalischen Grundpraktikum gedacht ist, nicht zu sprengen, können wir die Ideen der Bayes'schen Statistik, wie sie für die Unsicherheitsbetrachtungen relevant sind, im Abschnitt 3 nur andeuten und werden uns da auf einige wenige handwerkliche Aspekte beschränken. Zuvor sollen zunächst einige Grundlagen der konventionellen mathematischen Statistik zusammengestellt werden (Abschnitt 2), bevor schließlich im Abschnitt 4 anhand weniger Beispiele ein Einblick in die sog. beurteilende Messdatenanalyse gegeben werden soll, die z. B. untersucht, ob ein theoretisches Modell mit gegebenen Messdaten verträglich ist oder nicht. Auch hier liegt zunächst die Betonung auf den Verfahren der konventionellen Statistik. Im Abschnitt 4.3 wird dann abschließend anhand eines Beispiels noch ein kurzer Ausblick auf die entsprechenden Verfahren der Bayes'schen Statistik gegeben.

2. Grundlagen der konventionellen Statistik

2.1. Charakterisierung eines Datensatzes: Mittelwert und Streuung

Wir beginnen mit der Annahme, dass die interessierende physikalische Größe X sorgfältig gemessen wurde, und nun liege der Datensatz $\{x_i\} = \{x_1, x_2, \dots, x_n\}$ vor.

Bei der Durchsicht der Messdaten kann nun auffallen, dass einzelne Werte besonders augenfällig vom Rest der anderen abweichen. In diesem Fall ist es entscheidend zu prüfen, ob der begründete Verdacht besteht, dass bei der entsprechenden Messung ein Fehler unterlaufen ist. Es wäre unsinnig, Messwerte in der Analyse zu verwenden, die sachlich begründet als fehlerhaft verdächtigt werden und so bleiben solche Werte in der Regel unberücksichtigt. Andererseits ist jede Form der Manipulation an einem Datensatz als grundsätzlich fragwürdig einzuschätzen: Ein stark abweichender Messwert kann, wenn er „echt“ ist, auch auf einen vielleicht unerwarteten physikalischen Effekt hindeuten. *In jedem Fall sind daher ermittelte Ausreißer als solche zu dokumentieren!* Für das Folgende setzen wir voraus, dass die betrachteten Datensätze von solchen Ausreißern korrekt befreit wurden.

In diesem Datensatz gibt es einen minimalen und einen maximalen Wert, x_{\min} bzw. x_{\max} zwischen denen die erhaltenen Messwerte streuen. In Rahmen der konventionellen Statistik versteht man nun den vorliegenden Datensatz als *Stichprobe* aus der Grundgesamtheit unendlich vieler Messungen der Größe X , die unter den gleichen Bedingungen durchgeführt wurden. Der unbekannte Erwartungswert oder Mittelwert dieser Grundgesamtheit, oft auch als “wahrer Wert” der Messgröße X bezeichnet, kann nun durch unseren Datensatz abgeschätzt werden. Es zeigt sich, dass der *empirische Mittelwert* des Datensatzes den besten Schätzwert dafür liefert:

$$\text{empirischer Mittelwert des Datensatzes: } \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

Die Schwankungsbreite, *Streuung*, der Daten um den Mittelwert \bar{x} wird durch ihr mittleres Abstandsquadrat zum Mittelwert, der *Standardabweichung* s charakterisiert:

$$\text{empirische Standardabweichung des Datensatzes } s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (3)$$

Das Quadrat der Standardabweichung heißt *Varianz* s^2 . Der Grund dafür, dass im Nenner von (3) der Faktor $n - 1$ anstelle von n auftaucht liegt im Prinzip darin, dass die Summe wegen des darin enthaltenen arithmetischen Mittels \bar{x} , welches die Daten x_i des Datensatzes bereits enthält, nur $n - 1$ unabhängige Summanden enthält. Eine genauere Begründung gibt es weiter unten im Abschnitt 2.2.1. Dort zeigt sich dann, dass s^2 gemäß Glg. (3) definiert, den besten Schätzwert für die Varianz σ^2 der Grundgesamtheit liefert.

Anschaulich ist zunächst klar, dass ein vertrauenswürdiger Datensatz durch eine kleine Streuung repräsentiert wird, d.h. die Standardabweichung s ist ein Maß für die Qualität des vorliegenden Datensatzes. Will man nun den Mittelwert als Schätzwert für die Messgröße, als Ergebnis angeben, müsste man sich aber eigentlich mit der *Qualität des Mittelwertes* befassen.

2.1.1. Schätzung der Qualität des Mittelwertes

Wie könnte man also die Qualität des Mittelwertes ermitteln? Warum nicht so, wie wir die Qualität des Datensatzes ermittelt haben? Also: Wir müssen uns einen Datensatz von Mittelwerten beschaffen. Nun, dies bedeutet Arbeit, denn wir müssen die ursprüngliche Messreihe wiederholen. Damit erhalten wir einen zusätzlichen Datensatz, der uns einen neuen Mittelwert liefert. Dann könnten wir diese Prozedur wiederholen und nochmals wiederholen und ... Am Ende hätten wir eine große Anzahl Datensätze mit dazugehörigen Mittelwerten vorliegen. Die so ermittelte Folge von Mittelwerten wird immer noch um den (unbekannten) Mittelwert der Grundgesamtheit ("wahrer Wert") schwanken, aber sicher in einem geringeren Maße, als das die einzelnen Datensätze tun. Die *Schwankungsbreite der Mittelwerte* $s_{\bar{x}}$ wird also kleiner sein als s und abnehmen, je mehr einzelne Datenpunkte in die Mittelung eingehen.

Die mathematische Statistik erlaubt es uns nun, auf diese umständliche Prozedur, eine Vielzahl von Messreihen durchzuführen, zu verzichten. Die Theorie sagt, dass die *Schwankungsbreite der Mittelwerte* $s_{\bar{x}}$ bereits aus einem einzigen Datensatz abgeschätzt werden kann. Es ist nämlich $s_{\bar{x}}$ mit der Standardabweichung s eines einzelnen Datensatzes auf eine einfache Weise verknüpft, wie man leicht aus der Varianz einer Summe von Zufallsgrößen ableiten kann, siehe Abschnitt 2.2.1 oder auch [7, 8]:

$$\begin{aligned} \text{Schätzwert für die Schwankungsbreite des Mittelwertes} \quad s_{\bar{x}} &= \frac{s}{\sqrt{n}} \\ s_{\bar{x}} &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}} \end{aligned} \quad (4)$$

Nun sind wir in der Lage, den besten Schätzwert der Messgröße \bar{x} zusammen mit der *Unsicherheit* anzugeben, also ein Intervall anzugeben, das die Streuung derjenigen Werte kennzeichnet, die der Messgröße sinnvoll als Schätzwerte zugewiesen werden können.

$$x = \bar{x} \pm s_{\bar{x}} \quad (5)$$

Welche genaue Bedeutung diese Angabe hat werden wir später im Zusammenhang mit der Normalverteilung und dem Zentralen Grenzwertsatz in Abschnitt 2.2.6 diskutieren.

Aufgabe:

Jeder wissenschaftliche Taschenrechner hat Funktionen für Mittelwert und Standardabweichung eingebaut. Machen Sie sich anhand des unten stehenden Beispiels mit Ihrem Modell vertraut, so dass Sie es im Praktikum bedienen können. Schauen Sie insbesondere im Handbuch Ihres Rechners nach, nach welcher Formel der Taschenrechner die Standardabweichung ausgibt. Je nach Fabrikat wird Formel (3) oder (4) angewendet. Dem entsprechend muss man eventuell die Taschenrechnerausgabe korrigieren, wenn man das Ergebnis nach (5) angibt.

Beispiel: Gegeben sei folgender Datensatz von Messwerten:

$$\{x_i\} = \{2.5, 2.8, 2.2, 2.3, 2.2, 2.7, 2.6, 2.4\}.$$

Benutzen Sie die Mittelwert-Varianz-Funktion Ihres Taschenrechners, um $x = \bar{x} \pm s_{\bar{x}}$ anzugeben.

Ergebnis: Die korrekte Taschenrechnerausgabe lautet $\bar{x} = 2.46250$, $s = 0.22638$. Eine sinnvolle Angabe des Ergebnisses ist damit:

$$x = 2.46 \pm 0.08.$$

2.2. Häufigkeitsverteilungen

Nun ist es an der Zeit, sich zu überlegen, in welcher Weise die Daten in einem Datensatz schwanken, d. h. wie sie um den Mittelwert herum verteilt sind. Diese Verteilung wird sicher davon abhängen, von welcher Art die Messgröße ist, ob sie kontinuierliche Werte annehmen kann, oder ob sie nur zwei diskrete Werte annehmen kann, z. B. Teilchen registriert, ja oder nein. Zunächst schicken wir einige allgemeine Bemerkungen voran, die für alle Verteilungen gelten. Dann besprechen wir die Wahrscheinlichkeitsverteilung eines Merkmals mit komplementären Ausprägungen, nämlich die *Binomialverteilung* und die *Poissonverteilung*. Im Falle einer kontinuierlich veränderlichen Messgröße hingegen wird der Experimentator sich zunächst ein graphisches Bild der Datenverteilung machen wollen. Dazu dient das *Histogramm*. Anschließend behandeln wir die Normalverteilung und den Zentralen Grenzwertsatz, der die Universalität dieser für den wissenschaftlichen Alltag so bedeutenden Verteilungsfunktion begründet.

2.2.1. Allgemeines über Zufallsgrößen und deren Verteilung

Alle der nun folgenden Details lassen sich bestens in [7] oder auch [8] nachlesen, auch wenn man im Einzelfall vielleicht etwas danach stöbern muss. Hier also eine Zusammenfassung der für uns wichtigen Aspekte:

- Eine Verteilung heißt *diskret*, wenn es eine endliche, oder abzählbar unendliche Menge von Werten x_i gibt, die mit einer Wahrscheinlichkeit $P(x_i) > 0$ realisiert werden, wobei gelten muss $\sum_i P(x_i) = 1$.
- Eine Verteilung heißt *stetig*, wenn eine Funktion $f(x)$, ($f(x) \geq 0, \forall x$) mit $\int_{-\infty}^{\infty} f(x) dx = 1$ existiert, für die gilt: $P(a < X < b) = \int_a^b f(x) dx$ gibt die Wahrscheinlichkeit dafür an, bei einer Messung einen Wert für X zu erhalten, der im Intervall (a, b) liegt. $f(x)$ heißt dann Wahrscheinlichkeitsdichte (auch PDF: probability density function) der stetigen Verteilung. Man beachte, dass für stetig verteilte Zufallsgrößen $P(x) = 0, \forall x$ gilt.
- Die Summenfunktion einer Wahrscheinlichkeitsdichteverteilung heißt *Verteilungsfunktion* oder *kumulative Wahrscheinlichkeitsverteilung* $F_X(x)$ der Zufallsvariablen X . Die Verteilungsfunktion beschreibt die Wahrscheinlichkeit dafür, einen Messwert unterhalb einer wählbaren Schranke x zu finden:

$$F_X(x) = P(X < x) = \sum_{x_i < x} P(x_i) \quad \text{für diskrete Verteilungen} \quad (6)$$

$$F_X(x) = P(X < x) = \int_{-\infty}^x f(\xi) \cdot d\xi \quad \text{für stetige Verteilungen} \quad (7)$$

- Zwei Zufallsgrößen X und Y heißen *statistisch unabhängig* wenn die Verteilungsfunktion der zweidimensionalen Zufallsvariablen (X, Y) gleich dem Produkt der einzelnen Verteilungsfunktionen ist, also:

$$F(x, y) = F_X(x) \cdot F_Y(y) \quad (8)$$

gleiches gilt in diesem Fall auch für die Wahrscheinlichkeitsdichten:

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad (9)$$

oder auch die Wahrscheinlichkeiten, im Falle diskreter Zufallsgrößen: Die Wahrscheinlichkeit, dass gleichzeitig der Fall $X = x_i$ und $Y = y_j$ eintritt ist dann gleich dem Produkt aus den Einzelwahrscheinlichkeiten:

$$P(X = x_i, Y = y_j) = P_X(x_i) \cdot P_Y(y_j)$$

- Es sei $m(X)$ eine Funktion der Zufallsgröße X , deren Werte gemessen wurden. Dann bezeichnen wir mit $E(m(X))$ den Erwartungswert der Funktion $m(X)$. Dieser ist definiert als:

$$E(m(X)) = \sum_i m(x_i) \cdot P(x_i) \quad \text{für eine diskret verteilte Größe} \quad (10)$$

$$E(m(X)) = \int_{-\infty}^{\infty} m(x) \cdot f(x) \cdot dx \quad \text{für eine stetig verteilte Größe} \quad (11)$$

- Insbesondere erhält man

$$\text{mit } m(X) = X \quad \Rightarrow \quad E(X) = \mu \quad \text{Mittelwert } \mu \quad (12)$$

$$\text{mit } m(X) = (X - \mu)^2 \quad \Rightarrow \quad E((X - \mu)^2) = \sigma^2 \quad \text{Varianz } \text{Var}(X) \equiv \sigma^2 \quad (13)$$

- Aus den obigen Definitionen ergibt sich der Erwartungswert einer Summe von Zufallsvariablen $\{X_1, \dots, X_n\}$ als:

$$E(a_1 X_1 + \dots + a_n X_n) = a_1 E(X_1) + \dots + a_n E(X_n) \quad (14)$$

mit reellen Zahlen a_i . Insbesondere ergibt sich daraus der Verschiebungssatz

$$E((X - \mu)^2) = E(X^2) - \mu^2, \quad (15)$$

der häufig die praktische Berechnung der Varianz erleichtert.

- Sind die Zufallsgrößen $\{X_1, \dots, X_n\}$ statistisch unabhängig, dann gilt für die Varianz der Summe:

$$\text{Var}(a_1 X_1 + \dots + a_n X_n) = a_1^2 \text{Var}(X_1) + \dots + a_n^2 \text{Var}(X_n) \quad (16)$$

Wer sich davon überzeugen möchte setze die Summe $\sum a_i X_i$ mit (14) in die Definition (13) ein, multipliziere aus und stelle fest, dass unter der Bedingung (9) für die Mischterme $i \neq j$ gilt: $E((X_i - \mu_i) \cdot (X_j - \mu_j)) = E((X_i - \mu_i)) \cdot E((X_j - \mu_j)) = 0$.

Aus (16) folgt dann für arithmetische Mittelwerte unabhängiger Messungen $\bar{X} = \sum X_i/n$

$$\text{Var}(\bar{X}) = \text{Var}(X)/n^2 \quad (17)$$

und somit folgt Glg. (4) aus (3).

Nun können wir auch den Faktor $n - 1$ verstehen, der in Glg (3) eingeführt wurde: Es gilt nämlich nach (16): $\text{Var}(X_k - \bar{X}) = \text{Var}(X_1/n + \dots + (1/n - 1)X_k + \dots + X_n/n) = \frac{n-1}{n} \text{Var}(X)$ (nachrechnen!) und somit zeigt sich für Glg. (3), dass

$$E\left(\sum_i (X_i - \bar{X})^2\right) = \sum_i E\left((X_i - \bar{X})^2\right) = n \text{Var}(X_i - \bar{X}) = (n - 1) \text{Var}(X) \quad (18)$$

Somit ist also der Erwartungswert der Größe S^2 definiert gemäß Glg. (3) gleich der Varianz der Größe X genau dann, wenn im Nenner von (3) der Faktor $n - 1$ und nicht n steht! Man spricht dann auch davon, dass S^2 einen *erwartungstreuen Schätzer* für σ^2 darstellt.

2.2.2. Die Binomialverteilung $B_n(p)$

Die Messung bestehe darin, in einem System zwei sich gegenseitig ausschließende Merkmale A bzw. \bar{A} zu unterscheiden und entsprechend zu klassifizieren. Typisches Beispiel in der Physik sind Experimente, in denen Teilchen (Photonen, Elektronen, Protonen, etc.) gezählt werden; A bedeutet dann, dass das Teilchen in einem vorgegebenen Zeit-, Energie-, oder Impulsfenster registriert wurde, bzw. \bar{A} nicht registriert wurde. Die Wahrscheinlichkeit für das Auftreten der Ausprägung A sei p und entsprechend $1 - p$ für die komplementäre Ausprägung \bar{A} .

Annahme: Eine Datenanalyse oder eine andere a priori Information liefere uns einen soliden Schätzwert für p .

Frage: Wie groß ist die Wahrscheinlichkeit $P_n(k)$, dass bei n Realisationen der Messung k mal die Ausprägung A registriert wird? Antwort:

$$P_n(k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \quad (19)$$

Die daraus gewonnene Verteilung heißt Binomialverteilung $B_{np}(k)$. Sie ist naturgemäß diskret und wird durch die zwei Parameter n und p festgelegt.

$$\text{Ihr Mittelwert ist:} \quad \mu = np \quad (20)$$

$$\text{Ihre Varianz ist:} \quad \sigma^2 = np(1 - p) \quad (21)$$

$$P_n(k) \text{ wird maximal für} \quad np + p - 1 \leq k \leq np + p \quad (22)$$

- Mit Hilfe der Stirling'schen Formel kann gezeigt werden, dass sich für große n und nicht zu extreme p die Binomialverteilung $B_{np}(k)$ einer Normalverteilung $N(x; np, np(1 - p))$ (s. Abschnitt 2.2.5) nähert, d. h.:

$$\sum_{k < x} P_n(k) = \sum_{k < x} \binom{n}{k} p^k (1 - p)^{n-k} \longrightarrow \int_{-\infty}^x \frac{1}{\sqrt{2\pi} \sqrt{np(1-p)}} \exp\left\{-\frac{(\xi - np)^2}{2(np(1-p))}\right\} d\xi \quad (23)$$

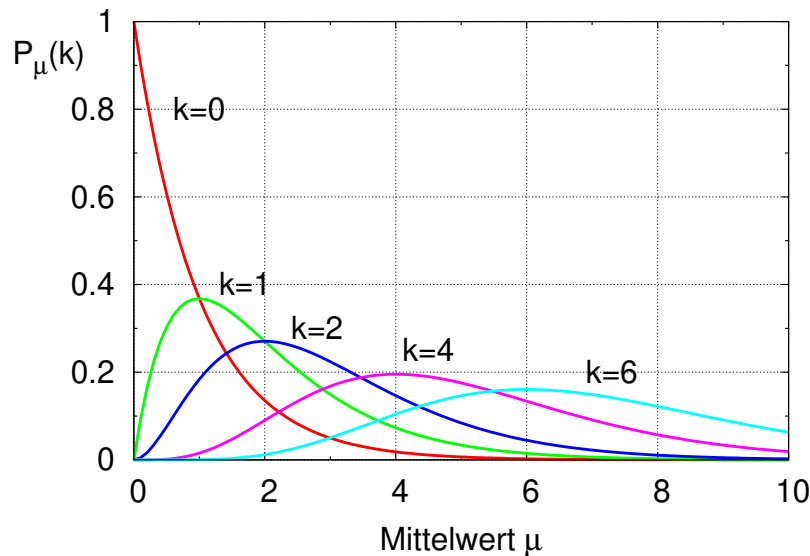


Abbildung 2: Darstellung der Poissonverteilung (29). Man sollte sich aber durch diese Darstellung nicht verwirren lassen: Die Poissonverteilung ist diskret in k . Aber es ist üblich, sie über ihren Parameter Mittelwert μ aufzutragen.

- Bei kleinen Wahrscheinlichkeiten p und großen n , kann mit der Poissonverteilung (s. folgender Abschnitt) mit Mittelwert np gerechnet werden, denn es gilt

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \longrightarrow \frac{(np)^k}{k!} \exp(-np) = P_{\mu=np}(k) \quad (24)$$

2.2.3. Die Poissonverteilung $P_\mu(k)$

Betrachtet wird, wie bei der Binomialverteilung, eine zufällige Variable mit zwei sich ausschließenden Ausprägungen A und \bar{A} , wie z. B. in einem Teilchenzählprozess (vgl. Absatz 2.2.2). Dabei soll die Anzahl der vorliegenden Messungen n sehr groß sein, $n \rightarrow \infty$. Weiter ist erforderlich, dass ein guter Schätzwert μ für die Anzahl Experimente mit Ausprägung A (z.B. A : Teilchen wurde registriert) vorliegt. Die Wahrscheinlichkeit p für die Messung von A ist also

$$p = \frac{\mu}{n} \quad (25)$$

Frage: Wie groß ist die Wahrscheinlichkeit bei einer Wiederholung der Messung k -mal das Merkmal A zu messen?

Konkret z. B. die Messung von radioaktiven Kernzerfällen: Wie groß ist die Wahrscheinlichkeit bei Vorliegen von n radioaktiven Atomkernen in einem gegebenen Zeitfenster genau k Zerfallsereignisse zu messen? Die entsprechende Wahrscheinlichkeit ist durch die Binomialverteilung (19) mit $p = \mu/n$

(25) gegeben:²

$$P_n(k) = \binom{n}{k} \cdot \left(\frac{\mu}{n}\right)^k \cdot \left(1 - \frac{\mu}{n}\right)^{n-k} \quad (26)$$

Nun ist n in einem kernphysikalischen Experiment typischer Weise eine sehr große Zahl: selbst bei wenigen Mikrogramm eines radioaktiven Elementes ist n von der Größenordnung 10^{15} . Im Vergleich dazu sind effektive Zählereignisse sehr selten. Das heißt es liegen folgende Verhältnisse vor: $k/n \ll 1$, $\mu/n \ll 1$ und $n \rightarrow \infty$. Damit können wir folgende Näherungen machen:

$$\text{Einsetzen in (26) von } \left(1 - \frac{\mu}{n}\right)^{n-k} \longrightarrow \exp(-\mu) \quad (27)$$

$$\text{und } \binom{n}{k} \frac{1}{n^k} \longrightarrow \frac{1}{k!} \quad (28)$$

$$\text{ergibt die Poissonverteilung : } P_\mu(k) = \frac{\mu^k}{k!} \exp(-\mu) \quad (29)$$

Die Poissonverteilung $P_\mu(k)$ ist eine diskrete Verteilung und wird durch einen einzigen Parameter, ihren Mittelwert μ festgelegt. Ihre Varianz beträgt

$$\text{Varianz der Poissonverteilung: } \sigma^2 = \mu \quad (30)$$

Mittelwert und Varianz der Poissonverteilung sind also gleich groß. Für große Mittelwerte lässt sich die Poissonverteilung durch eine Normalverteilung approximieren, die dann ebenfalls nur einen freien Parameter aufweist:

$$\sum_{k < x} P_\mu(k) = \sum_{k < x} \frac{\mu^k}{k!} \exp(-\mu) \longrightarrow \int_{-\infty}^x \frac{1}{\sqrt{\mu \cdot 2\pi}} \exp\left(-\frac{(\xi - \mu)^2}{2\mu}\right) d\xi \quad (31)$$

Beispiel:

Sie haben in einem kernphysikalischen Experiment in einem Zeitintervall von 6 Minuten 534 Zählereignisse registriert. Wie geben Sie die Zählrate \dot{N} und deren Unsicherheit an?

Ergebnis: Es handelt sich hierbei um eine Einzelmessung. Also ist $k = 534$ und wir nähern $\mu \approx k$, sowie $\sigma \approx \sqrt{k}$. Damit ergibt sich:

$$\dot{N} = \frac{\Delta k}{\Delta t} = (1.48 \pm 0.06) \text{ s}^{-1}$$

2.2.4. Das Histogramm

Wir nehmen an, dass ein Experiment durchgeführt wurde, bei dem die *kontinuierliche* Größe X gemessen wurde. Das Ergebnis liegt als Messdatensatz $\{x_i\} = \{x_1, \dots, x_n\}$ vor. Als erstes wird der

²Die Binomialverteilung sieht streng genommen nur Experimente „mit Zurücklegen“ vor, damit die Wahrscheinlichkeit p für ein einzelnes Ereignis sich im Verlauf des Versuches nicht ändert. Im vorliegenden Fall ist aber die Anzahl k der zerfallenen Kerne gegen n eine sehr kleine Zahl, so dass die Bedingung einer unveränderten Wahrscheinlichkeit p auch ohne „Zurücklegen“ erfüllt bleibt.

geübte Experimentator sich einen Überblick über die Daten verschaffen wollen. Dazu sucht er sich im Datensatz den Minimal- und den Maximalwert. Danach teilt er dieses Intervall in m Klassen ein, wobei er $m \ll n$ wählt, so dass in jede Klasse eine genügende Anzahl Messwerte fallen (genügend: mind. 5). Danach zeichnet er sich ein *Histogramm* des Datensatzes, indem er jede Klasse als vertikalen Balken mit einer Länge, die proportional der Anzahl enthaltener Messwerte ist, darstellt; In Abb. 3 ist eine „einhöckrige“ *unimodale*, Verteilung dargestellt. Wenn wir davon ausgehen, dass die gemessene physikalische Größe durch einen einzigen Wert charakterisiert wird, erwarten wir eine unimodale Messdatenverteilung. Zeigt aber das Histogramm zwei- oder mehr Höcker (Moden), so muss am Experiment etwas schief gegangen sein, und der ganze Datensatz muss als ungültig verdächtigt werden.

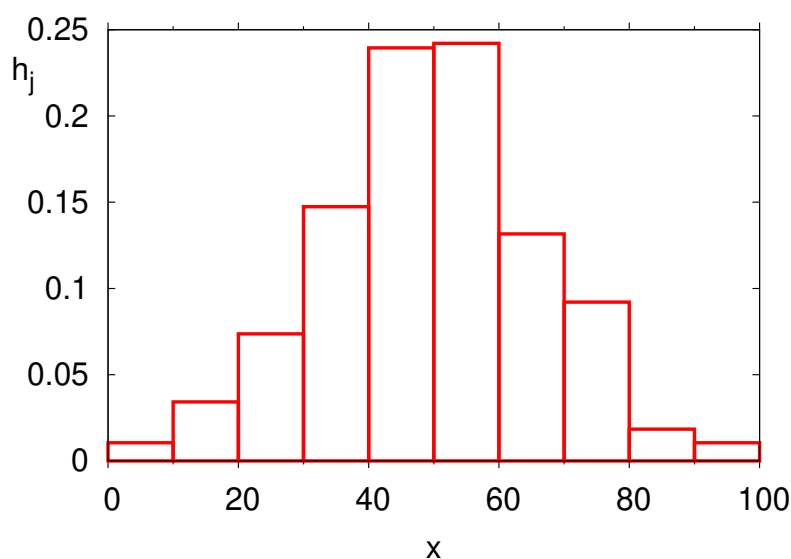


Abbildung 3: Histogramm eines Datensatzes $\{x_i, i = 1 \dots 380\}$. Die Messwerte x_i wurden in $m = 10$ Klassen eingeteilt. h_j ($j = 1 \dots 10$) bezeichnet die relative Häufigkeit, mit der ein Messwert in die Klasse j fällt.

Je größer der Messdatensatz ist, desto feiner können wir die Klassen unterteilen. Allerdings strebt dann das individuelle h_j der einzelnen Klasse gegen Null. Daher dividert man durch die Intervallbreite einer jeweiligen Klasse und erhält so eine Wahrscheinlichkeitsdichte $f_j = h_j / \Delta x_j$. Für einen sehr großen Messdatensatz und immer feinere Klassenunterteilung nähern sich so die Treppenstufen einer kontinuierlichen Kurve. Ohne genauere Kenntnisse der experimentellen Bedingungen können wir jedoch über die Form dieser Grenzkurve keine zuverlässigen Aussagen machen, d. h. wir verfügen im Allgemeinen nicht über ein mathematisches Modell der Verteilung von Messdaten. In vielen Fällen zeigt sich jedoch, dass eine *Normalverteilung* die Dichtefunktion f_j der Messdaten gut beschreibt. Der Grund dafür liegt im zentralen Grenzwertsatz dem wir uns im nächsten Abschnitt zuwenden. Zunächst jedoch noch einige Bemerkungen zur Normalverteilung.

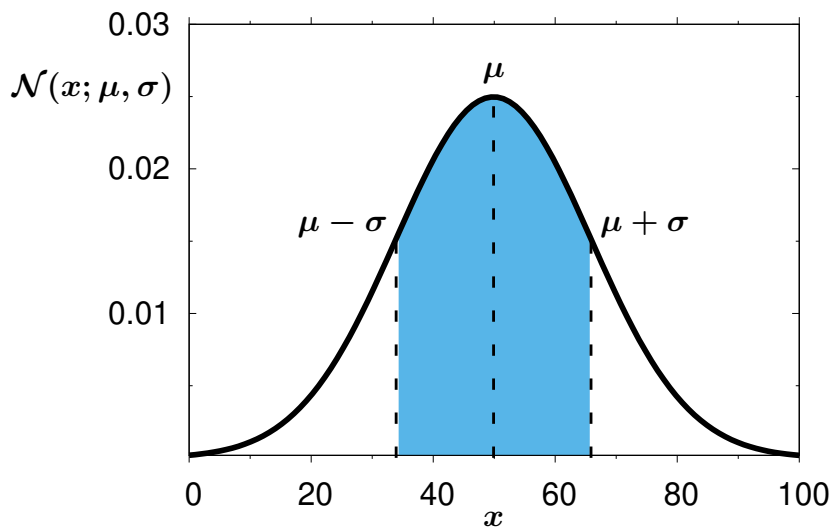


Abbildung 4: Normalverteilung eines Datensatzes. Im Intervall $(\mu - \sigma, \mu + \sigma)$ liegen 68.3% der Messdaten.

2.2.5. Die Normalverteilung $\mathcal{N}(x; \mu, \sigma)$

Eine Zufallsgröße heißt normal- oder gaußverteilt wenn die zugehörige Dichtefunktion folgende Form hat:

$$\text{Normalverteilung: } \mathcal{N}(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} \quad (32)$$

$\mathcal{N}(x; \mu, \sigma)$ oder auch kurz $\mathcal{N}(\mu, \sigma)$ beschreibt eine symmetrische Glockenkurve, die durch die beiden Parameter μ und σ eindeutig festgelegt ist. Die Kurve hat ihr Maximum an der Stelle $x = \mu$ und die Breite 2σ an den Wendepunkten, siehe Abb. 4. Der Parameter σ ist dabei gleichzeitig die Standardabweichung der Verteilung gemäß Glg. (13) und μ ist der Mittelwert der Normalverteilung gemäß Glg. (12). Wenn eine Messgröße gemäß $\mathcal{N}(\mu, \sigma)$ normalverteilt ist, so kann der Verteilungsparameter σ , gemäß unseren Überlegungen zu Glg. (18), aus dem Messdatensatz $\{x_i\}$ nach Formel (3) abgeschätzt werden, also $\sigma = s$. Beschreibt hingegen die Normalverteilung eine Verteilungsfunktion von Mittelwerten, so sind die Verteilungsparameter nach Glg. (2) und (4) zu ermitteln, also $\sigma = s/\sqrt{n}$.

Bemerkungen

1. Die Normalverteilung ist eine Wahrscheinlichkeitsdichteverteilung, das bedeutet sie ist normiert $\int_{-\infty}^{\infty} \mathcal{N}(x; \mu, \sigma) dx = 1$ und $\mathcal{N}(x; \mu, \sigma) dx$ beschreibt die Wahrscheinlichkeit dafür, einen Wert der Größe X innerhalb des Intervalls $(x, x + dx)$ zu finden, also

$$P(x < X < x + dx) = \mathcal{N}(x; \mu, \sigma) dx. \quad (33)$$

Dem entsprechend ergibt sich für die Wahrscheinlichkeit P einen Wert der Größe X im Intervall (a, b) anzutreffen,

$$P(a < X < b) = \int_a^b \mathcal{N}(x; \mu, \sigma) dx. \quad (34)$$

Also ist z. B. die Wahrscheinlichkeit, einen Datenpunkt im Intervall $(\mu - \sigma, \mu + \sigma)$ zu erhalten:

$$P(|X - \mu| < \sigma) = \int_{\mu - \sigma}^{\mu + \sigma} \mathcal{N}(x; \mu, \sigma) dx = 0.683 \quad (35)$$

Für die Praxis wichtig sind auch die 2- σ -Werte $P(|X - \mu| < 2\sigma) = 0.9545$, bzw. die 3- σ -Werte $P(|X - \mu| < 3\sigma) = 0.9973$; siehe Diskussion in Abschnitt 2.3.

2. Durch die Koordinatentransformation

$$z = \frac{x - \mu}{\sigma} \quad (36)$$

gewinnt man die Darstellung der *standardisierten Normalverteilung* $\mathcal{N}(0, 1)$ mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$

$$\mathcal{N}(z; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \quad (37)$$

3. Die kumulative Verteilungsfunktion der Normalverteilung lautet:

$$\Phi(x; \mu, \sigma) = \int_{-\infty}^x \mathcal{N}(\xi; \mu, \sigma) d\xi \quad (38)$$

Da sich dieses Integral nicht elementar lösen lässt, ist die Verteilungsfunktion der standardisierten Normalverteilung $\Phi(x; 0, 1) = \int_{-\infty}^x \mathcal{N}(\xi; 0, 1) d\xi$ tabelliert.

4. Falls wir annehmen können, dass die Werte einer Messreihe normalverteilt sind, so sind die nach (2) bzw. (3) bestimmten Werte umso bessere Schätzwerte für den Mittelwert μ und die Standardabweichung σ der Verteilung, je größer die Anzahl Messdaten n ist. Da es aufwändig ist, viele Daten zu messen, wird man oft mit kleinen n auskommen wollen. Wie bei Glg. (4) diskutiert, geht aber die Anzahl der Messwerte $\propto 1/\sqrt{n}$ in die Unsicherheit der geschätzten Lage des Mittelwertes ein. In der Praxis gilt daher: die Anzahl Messwerte in einer Messreihe sollte vernünftigerweise 10 nicht unterschreiten. Verwendet man kleinere Datensätze zur Abschätzung von Mittelwert und Standardabweichung, dann ist die Unsicherheit des Mittelwertes mit Hilfe des sog. t-Faktors zu korrigieren. Mehr dazu im Abschnitt 2.3.1

2.2.6. Der Zentrale Grenzwertsatz (Central Limit Theorem)

Worin begründet sich nun die hohe Universalität der Normalverteilung und die Tatsache, dass diese Verteilung so einen großen Stellenwert in der mathematischen Statistik wie auch in der praktischen experimentellen Arbeit einnimmt? Die Antwort darauf liefert der sogenannte Zentrale Grenzwertsatz. Bei unzähligen Problemen des experimentellen Alltags lässt sich die Verteilung einer interessierenden Zufallsgröße nur mit enormem Aufwand oder auch gar nicht exakt bestimmen. Der Zentrale Grenzwertsatz liefert nun in vielen Fällen eine asymptotische Aussage über die Verteilung, die in der Praxis meist völlig ausreichend ist und dem Experimentator häufig weitreichende Möglichkeiten an die Hand gibt, seine Resultate zu beurteilen.

Es gibt den Zentralen Grenzwertsatz in verschiedenen Varianten. Eine einfache Form lautet wie folgt, für einen Beweis wende man sich an die Standardlehrbücher [7, 8]:

Seien $\{X_1, X_2, \dots, X_n\}$ **unabhängige** und **identisch verteilte** Zufallsvariablen³ mit Mittelwert μ und Standardabweichung σ . Dann konvergiert die Verteilung der Mittelwerte $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$ für $n \rightarrow \infty$ gegen die Normalverteilung $\mathcal{N}(\bar{X}_n; \mu, \sigma/\sqrt{n})$.

Das bedeutet, dass wir *ohne* eine Annahme über die Form der tatsächlichen Verteilung der Einzelmessungen $\{x_i\}$ zu machen, wissen, dass die daraus gewonnenen Mittelwerte normalverteilt sind, sofern sie nur aus einer ausreichenden Anzahl n von Einzelmessungen gewonnen wurden. Wichtig dabei ist lediglich, dass die beiden Voraussetzungen „*unabhängig*“ und „*identisch verteilt*“ erfüllt sind, was bei einer Mittelwertbildung aus unter gleichen Bedingungen durchgeführten Einzelmessungen für gewöhnlich der Fall ist. Was „ausreichende Anzahl“ genau bedeutet hängt im Allgemeinen von der konkreten Verteilung der $\{x_i\}$ ab. In der Regel ist jedoch eine Normalverteilung der Mittelwerte bereits ab $n \geq 10$ akzeptabel erfüllt.

Wir betrachten dazu als Beispiel eine Zufallsgröße X , die der Einfachheit halber alle Werte im Intervall $[0, 1]$ mit gleicher Wahrscheinlichkeit annehmen soll, d. h. es liegt eine gleichverteilte Zufallsgröße vor und die Wahrscheinlichkeitsdichte der Grundgesamtheit ist:

$$f_{\text{gleich}}(x) = \begin{cases} 1 & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases} \quad (39)$$

mit einem Mittelwert $\mu_g = 0.5$ und einer Standardabweichung $\sigma_g = 1/(2\sqrt{3})$. Nun bestimmen wir in einem Computereperiment die Verteilung von Mittelwerten $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$, die jeweils aus n gemäß (39) verteilten Zufallsgrößen gebildet wurden. Abb. 5 zeigt nun, neben der Dichtefunktion der Grundgesamtheit, die Dichte, die aus der Häufigkeitsverteilung von Mittelwerten \bar{X}_n gewonnen wurde, jeweils als Histogramm. Dabei gibt \bar{X}_1 in brauchbarer Näherung die Wahrscheinlichkeitsdichte der Gleichverteilung wider und der Unterschied zu einer Gaußverteilung mit gleicher Schwankungsbreite ist sehr auffällig. Je mehr unabhängige und gleichverteilte X_i jedoch in die Mittelwertbildung eingehen, desto mehr nähert sich die Verteilung dieser Mittelwerte einer Normalverteilung $\mathcal{N}(\mu_g, \sigma_g/\sqrt{n})$. Bereits für $n = 10$ sind dabei die Abweichungen praktisch zu vernachlässigen.

Nun gibt es Verallgemeinerungen des Zentralen Grenzwertsatzes, die sogar noch weiterreichende Schlussfolgerungen zulassen. Typisch dabei ist, dass eine der beiden Voraussetzungen „*unabhängig*“ und „*identisch verteilt*“ zugunsten einer schwächeren Forderung an die Zufallsvariablen X_i fallengelassen wird. Dann lässt sich z. B. zeigen, dass die Summe von unabhängigen und beliebig (auch unterschiedlich) verteilten Zufallsgrößen mit endlicher Varianz im Grenzfall $n \rightarrow \infty$ normalverteilt ist, solange eine weitere (schwache) Bedingung für die X_i gilt, die im Wesentlichen sicherstellt, dass nicht ein Teil der Summanden alle anderen dominiert (z. B. die Ljapunow Bedingung, s. [7]). Physikalisch gesprochen heißt das: Setzt sich eine Messgröße aus einer großen Anzahl von unabhängigen unterschiedlich aber „vernünftig“ verteilten Einzeleffekten additiv zusammen, so ist diese normalverteilt. Da dies in der Natur sehr häufig zutrifft, erklärt sich so die ubiquitäre Erscheinung der Normalverteilung als (Näherungs-)Verteilung physikalischer Größen.

³Hier wird also nichts über die konkrete Form der Verteilung vorausgesetzt, diese ist beliebig!

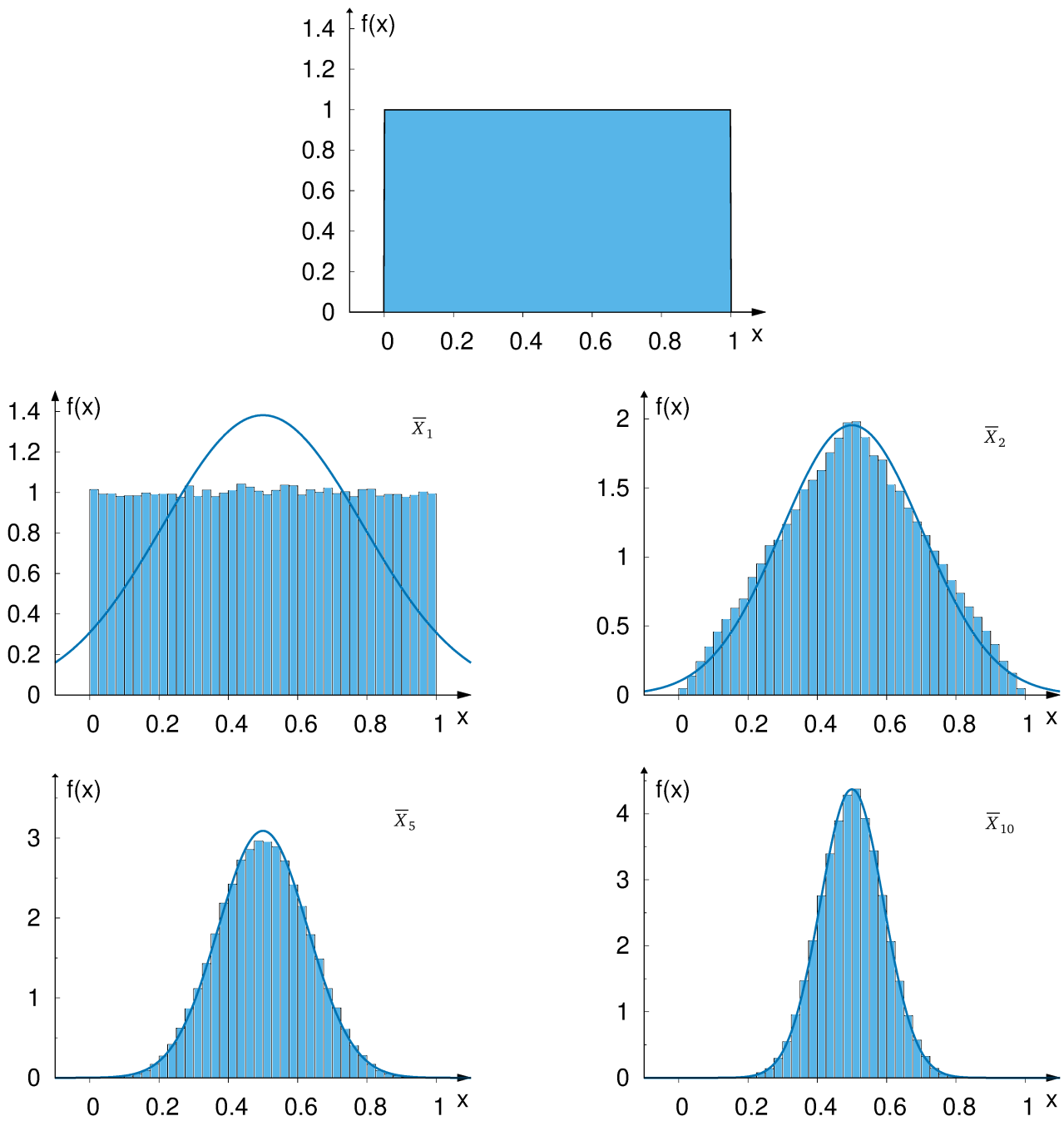


Abbildung 5: Beispiel für den Zentralen Grenzwertsatz. Für eine gleichverteilte Zufallsgröße X , $x \in [0, 1]$ wird die Verteilung verschiedener Mittelwerte $\bar{X}_n = (X_1 + X_2 + \dots + X_n)/n$ durch ein „Computereperiment“ bestimmt. Die Stichprobengröße ist jeweils 10^5 . Ganz oben: die Verteilung der Grundgesamtheit mit einem Mittelwert von $\mu_g = 0.5$ und einer Standardabweichung von $\sigma_g = 1/(2\sqrt{3})$. Die „experimentell“ ermittelten Stichprobenverteilungen sind als Histogramme dargestellt. Zum Vergleich ist jeweils eine Normalverteilung mit gleichem Mittelwert und gleicher Standardabweichung σ_g/\sqrt{n} gezeigt. Je mehr unabhängige Zufallsgrößen in die Mittelwertbildung eingehen, desto schmaler wird die Verteilung der Mittelwerte und desto mehr nähert sich diese einer Normalverteilung.

2.3. Überdeckungsintervall und Erweiterungsfaktor

Nehmen wir an während eines Experiments, in dem die physikalische Größe X ermittelt werden soll, wird ein Datensatz $\{x_i\}$ bestehend aus n Einzelmessungen gewonnen. Dieser Datensatz wird mit den in Absatz 2.1 dargestellten Methoden analysiert, und ergibt als Resultat $\bar{x} \pm s_{\bar{x}}$, vgl. (5).

Der Zentrale Grenzwertsatz sagt uns nun, dass bei ausreichend großem n der gemessene Mittelwert \bar{x} aus einer normalverteilten Gesamtheit $\mathcal{N}(\bar{X}; \mu, \sigma)$ von Mittelwerten stammt deren Breite durch $s_{\bar{x}} = s/\sqrt{n}$ abgeschätzt werden kann. Dann gibt nach Glg. (35)

$$P(\bar{x} - s/\sqrt{n} \leq \mu \leq \bar{x} + s/\sqrt{n}) = 0.683$$

die relative Häufigkeit an, mit der durch ein solches Experiment festgelegte Grenzen $\bar{X} \pm S/\sqrt{n}$ den unbekanntem Mittelwert μ der Grundgesamtheit enthalten. Man spricht deshalb von 68% *Überdeckungswahrscheinlichkeit*, die dem Intervall $\bar{x} \pm s/\sqrt{n}$ zugeordnet ist und daher auch von einem 68%-*Überdeckungsintervall*.

Dies ist *nicht* gleichbedeutend damit, dass für eine konkrete Realisierung des Intervalls $\bar{x} \pm s/\sqrt{n}$ in einem bestimmten Experiment die Wahrscheinlichkeit angegeben wird, mit der der “wahre Wert” in diesem Intervall zu finden ist. Auch wenn es intuitiv naheliegend ist zu fragen: “Mit welcher Wahrscheinlichkeit befindet sich der wahre Wert der Messgröße in dem durch mein Experiment bestimmten Intervall?” so kann die konventionelle Statistik darauf jedoch keine Antwort geben, denn μ ist *keine Zufallsgröße* sondern der feststehende, aber unbekannte Erwartungswert der Grundgesamtheit. Gelöst wird dieses Problem erst im Rahmen der Bayes’schen Statistik, im Rahmen derer auf Grundlage der verfügbaren Information auch dem wahren Wert μ eine Verteilung zugeordnet werden kann (siehe Kap. 4.3).

Um nochmals die Bedeutung des Überdeckungsintervalls zu veranschaulichen: Nehmen wir an, wir würden das Experiment noch einige Male wiederholen. Jede Wiederholung liefert einen neuen Datensatz $\{x_i\}_1, \{x_i\}_2, \dots, \{x_i\}_n$. In Abb. 6 sind $n = 8$ Datensätze dargestellt mit ihren jeweiligen Mittelwerten \bar{x}_i und den 68%-Überdeckungsintervallen als Unsicherheitsbalken. Der im Normalfall unbekannte, wahre Wert μ der gemessenen Größe X wird dann in 68%, also ca. 2/3, der dargestellten Balken zu finden sein.

Wenn uns 68% zu unsicher sind, sagt uns die Normalverteilung $\mathcal{N}(\mu, s/\sqrt{n})$, wie groß wir in Einheiten von s/\sqrt{n} das Intervall wählen müssen, um eine bestimmte Überdeckungswahrscheinlichkeit zu gewährleisten. Wollen wir diese z. B. mit 95% festlegen, so wird das entsprechende Intervall eine Breite von $k \cdot s/\sqrt{n}$ mit $k = 1.96$ aufweisen. Der Faktor k wird dabei auch *Erweiterungsfaktor* genannt und bei einer Unsicherheitsangabe von $\pm k \cdot s/\sqrt{n}$ von einer *erweiterten Unsicherheit* gesprochen.

Hinweis: Es ist in der Physik üblich, in grafischen Darstellungen von Experimentdaten die Messpunkte mit sog. *Unsicherheitsbalken* (früher: Fehlerbalken) zu versehen. Diese Unsicherheitsbalken haben standardmäßig die Länge von $\pm s_{\bar{x}} = \pm s/\sqrt{n}$, vgl. (4). Sie repräsentieren also im Falle einer Normalverteilung eine Überdeckungswahrscheinlichkeit von 68%. Im Falle der Angabe von erweiterten Unsicherheiten mit $k > 1$ ist daher stets der Erweiterungsfaktor explizit anzugeben.

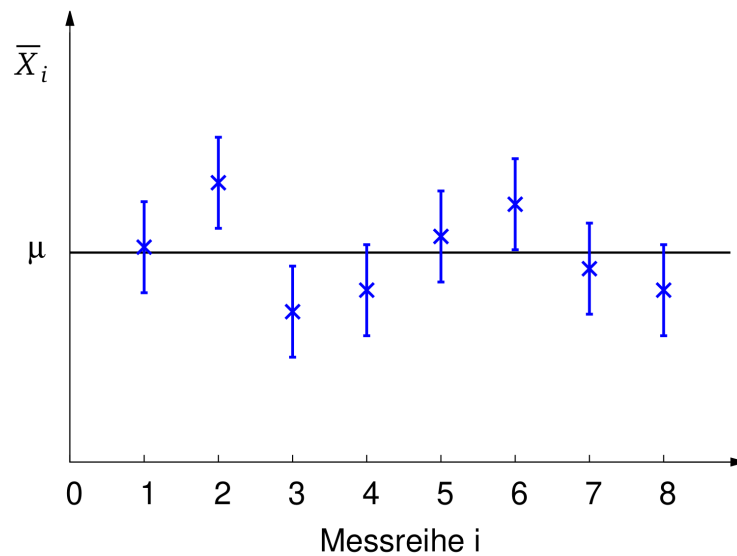


Abbildung 6: Darstellung der Resultate eines Experiments, in dem die Größe X in 8 Messreihen ermittelt wurde. Jede Messserie i liefert einen Datensatz, der hier durch seinen Mittelwert \bar{x}_i und das 68%-Überdeckungsintervall als Unsicherheitsbalken repräsentiert wird. Die Grafik sollte so interpretiert werden: Im Mittel werden 68% der Unsicherheitsintervalle den (unbekannten) wahren Wert μ der untersuchten Größe X enthalten.

2.3.1. Die Student'sche t -Verteilung

Liegt den Schätzwerten für Mittelwert und Standardabweichung einer normalverteilten Zufallsgröße X nur ein geringer Stichprobenumfang zugrunde, was im Grundpraktikum durchaus vorkommen kann, so muss zur Angabe der Unsicherheit des Mittelwertes bzw. zur Angabe eines Überdeckungsintervalles zu einer gegebenen Wahrscheinlichkeit, ein Korrekturfaktor berücksichtigt werden.

Der Grund ist folgender: Wird von einem Datensatz aus n unabhängigen Messwerten $\{x_i\}$ einer gemäß $\mathcal{N}(\mu, \sigma)$ normalverteilten Zufallsgröße X , der empirische Mittelwert \bar{X} gebildet, so ist die Zufallsgröße $Y = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ nach Glg. (37) standardnormalverteilt gemäß $\mathcal{N}(Y; 0, 1)$. Schätzt man nun jedoch die Standardabweichung gemäß S/\sqrt{n} ab, dann ist die Größe $T = (\bar{X} - \mu)/(S/\sqrt{n})$ nicht mehr normalverteilt, da S neben \bar{X} eine weitere Zufallsvariable darstellt, zu der ebenfalls eine Verteilung gehört. Die Größe T folgt in diesem Fall einer *Student'schen t -Verteilung* mit $n - 1$ Freiheitsgraden (Details siehe z. B. [7]). Diese Verteilung geht für große n in eine $\mathcal{N}(0, 1)$ Normalverteilung über. Die Varianz von T hängt von der Stichprobengröße n ab und es gilt für $n \geq 4$:

$$\text{Var}(T) = \frac{n-1}{n-3}.$$

Nach den Rechenregeln zur Varianz Glg. (16) ergibt sich daher für ein Messergebnis inklusive Standardabweichung zur Angabe der Unsicherheit:

$$\bar{x} \pm \sqrt{\frac{n-1}{n-3}} \cdot s/\sqrt{n}. \quad (40)$$

n	t_p für $p = 0.683$	t_p für $p = 0.95$
3	1.32	4.3
4	1.20	3.18
5	1.14	2.78
6	1.11	2.57
8	1.08	2.37
10	1.06	2.26
1000	1.00	1.96

Tabelle 1: Der Student'sche t -Faktor für einige ausgewählte Stichprobengrößen n und zwei verschiedene Überdeckungswahrscheinlichkeiten p .

Insbesondere wird hier deutlich, dass für $n \leq 3$ die Angabe einer Standardabweichung *nicht sinnvoll möglich* ist.

Analog ist vorzugehen, möchte man bei kleinen Stichprobengrößen ein Überdeckungsintervall angeben. Meist sind die p -Quantile der t -Verteilung tabelliert, d. h. die Werte t_p mit $P(|T - \mu| \leq t_p) = p$. Das Überdeckungsintervall für \bar{X} ergibt sich dann mit Hilfe dieses sog. t -Faktors und der empirischen Standardabweichung der Messdaten s zu:

$$\bar{x} \pm t_p \cdot s / \sqrt{n} \quad (41)$$

Für $p = 0.683$ und $p = 0.95$ kann der t -Faktor für einige Werte von n der Tabelle 1 entnommen werden. Für große n , so wird an dieser Tabelle auch deutlich, sind die aus der Student'schen t -Verteilung und aus der entsprechenden Normalverteilung ermittelten Überdeckungsintervalle identisch. In der Praxis berücksichtigt man typischer Weise die in Gln. (40) und (41) angegebenen Vorfaktoren für $n \leq 10$.

3. Angabe von Messunsicherheiten nach GUM

3.1. Typ A Messunsicherheiten

Zum Zwecke der Angabe von Messunsicherheiten mit statistischen Mitteln (Typ A Unsicherheiten nach GUM) können wir das bisher Gesagte wie folgt zusammenfassen: Die physikalische Größe X wird durch einen Datensatz bestehend aus n *unabhängigen* Messungen $\{x_i\} = \{x_1, x_2, \dots, x_n\}$ repräsentiert. Als bester Schätzwert für die Messgröße wird der empirische Mittelwert nach Glg. (2) angegeben:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

und die Streuung des Datensatzes um den Mittelwert durch die empirischen Standardabweichung nach Glg. (3) charakterisiert:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}.$$

Ist n dabei ausreichend groß, ist \bar{x} nach dem zentralen Grenzwertsatz der beste Schätzer für den Erwartungswert einer Normalverteilung von Mittelwerten \bar{X} deren Standardabweichung durch s/\sqrt{n} abgeschätzt werden kann. Also ist das Ergebnis als

$$\bar{x} \pm s/\sqrt{n}$$

anzugeben. Ist die Stichprobengröße klein und die Größe X normalverteilt, dann wird die Verteilung der Mittelwerte besser durch eine t -Verteilung abgeschätzt und das Ergebnis für $4 \leq n \leq 10$ nach Glg. (40) angegeben:

$$\bar{x} \pm \sqrt{\frac{n-1}{n-3}} \cdot s/\sqrt{n}$$

In manchen Fällen, wie z. B. in einem kernphysikalischen Zählexperiment, dem eine Poissonverteilung zugrunde liegt, ist die Abschätzung von Mittelwert und Standardabweichung aufgrund einer einzelnen Messung möglich, deren Ergebnis angegeben wird als:

$$x \pm \sqrt{x}.$$

In jedem Falle wird der Messgröße X also eine Wahrscheinlichkeitsverteilung zugeordnet. Die Standardabweichung dieser Verteilung wird dem Messergebnis laut GUM als sog. *Standardunsicherheit* zugewiesen:

$$X \pm \sqrt{\text{Var}(X)} \quad (42)$$

3.2. Typ B Messunsicherheiten

Wie bereits in Abschnitt 1 erwähnt, ist die Durchführung häufiger Wiederholmessungen nicht immer möglich und praktikabel. Im Gegenteil müssen im experimentellen Alltag sehr häufig Unsicherheitsabschätzungen auf andere Weise erfolgen. Diese können durch die konventionelle frequentistische

Statistik jedoch nicht erfasst werden, da letztere nur Aussagen über relative Häufigkeiten treffen kann. Eine einheitliche Beschreibung beider Arten von Unsicherheiten ist jedoch im Rahmen der Bayes'schen Statistik möglich. Hier werden Wahrscheinlichkeiten nicht mehr als relative Häufigkeiten verstanden, sondern charakterisieren unsere *unvollständige Kenntnis* über das Eintreten von Ereignissen auf der Basis der vorliegenden Information. Diese zugrunde liegende Information kann durchaus eine durch einen Datensatz gegebene Verteilung relativer Häufigkeiten sein, aber eben auch in anderer Art vorliegen, wie z. B. als Herstellerangabe über Toleranzgrenzen eines Messgerätes, als Literaturangabe aus vergangenen Messungen oder auch als eine (zwangsläufig subjektive) Einschätzung eines Experten gewonnen aus der langjährigen Erfahrung mit einem Experiment oder einem Material.

Damit eine einheitliche Behandlung der Messunsicherheiten erfolgen kann, wird auch den gemäß Typ B ermittelten Größen eine Wahrscheinlichkeitsverteilung $f(x)$ zugeordnet. Im Gegensatz zur frequentistischen Statistik trifft diese dann keine Aussage über eine (abgeschätzte) Verteilung relativer Häufigkeiten der Werte, die die Größe X annehmen kann, sondern gibt ein Maß dafür an, wie sehr wir, aufgrund der uns vorliegenden unvollständigen Information, vermuten, dass die Größe X einen bestimmten Wert hat. Dabei mag der Größe X physikalisch gesehen ein Zufallsprozess zugrunde liegen oder auch nicht.

Beispielsweise könnten wir auf der Grundlage vorliegender Informationen eine Wahrscheinlichkeit dafür angeben, dass der Wert einer bestimmten physikalischen Fundamentalkonstante zwischen zwei vorgegebenen Grenzen liegt. So eine naheliegende und intuitiv sinnvolle Aussage ist im Rahmen der konventionellen Statistik jedoch undenkbar, da es sich bei einer Fundamentalkonstante nicht um eine Zufallsvariable handelt. Zur Beschreibung der *uns vorliegenden Information* über diese Größe, ist die Angabe einer Wahrscheinlichkeitsverteilung im Rahmen der Bayes'schen Statistik jedoch sinnvoll. Diese Verteilung wird um so breiter sein, je ungenauer die uns vorliegende Information darüber ist. Im dem praktisch nie eintretenden Fall, dass wir über vollständige Information verfügen, wird die Verteilung hingegen die Form einer Dirac'schen Deltafunktion $\delta(x - x_0)$ annehmen, d. h. unendlich schmal werden.

3.3. Das Prinzip der maximalen Informationsentropie (PME)

Wie lässt sich nun bei gegebener Information eine Wahrscheinlichkeitsverteilung eindeutig zuordnen? Diese Zuordnung erfolgt unter der Maßgabe, dass die resultierende Verteilung *ausschließlich* die vorliegende Information widerspiegeln soll. Solch eine Verteilung wird dann *vorurteilsfreie Verteilung* genannt. Das bedeutet z. B., dass wenn gegebene Information es nicht zulässt, zwischen verschiedenen sich ausschließenden möglichen Ereignissen in einem Problem zu unterscheiden, dann müssen diese Ereignisse als gleich wahrscheinlich gewertet werden. Eine Verallgemeinerung dieses Prinzips ist das Prinzip der *maximalen Informationsentropie*. Einer Größe X wird eine Verteilung $f(x)$ dann folgendermaßen zugeordnet:

- Die Verteilung soll mit der gegebenen Information übereinstimmen, wird also z. B. so gewählt, dass ihr Erwartungswert mit dem besten Schätzer für den Messwert übereinstimmt und auch dass die Verteilung nicht Werten von X eine endliche Wahrscheinlichkeit zuweist, die von vornherein ausgeschlossen sind.

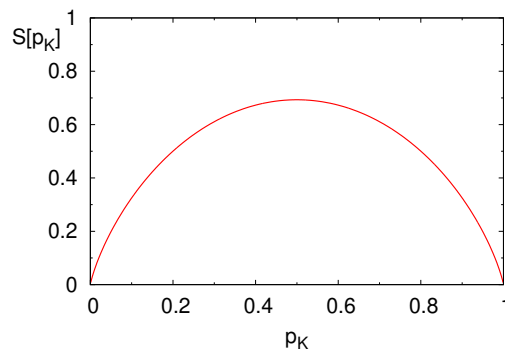


Abbildung 7: Die Informationsentropie $S[p_K]$ für einen einzelnen Münzwurf mit der Randbedingung $p_K + p_Z = 1$ ist für $p_K = 1/2$ maximal; also ist $p_K = p_Z = 1/2$ die vorurteilsfreie Verteilung für einen Münzwurf.

- Die Verteilung wird so gewählt, dass sie gleichzeitig die sog. Informationsentropie

$$S[f] = - \int f(x) \ln(f(x)) dx \quad (43)$$

maximiert. Wie sich zeigen lässt ist dann sicher gestellt, dass die Verteilung vorurteilsfrei ist, d. h. *nur* die gegebene Information widerspiegelt. Zur Begründung von (43) und für weitere Details sei der Leser auf [3] und [9] verwiesen.

Beispiel: Um die Vorgehensweise in einem ganz einfachen Fall zu demonstrieren, betrachten wir einen Münzwurf: die möglichen Ergebnisse des Münzwurfes sind „Kopf“ (K) oder „Zahl“ (Z). Intuitiv erwarten wir die Wahrscheinlichkeit $p = 1/2$ für jedes Ereignis. Wie würde man diese Wahrscheinlichkeit nach dem PME ermitteln? Die vorliegende Information besagt, dass mit Sicherheit eines der beiden Ereignisse eintritt, also $p_K + p_Z = 1$. Weitere Information liegt uns nicht vor. Die Informationsentropie ist dann in diesem Fall diskreter Ereignisse: $S[p_K, p_Z] = -p_K \ln(p_K) - p_Z \ln(p_Z)$. Die Randbedingung berücksichtigen wir durch Einsetzen von $p_Z = 1 - p_K$ und erhalten: $S[p_K] = -p_K \ln(p_K) - (1 - p_K) \ln(1 - p_K)$. Wie Abb. 7 zeigt und wie sich auch durch Nullsetzen der Ableitung analytisch ergibt, hat $S[p_K]$ ein Maximum bei $p_K = 1/2$, wie erwartet. Also ist $p_K = p_Z = 1/2$ die vorurteilsfreie Verteilung für einen Münzwurf.

Im allgemeinen Fall ist die Mathematik natürlich etwas komplizierter. Die kontinuierliche Funktion $f(x)$ zu finden, die unter gegebenen Randbedingungen den Ausdruck Glg. (43) maximiert, stellt ein typisches Variationsproblem der Funktionalanalysis dar, das sich mit Hilfe sogenannter Lagrangemultiplikatoren lösen lässt. Ohne dass wir an diese Stelle auf weitere Details eingehen können (siehe z. B. [3, 9]), soll ein weiteres Beispiel die Relevanz für die Anwendung in der Datenanalyse verdeutlichen:

Beispiel: Sie messen im Grundpraktikum mit einem Drehspulmessgerät eine Spannung von 100 V. Die vom Hersteller angegebene Toleranzklasse des Gerätes ist einem Aufdruck auf der Skala zu entnehmen und beträgt 2.5% des Messbereiches. Bei einem Messbereich von 600 V sind dies also 15 V. Wie groß ist der Unsicherheitsbeitrag aufgrund der vom Hersteller angegebenen Toleranz?

Nach vorliegender Information liegt die gesuchte Spannung also im Intervall zwischen 85 . . . 115 V, da die Toleranzklasse maximale Abweichungen charakterisiert. Weitere Informationen liegen uns nicht vor. Welche Verteilung soll nun unserer Messgröße sinnvoller Weise zugewiesen werden? Eine Normalverteilung scheidet beispielsweise aus, da sie nicht mit der gegebenen Information übereinstimmt: Sie weist nämlich auch beliebig großen Abweichungen eine evtl. kleine aber endliche Wahrscheinlichkeit zu, was der vorliegenden Information widerspricht.

Wird das Variationsproblem Glg. (43) unter den Randbedingungen $\int f(x)dx = 1$ (Verteilung ist normiert) und mit der Information $a \leq x \leq b$ (Wert liegt zwischen zwei Grenzen), d. h. $f(x < a) = 0$ und $f(x > b) = 0$ gelöst, dann ergibt sich als vorurteilsfreie Verteilung eine Gleichverteilung:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \quad (44)$$

Die Varianz der Gleichverteilung beträgt $\sigma^2 = (a - b)^2/12$. Übertragen auf die Charakterisierung des Drehspulmessgerätes bedeutet dies: Die vorliegende Information gibt eine obere Schranke für die Messabweichungen vor gemäß $U_{\min} \leq U \leq U_{\max}$. Wenn keine weitere Information vorliegt, ergibt sich aus dem PME Glg. (43) als vorurteilsfreie Verteilung, die den Stand der unvollständigen Kenntnis über die Messgröße wiedergibt, eine Gleichverteilung auf dem Intervall $U_{\min} \leq U \leq U_{\max}$. Die zugehörige Standardabweichung ergibt dann den gesuchten Unsicherheitsbeitrag aufgrund der Toleranz zu $u_{\text{Tol}}(U) = (U_{\max} - U_{\min})/(2\sqrt{3}) = 9 \text{ V}$ und das Ergebnis ist unter Vernachlässigung eventueller weiterer Beiträge $U = (100 \pm 9) \text{ V}$.

Auf diese Weise ist es möglich in verschiedenen Fällen einer Messgröße aufgrund vorliegender nicht-statistischer Information eine Verteilung zuzuweisen. Tabelle 2 zeigt dabei welche Verteilungen für einige typische Fälle von vorliegender Information durch das PME zugewiesen werden. Erneut soll betont werden, dass die zugewiesene Verteilung dann *keine Verteilung relativer Häufigkeiten* mehr darstellt, sondern die uns vorliegende unvollständige Information über die Messgröße abbildet.

Da nun sowohl die statistische (Typ A) als auch die nicht-statistische Information (Typ B) über eine Messgröße durch die jeweilig zugeordnete Wahrscheinlichkeitsdichte ausgedrückt wird, ist es einfach, beide Arten von Information in einer gemeinsamen Unsicherheitsangabe zu kombinieren. Jeder Unsicherheitsbeitrag ist nun durch die Varianz (oder die Standardabweichung) einer Verteilung charakterisiert. Und wie Varianzen von Verteilungen miteinander zu kombinieren sind, verrät uns Glg. (16): Setzt man die Additivität der zugehörigen Einflussgrößen voraus, dann addieren sich die Unsicherheiten (Standardabweichungen) quadratisch: Sind also $u_A(x)$ und $u_B(x)$ nach Typ A und B ermittelte Unsicherheiten, ergibt sich die gesamte Unsicherheit zu:

$$u_{\text{gesamt}}(x) = \sqrt{u_A^2(x) + u_B^2(x)} \quad (45)$$

Im Folgenden ein Beispiel für die Kombination von TypA/B Messunsicherheiten, wie es im Praktikum vorkommen kann. Viele weitere ebenfalls fürs Praktikum relevante Beispiele finden Sie in [5].

Beispiel: Mit dem Digitalmultimeter PeakTech 2010 DMM aus dem Praktikum haben Sie Hallspannungen gemessen. Für gewisse Parameter von Probenstrom und Magnetfeld haben Sie als Mittelwert aus fünf Messungen eine Spannung von $\bar{U}_H = 45.3 \text{ mV}$ erhalten. Die Standardabweichung wurde

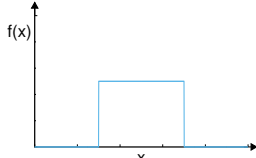
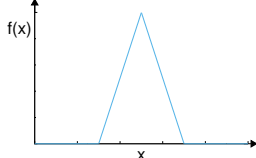
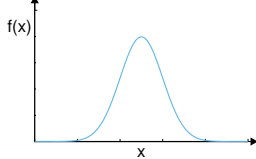
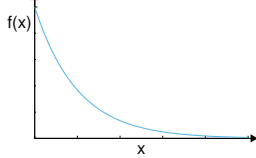
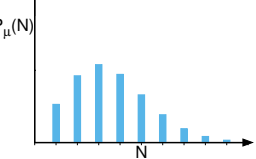
vorliegende Information über die Messgröße X	zugeordnete Verteilung	Erwartungswert Standardabweichung	
obere und untere Grenze $a \leq x \leq b$	Gleichverteilung: $f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$		$\mu = \frac{a+b}{2}$ $\sigma = \frac{b-a}{2\sqrt{3}}$
obere und untere Grenze $a \leq x \leq b$ Randwerte sehr unwahrscheinlich	Dreiecksverteilung: $f(x) = \begin{cases} \frac{4(x-a)}{(b-a)^2} & \text{für } a \leq x \leq \frac{a+b}{2} \\ \frac{4(b-x)}{(b-a)^2} & \text{für } \frac{a+b}{2} \leq x \leq b \\ 0 & \text{sonst} \end{cases}$		$\mu = \frac{a+b}{2}$ $\sigma = \frac{b-a}{2\sqrt{6}}$
bester Schätzwert x_0 , Standardunsicherheit $u_s(x)$	Normalverteilung: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$		$\mu = x_0$ $\sigma = u_s(x)$
bester Schätzwert x_0 , $x \geq 0$	Exponentialverteilung: $f(x) = \frac{1}{\mu} e^{-x/\mu}$		$\mu = x_0$ $\sigma = \mu = x_0$
diskrete Zählereignisse, Einzelmessung N_0	Poissonverteilung: $P_\mu(N) = \frac{\mu^N}{N!} e^{-\mu}$		$\mu = N_0$ $\sigma = \sqrt{\mu} = \sqrt{N_0}$

Tabelle 2: Die eindeutige Zuordnung einer Wahrscheinlichkeitsdichte bzw. Wahrscheinlichkeitsverteilung für einige Beispiele von vorliegender Information über eine kontinuierliche Messgröße X bzw. diskrete Größe N .

gemäß Glg. (40) zu $u_S(\bar{U}_H) = 0.23 \text{ mV}$ ermittelt. Zusätzlich kennen Sie die Toleranzangabe des Herstellers für das Multimeter im Messbereich 200 mV DC: $\pm(0.5\%Mw + 5\text{Dgt.})$. Bei einem Digitalmultimeter bedeutet dies: 0.5% des Messwertes plus 5 mal die kleinste signifikante Stelle auf der Digitalanzeige. Bei einem Messwert von 45.3 mV ist dies: $0.005 \cdot 45.3 \text{ mV} + 0.5 \text{ mV} = 0.73 \text{ mV}$. Die nach dem PME zugeordnete Verteilung ist wieder eine Gleichverteilung und die zugehörige Standardabweichung ist $u_{\text{Tol}}(U) = 0.73/\sqrt{3} = 0.42 \text{ mV}$. Damit ist die gesamte Unsicherheit gemäß Glg. (45): $u_{\text{gesamt}}(U) = \sqrt{u_S^2(\bar{U}_H) + u_{\text{Tol}}^2(U)} = 0.48 \text{ mV}$. Das Ergebnis lautet also: $U_H = (45.3 \pm 0.5) \text{ mV}$.

3.4. Die Fortpflanzung von Messunsicherheiten

In physikalischen Experimenten kann häufig eine interessierende Größe g nicht direkt gemessen werden, sondern muss aus Messungen verschiedener Hilfsgrößen x, y, z, \dots abgeleitet werden. Naturgemäß sind die gemessenen besten Schätzwerte $\hat{x}, \hat{y}, \hat{z}, \dots$ der Hilfsgrößen alle mit Unsicherheiten behaftet: $x = \hat{x} \pm u(\hat{x})$, $y = \hat{y} \pm u(\hat{y})$, $z = \hat{z} \pm u(\hat{z})$, \dots . Die Frage, die sich nun stellt, ist, wie ermittelt man \hat{g} , also den besten Schätzwert, bzw. Erwartungswert von g , und wie propagieren die Unsicherheiten in den Hilfsgrößen in die Unsicherheit der interessierenden Größe g ?

Die Größe g sei dabei mit den gegenseitig unabhängigen Hilfsgrößen über das Gesetz

$$g = g(x, y, z, \dots) \quad (46)$$

verknüpft. Beispiel: Die kinetische Energie $g \equiv E_{\text{kin}}$ sei über die gemessenen Größen m, v via $E_{\text{kin}} = (m/2)v^2$ zu bestimmen. Die Messung der Masse m und der Geschwindigkeit v erfolgt mit völlig verschieden konstruierten Instrumenten und Verfahren. Daher ist eine gegenseitige Beeinflussung der Messungen nicht zu erwarten und m, v dürfen als unabhängig gelten.

Erwartungswert \hat{g} :

In erster Näherung gilt, dass der beste Schätzwert von g , also der Erwartungswert \hat{g} , sich aus den gemessenen Hilfsgrößen $\hat{x}, \hat{y}, \hat{z}, \dots$ bildet gemäß:

$$\hat{g} = \hat{g}(x, y, z, \dots) \approx g(\hat{x}, \hat{y}, \hat{z}, \dots) \quad (47)$$

Streng ist dies nur für eine lineare Beziehung $g(x, y, z, \dots)$ richtig. In der Praxis akzeptabel ist Glg. (47) aber auch dann, wenn $g(x, y, z, \dots)$ sich im Bereich der Unsicherheiten $\hat{x} \pm u(\hat{x}), \hat{y} \pm u(\hat{y}), \dots$ ausreichend linearisieren lässt.

Messunsicherheit (Standardabweichung) $u(\hat{g})$:

Abb. 8 demonstriert anhand einer Variablen x , wie sich die Messunsicherheit von x auf die Unsicherheit der Größe g auswirkt abhängig von der lokalen Steigung der Funktion an der Stelle \hat{x} bzw. \tilde{x} . Obwohl die Unsicherheit von x für alle Werte gleich angenommen wird, $u(x) = u(\hat{x}) = u(\tilde{x})$, ergeben sich drastisch verschiedene Unsicherheiten für g aufgrund der lokalen Steigung:

$$u(\hat{g}) = \Delta \hat{g} = \left. \frac{\partial g}{\partial x} \right|_{\hat{x}} \cdot \Delta x = \left. \frac{\partial g}{\partial x} \right|_{\tilde{x}} \cdot u(\tilde{x}), \quad (48)$$

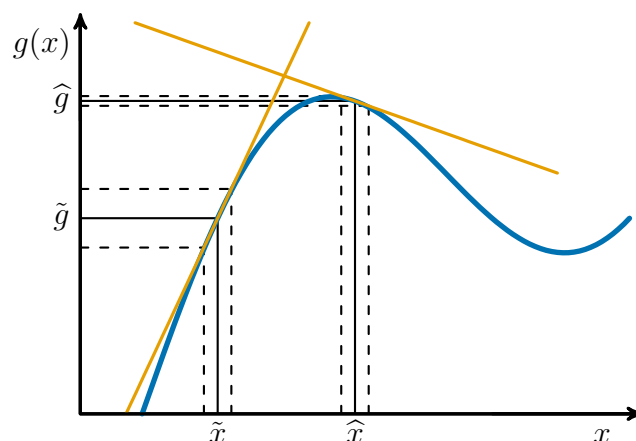


Abbildung 8: Die Auswirkung der Messunsicherheit $u(x)$ einer gemessenen Hilfsgröße x auf die Unsicherheit einer interessierenden physikalischen Größe g ist abhängig von der lokalen Steigung, d.h. der Ableitung der Funktion $\partial g(x)/\partial x$ an der jeweiligen Stelle \hat{x} bzw. \tilde{x} .

und analog für $u(\tilde{g})$. Wichtig zu beachten ist dabei, dass die Funktion $g(x)$ im Bereich $\hat{x} \pm u(\hat{x})$ ausreichend durch ihre Tangente an der Stelle \hat{x} approximierbar, d. h. linearisierbar sein muss. Ist die Funktion im Intervall $\hat{x} \pm u(\hat{x})$ also „zu nervös“ und wechselt z. B. in dem bezeichneten Intervall mehrfach das Vorzeichen ihrer Ableitung, dann sind der folgende Ausdruck, wie auch Glg. (47), *nicht* brauchbar. In diesem Fall müsste man die aus den Verteilungen der Hilfsgrößen resultierende Verteilung der Größe g numerisch, z. B. mit Hilfe einer Monte Carlo Methode [6], ermitteln.

Für die Belange des Praktikums ist es ausreichend anzunehmen, dass $g(x, y, \dots)$ sich in dieser Hinsicht ausreichend gutmütig verhält. Setzt man nun weiter voraus, dass die Messungen der Hilfsgrößen x, y, z, \dots sich gegenseitig *nicht* beeinflussen, diese also untereinander statistisch unabhängig sind, dann folgt aus dem Additionstheorem für Varianzen Glg. (16) das Gauß'sche Fortpflanzungsgesetz für Messunsicherheiten:

$$u(\bar{g}) = \sqrt{\left(\left.\frac{\partial g}{\partial x}\right|_{\hat{x}} \cdot u(\hat{x})\right)^2 + \left(\left.\frac{\partial g}{\partial y}\right|_{\hat{y}} \cdot u(\hat{y})\right)^2 + \left(\left.\frac{\partial g}{\partial z}\right|_{\hat{z}} \cdot u(\hat{z})\right)^2 + \dots} \quad (49)$$

Der Ausdruck $\left.\frac{\partial g}{\partial x}\right|_{\hat{x}}$ bedeutet dabei, dass die partielle Ableitung von g an der Stelle des Wertes \hat{x} auszuwerten ist.

Die Einsicht von Gauß, dass sich die Standardabweichungen wie in (49) quadratisch fortpflanzen, ist für den Experimentator äußerst wichtig. Setzen wir die entsprechenden Zahlenwerte ein, so offenbart die Betrachtung der einzelnen Summanden unter der Wurzel, ob ein Experiment gut konzipiert wurde. Überwiegt ein Summand die anderen deutlich, so zeigt dies, dass die entsprechende Größe zu wenig präzise gemessen wurde, bzw. dass die dafür verwendeten Instrumente und Verfahren verbessert werden müssen. Da die Unsicherheiten quadratisch eingehen, ist es sinnlos zu versuchen, die Beiträge an sich bereits kleiner Summanden noch kleiner zu machen. An den großen Summanden muss angesetzt werden. In diesem Sinne ist (49) auch ein wichtiges Instrument für Planung, Beurteilung und Optimierung von Experimenten.

Abschließend soll darauf hingewiesen werden, dass die einheitliche Anwendung von Glg. (49) auf *alle* Messunsicherheitsbeiträge entscheidend auf den Voraussetzungen und dem Wahrscheinlichkeitsbegriff der Bayes'schen Statistik beruht. Glg. (49) setzt voraus, dass es sich bei allen Beiträgen um Varianzen bzw. Standardabweichungen handelt, d. h. dass allen Unsicherheitsangaben Verteilungen zugrunde liegen. Im Rahmen der konventionellen Statistik ist die Zuordnung einer Verteilung z. B. für systematische Messabweichungen meist nicht möglich. Die Angabe von Unsicherheiten dieser Art müsste dann stets getrennt von den statistischen Unsicherheiten erfolgen. Nur die Fortpflanzung von Typ A Unsicherheiten dürfte dann gemäß Glg. (49) erfolgen, alle anderen Beiträge müssten auf andere Weise, z. B. über den sog. Größtfehler abgeschätzt werden [2]. Allgemein gilt dieses Verfahren deshalb als unpraktikabel und so ist der Bayes'sche Ansatz Grundlage für das durch den GUM und seine Ergänzungen vorgeschlagene international einheitliche Verfahren zur Ermittlung und Angabe von Messunsicherheiten.

Beispiel: Bestimmung der Brennweite F einer Linse

Die Brennweite F einer Linse werde durch die unabhängige Messung der Gegenstandsweite a und der Bildweite b einer Abbildung eines Testgegenstandes ermittelt. Die Messung liefert die Information $a = \hat{a} \pm u(\hat{a})$ und $b = \hat{b} \pm u(\hat{b})$. Wie groß ist der sich daraus ergebende Schätzwert für die Brennweite der Linse \hat{F} und wie lässt sich die Standardabweichung der Verteilung von F abschätzen?

Die Brennweite F errechnet sich als $1/F = 1/a + 1/b$, also

$$F = F(a, b) = \frac{a \cdot b}{a + b} \quad (50)$$

Gemäß (47) folgt dann für den Erwartungswert \hat{f}

$$\hat{F} = \frac{\hat{a} \cdot \hat{b}}{\hat{a} + \hat{b}} \quad (51)$$

Die Standardabweichung $u(\hat{F})$ erhält man nach (49) als

$$u(\hat{F}) = \sqrt{\left(\left.\frac{\partial F}{\partial a}\right|_{\hat{a}} \cdot u(\hat{a})\right)^2 + \left(\left.\frac{\partial F}{\partial b}\right|_{\hat{b}} \cdot u(\hat{b})\right)^2} \quad (52)$$

$$= \frac{1}{(\hat{a} + \hat{b})^2} \sqrt{(\hat{b}^2 \cdot u(\hat{a}))^2 + (\hat{a}^2 \cdot u(\hat{b}))^2} \quad (53)$$

Bemerkung zur Unabhängigkeit von Messgrößen: Der Experimentator könnte auf die Idee kommen, statt jeweils die Strecken a und b zu messen, einmal den Abstand zwischen Gegenstands- und Bildebene $l = a + b$ zu messen und dann in der Folge nur noch eine Strecke, z. B. a zu bestimmen. Auf die Messung von b könnte er dann verzichten, da sich ja b aus a ergibt: $b = l - a$. Wenn dieser Weg gewählt wird, *ist aber die Messgröße b nicht mehr unabhängig von a* , und die obige Analyse nicht mehr zulässig!

Regel: Die interessierende Größe ist stets durch jene Größen darzustellen, die durch eine *unabhängige* Messung ermittelt werden. In diesem Fall ist also die Brennweite durch a und l auszudrücken:

$$F = F(a, l) = \frac{a \cdot (l - a)}{a + (l - a)} = \frac{a \cdot (l - a)}{l} \quad (54)$$

Für den Schätzwert der Brennweite $F = \hat{F} \pm u(\hat{F})$ erhalten wir dann nach (47) und (49)

$$\hat{F} = \frac{\hat{a}(\hat{l} - \hat{a})}{\hat{l}} \quad (55)$$

$$u(\hat{F}) = \sqrt{\left(\frac{\hat{l} - 2\hat{a}}{\hat{l}} u(\hat{a})\right)^2 + \left(\frac{\hat{a}^2}{\hat{l}^2} u(\hat{l})\right)^2} \quad (56)$$

3.5. Angabe des vollständigen Messergebnisses mit Unsicherheit

Das Endergebnis einer Datenanalyse besteht meist in der Angabe eines vollständigen Messergebnisses mit Unsicherheit. Die Zahlenwerte sind dabei sinnvoll auf- oder abzurunden. Auf wie viele signifikante Stellen genau das Messergebnis angegeben wird, hängt dabei von der Unsicherheit ab. Es gelten folgende Regeln:

- Die abschließende Angabe einer Unsicherheit hat mit höchstens zwei signifikanten Stellen zu erfolgen. Ist bei dem Zahlenwert der Unsicherheit die erste von Null verschiedene Ziffer von links eine 1 oder eine 2, dann soll die Angabe zwei signifikante Stellen enthalten, da sonst die Gefahr zu großer Rundungsabweichungen (bis zu 90%!) besteht.
- Das Messergebnis wird auf die selbe Stelle gerundet, wie die Unsicherheit.
- Beim Runden des Messergebnisses gilt: Ist die auf die Rundungsstelle folgende Ziffer ≤ 4 wird abgerundet, ist die Folgeziffer ≥ 5 wird aufgerundet.
- Beim Runden von Unsicherheiten gilt: Abgerundet wird nur, wenn sich der Unsicherheitsbetrag durch das Abrunden um höchstens 5% reduziert. Ansonsten wird *immer aufgerundet!*
- Mehrfachrundungen sind zu vermeiden. Werden gerundete Zwischenergebnisse angegeben, ist trotzdem mit der vollen Anzahl an Stellen weiter zu rechnen.

Beispiel: Bestimmung der Fallbeschleunigung $g \pm u(g)$. Rundung des vollständigen Messergebnisses für verschiedene Werte der Messunsicherheit:

Primärdaten:	g	9,814 73 m/s ²	9,814 73 m/s ²	9,814 73 m/s ²
	$u(g)$	0,063 42 m/s ²	0,083 42 m/s ²	0,015 34 m/s ²
Rundungsstelle:		0,01	0,01	0,001
gerundet:	$u(g)$	0,07 m/s ²	0,08 m/s ²	0,015 m/s ²
Ergebnis:		(9,81 ± 0,07) m/s ²	(9,81 ± 0,08) m/s ²	(9,815 ± 0,015) m/s ²

4. Beurteilende Datenanalyse

Die Methoden der Naturwissenschaften sind geprägt von einem Wechselspiel aus systematischen Naturbeobachtungen und daraus folgend der Entwicklung von Hypothesen und Theorien, meist als mathematische Modelle formuliert, die dann wiederum durch weitere Experimente und Beobachtungen überprüft werden. In diesem Zusammenhang ist es entscheidend, auf der Grundlage vorliegender Messdaten *trotz unvollständiger Information und vorhandenen Messunsicherheiten* zu objektiven Aussagen über die Vereinbarkeit von Theorie und experimentellen Daten zu gelangen.

Bereits in der Fragestellung zeigt sich hier wieder der Unterschied zwischen der Herangehensweise der konventionellen frequentistischen Statistik auf der einen Seite und der Bayes'schen Statistik auf der anderen, die wir in den vorangegangenen Abschnitten bereits diskutiert haben: Intuitiv wird man z. B. geneigt sein zu fragen: „Mit welcher Wahrscheinlichkeit ist bei vorliegendem Datensatz $\{x_i\}$ Theorie XY richtig?“ – Diese Frage lässt sich jedoch nur im Rahmen der Bayes'schen Statistik sinnvoll stellen. Dort lässt sich auch einer Hypothese, einer Aussage wie „Theorie XY trifft zu“ eine Wahrscheinlichkeit zuordnen, die dann keine relativen Häufigkeiten angibt, sondern den Stand der vorliegenden Information abbildet, wie wir bereits in Abschnitt 3.2 gesehen haben.

Im Rahmen der konventionellen Statistik, die sich stets auf relative Häufigkeiten bezieht, wird man deshalb die Frage eher indirekt stellen: „Angenommen Theorie XY ist richtig, wie groß ist dann die Wahrscheinlichkeit, den vorliegenden Datensatz $\{x_i\}$ zu messen?“ Diese Frage bezieht sich auf die Grundgesamtheit der Datensätze zur Größe X , deren relative Häufigkeiten wir im Rahmen der konventionellen Statistik immerhin abschätzen können. Ist die resultierende Wahrscheinlichkeit dann zu gering, werden wir die Theorie ablehnen, ist sie hoch werden wir geneigt sein, das Modell zu akzeptieren.

Die Grundidee von Hypothesentests wollen wir anhand einiger Beispiele zunächst im Rahmen der konventionellen frequentistischen Statistik erläutern. Abschnitt 4.3 gibt dann einen kleinen Ausblick darauf, wie Hypothesentests auch im Rahmen der Bayes'schen Statistik durchgeführt werden können.

4.1. Hypothesentests in der konventionellen Statistik

Unter einer *Hypothese* versteht man in der Datenanalyse eine Annahme über die Verteilung einer Messgröße, z. B. die Annahme dass die Verteilung der Schwingungsdauer T eines Pendels den Mittelwert $\bar{T} = 5.5$ s aufweist. Der *Test* einer Hypothese ist ein Prüfverfahren, das man anwendet, um die Entscheidung zu objektivieren, ob eine bestimmte Hypothese mit den vorliegenden Daten verträglich ist, oder nicht. Im ersten Fall wird die Hypothese *angenommen*, im zweiten Falle *verworfen*, denn „es besteht ein objektiver Grund zur Annahme, dass die Daten nicht mit der Hypothese verträglich sind“.

4.1.1. Hypothesentest in einer einfachen Situation

Wir betrachten die einfachste Variante eines solchen Tests und testen die Hypothese, wonach eine gemessene physikalische Größe einen spezifischen Wert aufweisen soll:

Beispiel: Das Stromnormal

Im Labor haben wir ein Stromnormal, das einen Strom von $15 \mu\text{A}$ liefern soll. Im Praktikum soll die Hypothese, wonach der Ausgangsstrom des Normals $15 \mu\text{A}$ betrage (Herstellerangabe), durch Messungen untermauert oder widerlegt werden. Dazu wird eine Serie von 50 Strommessungen durchgeführt und gemäß Abschnitt 2.1 analysiert. Das Resultat liefert einen mittleren Strom von $\bar{I} = 14.85 \mu\text{A}$ mit einer empirischen Standardabweichung von $s = 0.3 \mu\text{A}$. Muss daraus geschlossen werden, dass die Stromangabe $15 \mu\text{A}$ nicht stimmt?

Test: Wir betrachten den gemessenen Mittelwert als eine Stichprobe, die aus einer normalverteilten Gesamtheit gezogen wurde. Wir nehmen zunächst an, unsere Ausgangshypothese sei korrekt. Damit hat die Normalverteilung den Mittelwert $\mu = 15 \mu\text{A}$. Die Standardabweichung der Verteilung schätzen wir durch die empirische Stichprobenvarianz ab und erhalten $\frac{s}{\sqrt{50}} = \frac{0.3}{\sqrt{50}} = 0.042 \mu\text{A}$.

Nun wählen wir ein *Signifikanzniveau* (eine Vertrauensgrenze) α , z. B. $\alpha = 1\%$. Sollte die Wahrscheinlichkeit dafür, dass die gefundene Abweichung unseres Messwertes vom hypothetischen Mittelwert rein auf zufälligen Schwankungen beruht kleiner sein als α , dann betrachten wir diese Abweichungen als *statistisch signifikant* und lehnen die Hypothese ab.

Die Differenz des gemessenen Mittelwertes zum hypothetischen Wert beträgt in unserem Fall $(15.0 - 14.85) \mu\text{A} = 0.15 \mu\text{A}$, oder, ausgedrückt in Einheiten der Standardabweichung, $\frac{0.15}{0.042} = 3.5$.

Die Wahrscheinlichkeit dafür, experimentell einen Mittelwert \bar{I} zu erhalten, der kleiner oder gleich $14.85 \mu\text{A}$ ist, liegt, das zeigen Tabellen der Verteilungsfunktion der Normalverteilung, bei 0.02% also deutlich unter dem von uns gewählten Niveau α . Wir haben deshalb objektiven Grund anzunehmen, dass die Daten nicht mit der Hypothese verträglich sind.

***p*-Quantil:** In diesem Beispiel stellt der Wert 3.5 Einheiten der Standardabweichung, bzw. der entsprechende Wert auf der Stromskala, das „ 0.02% -Quantil“ dar. Das *p*-Quantil einer gegebenen Verteilung ist also ein Messwert mit der Eigenschaft, dass verglichen mit ihm im Mittel $p \cdot 100\%$ der Messwerte kleiner sind und mindestens $(1 - p) \cdot 100\%$ größer ausfallen.

Zusammenfassung: Schritte für den Hypothesentest

1. Zunächst spezifiziere man die Hypothese (oft besteht diese in der Angabe eines Parameterwertes), so dass daraus die Verteilungsfunktion bestimmt werden kann, aus der die Stichprobe stammen müsste sofern die Hypothese richtig ist.
2. Dann bestimme man die statistischen Maßzahlen der Stichprobe z. B. Mittelwert, Standardabweichung, etc.
3. Man wähle ein Signifikanzniveau α ; je empfindlicher man auf Abweichungen von der Hypothese testen will umso größer wähle man α ; je sicherer man hingegen sein will, dass gefundene Abweichungen systematischer Natur sind, desto kleiner wähle man α .
4. Nun bestimmt man das α -Quantil der Verteilung (oder das $(1 - \alpha)$ -Quantil, je nach Situation) und vergleicht es mit dem erhaltenen Messwert. Ist dieser weiter vom hypothetischen Mittelwert entfernt als das α -Quantil, so ist die Hypothese abzulehnen. Alternativ kann man auch einer Tabelle der *p*-Quantile den zum Messwert gehörigen *p*-Wert entnehmen und ihn mit dem gewählten Wert α vergleichen. Ist $p < \alpha$, so ist die Wahrscheinlichkeit die Stichprobe zu

realisieren zu klein, falls die Hypothese wahr ist. Es besteht dann ebenfalls Grund anzunehmen, dass die Daten nicht mit der Hypothese verträglich sind.

4.2. Der χ^2 -Test: Vergleiche Theorie und Experiment

Der χ^2 -Test (Chi-Quadrat-Test) liefert ein Maß für die Verträglichkeit eines Datensatzes mit einer theoretischen Vorhersage für den Fall dass Datenwerte y_i als Funktion eines gewissen Merkmals x_i erfasst wurden [14, 15]. Zum Beispiel könnte eine experimentell ermittelte Häufigkeitsverteilung $h(x)$ einer Messgröße X mit einer hypothetischen Verteilungsfunktion $\varphi(x)$ verglichen werden oder die experimentell gefundene Temperaturabhängigkeit einer physikalischen Größe $c(T)$ zur Überprüfung einer Theorie herangezogen werden, die $\tilde{c}(T)$ vorhersagt. Der Grundgedanke des χ^2 -Tests ist dabei sehr einfach: Man quantifiziert die Abweichungen von theoretischer Vorhersage und experimentellen Daten mittels einer Größe χ^2 . Ist diese zu groß, so wird beispielsweise die Hypothese: „ $\varphi(x)$ repräsentiert die Verteilungsfunktion der Daten“ verworfen. Liegt der Unterschied unterhalb eines gewissen Wertes, der natürlich von der Größe des Datensatzes abhängt, so wird die Hypothese angenommen. Zunächst müssen wir also die Abweichung von Messwerten und theoretischer Erwartung quantifizieren und Aussagen über die Verteilung dieser „Abweichungsgröße“ χ^2 treffen.

Summe der quadratischen Abweichungen und χ^2 -Verteilung

Experimentell sei ein Datensatz $\{y_i\}$ mit zugehörigen Unsicherheiten $u(y_i)$ ermittelt worden, z. B. die empirische Verteilung von Häufigkeiten der erhaltenen Werte x_i einer Messgröße X . Dann beschreibt die Quadratsumme der gewichteten Abweichungen oder auch *Residuen* $R_i = (y_i - \varphi(x_i))/u(y_i)$ wie gut die Messwerte mit einem theoretisch erwarteten Verhalten $\varphi(x_i)$ verträglich sind, also:

$$\chi^2 = \sum_i^n R_i^2 = \sum_i^n \left(\frac{y_i - \varphi(x_i)}{u(y_i)} \right)^2 \quad (57)$$

Um als nächstes Aussagen über die Verteilung der Größe χ^2 treffen zu können, ist folgender Satz über die Verteilung der Quadrate von normalverteilten Zufallsgrößen der Ausgangspunkt:

Seien $\{R_1, R_2, \dots, R_k\}$ unabhängige und gemäß $N(0, 1)$ standardnormalverteilte Zufallsgrößen. Die Größe $Z_k = \sum_i^k R_i^2$ ist dann χ^2 -verteilt mit k Freiheitsgraden und der Dichtefunktion:

$$f_{\chi^2}(z, k) = \frac{z^{k/2-1} e^{-z/2}}{2^{k/2} \Gamma(k/2)} \quad (58)$$

Wichtig ist also, dass die R_i *statistisch unabhängig* und *standardnormalverteilt* sind. Sind dabei die Zufallsgrößen Y_i nach $N(\mu_i, \sigma_i)$ normalverteilt und ist die Theorie $\varphi(X_i)$ korrekt, dann beschreibt $\varphi(x_i)$ gerade den Erwartungswert von Y_i , also $\varphi(x_i) = \mu_i$ und $u(y_i) = \sigma_i$. Damit entsprechen die R_i aber gerade der Transformation aus Glg. (36) und sind somit $N(0, 1)$ normalverteilt. Ob die Größen Y_i allerdings normalverteilt sind muss im Einzelfall geklärt werden. Typisch wäre, dass es sich bei den experimentell ermittelten y_i selbst bereits um Mittelwerte handelt und so der Zentrale Grenzwertsatz zumindest die näherungsweise Normalverteilung sichert, oder dass Y_i binomialverteilt ist aber man

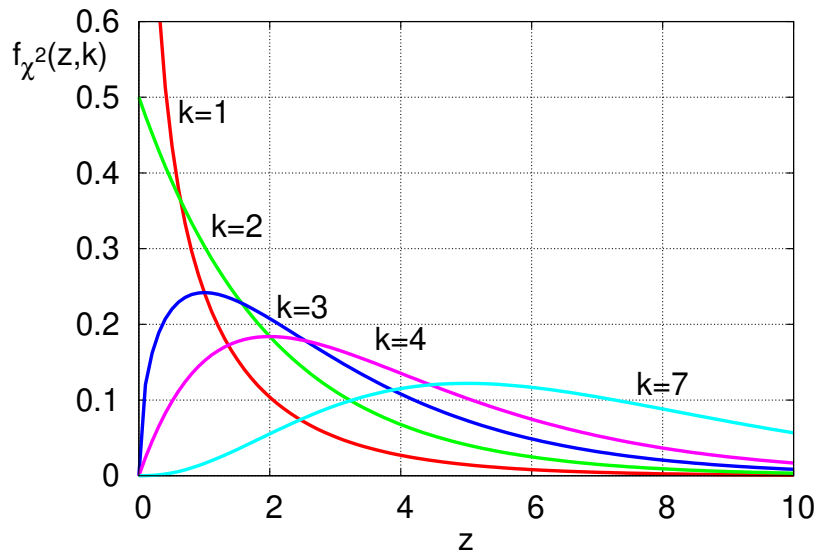


Abbildung 9: Die Dichtefunktion der χ^2 -Verteilung $f_{\chi^2}(z, n)$ für verschieden Freiheitsgrade n .

die Normalverteilungsnäherung für große n wie in Glg. (31) zugrunde legen kann. Für diese beiden praxisrelevanten Fälle lernen wir weiter unten je ein Beispiel kennen.

Eine weitere wichtige Voraussetzung ist die *Unabhängigkeit* der Residuen. Sie legt direkt die Zahl der Freiheitsgrade k der Verteilung fest. Diese Zahl kann durchaus kleiner sein als die Anzahl n der Summanden in Glg. (57), sei es weil die y_i z. B. relative Häufigkeiten sind und so wegen der Normierung $\sum_i^n y_i = 1$ die Zahl der Freiheitsgrade um einen reduziert wird (aus $n - 1$ bekannten y_i ließe sich y_n ausrechnen), oder weil die Funktion $\varphi(x)$ evtl. freie Parameter enthält, die durch den Datensatz $\{y_i\}$ erst festgelegt werden, was die Zahl der Freiheitsgrade eben um die Anzahl dieser freien Parameter verringert.

Für $k = 1$ lässt sich Glg. (58) übrigens leicht nachvollziehen, indem man in der Wahrscheinlichkeitsverteilung der Normalverteilung $z = x^2$ substituiert:

$$N(x, 0, 1) dx = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} dx^2 = \underbrace{\frac{z^{-1/2}}{\sqrt{2\pi}} e^{-z/2}}_{f_{\chi^2}(z, 1)} dz$$

Abb. 9 zeigt die Dichtefunktion $f_{\chi^2}(z, k)$ der χ^2 -Verteilung für einige Freiheitsgrade. Die p -Quantile der χ^2 -Verteilung sind tabelliert, siehe Anhang A.

4.2.1. Der χ^2 -Test: Vergleiche Datensatz mit hypothetischer Verteilung $\varphi(x)$

Folgendes sei die Ausgangslage: In einem Experiment hat man die Größe X gemessen und nach der Messung liegt ein Satz Daten $\{x_i, \dots, x_n\}$ vor. Hypothese: Wir verfügen über gute Gründe anzunehmen, dass die Daten einer bestimmten Verteilung $\varphi(x)$ folgen. Die grundsätzliche Strategie

wird nun sein, zunächst den Datensatz als Histogramm darzustellen und dann die so ermittelten Klassenhäufigkeiten b_j mit den aus $\varphi(x)$ ermittelten Häufigkeiten zu vergleichen.

Vorgehen:

1. Histogramm: Unterteile die x -Achse in k -Intervalle (Klassen) I_1, I_2, \dots, I_k derart, dass jedes Intervall mehr als 5...10 Datenwerte enthält. Eventuell müssen Randklassen zusammengefasst werden, damit diese Belegung zustandekommt. Dann bestimme man die Anzahl Daten b_j , die in jeder der Klassen I_j liegen. Die Summe aller b_j ergibt die Anzahl Daten $n = \sum_{j=1}^k b_j$. Diese Bedingung wird die Zahl der Freiheitsgrade auf $k - 1$ reduzieren.
2. Aus der hypothetischen Verteilung $\varphi(x)$ berechne man für jedes Intervall I_j die Wahrscheinlichkeit p_j , mit der ein Datenpunkt einen Wert innerhalb des j -ten Intervalls annimmt. Daraus berechne man die Anzahl der theoretisch in der Klasse I_j erwarteten Datenwerte $e_j = np_j$.
3. Bezüglich einer Klasse j sind die Häufigkeiten b_j Binomialverteilt (Wahrscheinlichkeit, dass bei n Messungen der Größe X der Messwert b_j -mal in die Klasse j fällt). Nehmen wir nun an es liegen genügend Messwerte in jeder Klasse vor (≥ 10 , geschickte Klasseneinteilung bei der Konstruktion des Histogramms wird also vorausgesetzt!), so dass gemäß Glg. (31) die Häufigkeiten b_j in jeder Klasse näherungsweise normalverteilt sind. Dann können wir die Unsicherheit wie folgt abschätzen: $u(b_j) \approx \sqrt{np_j} = \sqrt{e_j}$ und berechnen die Abweichung der Daten von der Hypothese gemäß:

$$\chi_0^2 = \sum_{j=1}^k \frac{(b_j - e_j)^2}{e_j} = \sum_{j=1}^k \frac{b_j^2}{e_j} - n. \quad (59)$$

4. Man wähle eine Signifikanzzahl (Vertrauensgrenze) α , z. B. 5%, 1%, oder ähnlich.
5. Mit Hilfe einer Tabelle der χ^2 -Verteilung mit $k - 1$ Freiheitsgraden (siehe Anhang) bestimme man das $(1 - \alpha)$ -Quantil $\chi_{1-\alpha}^2$, so dass mit der Wahrscheinlichkeit $1 - \alpha$ zufällige Abweichungen der experimentellen Häufigkeiten von der theoretisch erwarteten Verteilung kleiner sind als $\chi_{1-\alpha}^2$:

$$P(\chi^2 \leq \chi_{1-\alpha}^2) = 1 - \alpha \quad (60)$$

6. Ist $\chi_0^2 \leq \chi_{1-\alpha}^2$, so wird die Hypothese angenommen. Ist hingegen $\chi_0^2 > \chi_{1-\alpha}^2$, so wird sie verworfen, denn es besteht objektiv Grund zur Annahme, dass der Datensatz $\{x_i\}$ nicht mit der Verteilung $\varphi(x)$ kompatibel ist.

Ein Beispiel: Mendels Erbsen

Folgendes ist eines der klassischen Experimente, die Gregor Mendel 1865 publizierte um seine Vererbungstheorie zu bestätigen. Er untersuchte Erbsen um die dominant-rezessive Vererbung einzelner Merkmale zu belegen. Dabei betrachtete er in einem bestimmten Experiment bei 556 Erbsen die Merkmale Form (rund oder kantig) und Farbe (grün oder gelb) und beobachtete folgende Häufigkeiten der Merkmalskombinationen:

rund/gelb	rund/grün	kantig/gelb	kantig/grün
315	108	101	32

während seine Theorie Häufigkeiten dieser Merkmale im Verhältnis 9:3:3:1 vorhersagt. Kann die Theorie nun durch diese Daten bestätigt werden? Wir wählen als Signifikanzniveau $\alpha = 5\%$. Dies führt zu einem 95%-Quantil, welches sich für $k - 1 = 3$ Freiheitsgrade aus der Tabelle im Anhang ablesen lässt $\chi_{0.95}^2 = 7.81$. χ_0^2 ergibt sich andererseits nach (59) zu:

$$\chi_0^2 = \frac{16}{556} \left(\frac{315^2}{9} + \frac{108^2}{3} + \frac{101^2}{3} + \frac{32^2}{1} \right) - 556 = 0.47 < 7.81$$

Damit nehmen wir die Hypothese an und betrachten die Theorie als bestätigt. Jedoch ist der kleine Wert von χ^2 durchaus auffällig. Ein Blick in die χ^2 Tabelle im Anhang lehrt uns, dass im Falle von zufälligen Schwankungen um die von der Theorie vorhergesagten Mittelwerte solch ein kleines (oder noch kleineres) χ^2 nur in weniger als 10% aller Fälle vorkommt. Genauer gehört ein $\chi_p^2 = 0.47$ zu einer Wahrscheinlichkeit $p = P(\chi^2 \leq 0.47) = 0.07$. Das ist aber grenzwertig: Deutet ein unwahrscheinlich großes χ^2 darauf hin, dass die Theorie falsch ist, so folgt auch für ein unwahrscheinlich kleines χ^2 , dass es sich nicht um zufällige Abweichungen vom Mittelwert handelt. Vielmehr lässt sich schließen, dass entweder die Daten manipuliert wurden um eine bessere Übereinstimmung mit der Theorie zu erhalten (!), oder dass die Unsicherheit der Messwerte falsch eingeschätzt wurde. Im vorliegenden Fall ist tatsächlich mehrfach der Verdacht geäußert worden Gregor Mendel (oder sein Gärtner ;-) könnte die Daten manipuliert haben, um die Theorie besser aussehen zu lassen [16].

4.2.2. Der χ^2 -Test: Vergleich einer Messreihe $y_i(x_i)$ mit theoretischer Voraussage $\eta(x)$

Bei physikalischen Experimenten treffen wir häufig folgende Situation an: Wir messen eine Größe y , die von einer Variablen x abhängt. Zum Beispiel könnten es sich um die Temperaturabhängigkeit von Materialeigenschaften handeln, oder die Abstandsabhängigkeit der elektrischen Feldstärke von einem geladenen Objekt oder, wie im nächsten Beispiel, der Zerfall von Bierschaum als Funktion der Zeit.

In jedem Fall messen wir die Größe y für n verschiedene x_i und bestimmen die zugehörigen Unsicherheiten, so dass wir für jedes x_i einen Messwert y_i sowie die Unsicherheit $u(y_i)$ erhalten. Dann berechnen wir die Abweichung unserer Daten von der Theorie gemäß:

$$\chi^2 = \sum_i^n \left(\frac{y_i - \eta(x_i)}{u(y_i)} \right)^2 \quad (61)$$

Man beachte, dass dabei Messwerte mit einer großen Unsicherheit geringer gewichtet werden als solche mit kleinem $u(y_i)$.

Nun ist es in den meisten Fällen jedoch so, dass das mathematische Modell, das wir gerne überprüfen möchten, nicht vollständig durch die physikalische Theorie festgelegt ist, sondern freie Parameter enthält, die durch unsere Messdaten $y_i(x_i)$ erst bestimmt werden müssen. Beispielsweise könnte es sein, dass wir für $\eta(x)$ einen linearen Zusammenhang erwarten: $\eta(x) = a \cdot x + b$, die Parameter a und b jedoch unbekannt sind. Generell wird man die freien Parameter dann so wählen, dass χ^2 minimal wird. Für eine lineare Funktion $\eta(x)$ ist dieses Problem analytisch lösbar und das Verfahren

nennt sich *lineare Regression*. Durch Nullsetzen der Ableitungen: $\frac{\partial \chi^2(a,b)}{\partial a} = 0$ und $\frac{\partial \chi^2(a,b)}{\partial b} = 0$ erhält man unter der Voraussetzung, dass alle $u(y_i)$ gleich groß sind:

$$a = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} \quad (62)$$

$$b = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2} \quad (63)$$

Im allgemeineren Fall, dass die Funktion $\eta(x)$ nichtlinear ist oder dass die Messunsicherheiten $u(y_i)$ unterschiedlich ausfallen, wird man χ^2 numerisch minimieren. Auch wenn wir auf die dabei angewandten Methoden an dieser Stelle nicht näher eingehen können, sei als Stichwort der *Levenberg-Marquardt Algorithmus* erwähnt, der hier in ganz vielen praktischen Fällen Anwendung findet. Dem interessierten Leser sei dazu das Kapitel *Modeling of Data* in den *Numerical Recipes* empfohlen, welches nicht nur die mathematischen Zusammenhänge leicht verständlich aufbereitet sondern auch fertige programmierte Routinen anbietet [17].

Nehmen wir nun an, wir haben mit dem geeigneten Verfahren s freie Parameter durch Minimierung von χ^2 über unsere n Messpunkte festgelegt. Dann ergibt sich aus unseren Messdaten und Glg. (61) ein Wert χ_{\min}^2 für das so erhaltene minimale χ^2 und wir können den χ^2 Test durchführen, indem wir wieder eine Vertrauensgrenze α festlegen und das $1 - \alpha$ -Quantil der χ^2 Verteilung, diesmal für $n - s$ Freiheitsgrade nachschlagen. Und dann gilt wie immer: ist χ_{\min}^2 größer lehnen wir das Modell ab, ist es kleiner akzeptieren wir.

Auch hier nochmal eine Zusammenfassung des Verfahrens:

- Durch unser Experiment gewinnen wir n Messwerte der Größe Y mit den zugehörigen Unsicherheiten, also n Zahlentripel der Form $\{x_i, y_i, u(y_i)\}$.
- Mit einem geeigneten Verfahren wählen wir die s freien Parameter in unserem mathematischen Modell $\eta(x; a_1, \dots, a_s)$ so, dass χ^2 aus (61) minimal wird.
- Mit dem so erhaltenen χ_{\min}^2 führen wir einen χ^2 -Test mit $n - s$ Freiheitsgraden durch. Für zu große χ_{\min}^2 ist das Modell $\eta(x)$ abzulehnen, bei zu kleinen Werten erhebt sich u. U. der Verdacht auf Datenmanipulation oder die Messunsicherheiten wurden falsch eingeschätzt.

Ein Beispiel: Der Zerfall von Bierschaum

Betrachten wir ein Beispiel und untersuchen den Zerfall von Bierschaum. Dabei wird in einem zylindrischen Glas zu verschiedenen Zeiten t_i die Höhe h_i des Bierschaumes gemessen. Das Experiment wird wiederholt, am besten mit der bevorzugten Marke des Experimentators ;-), und zu jedem Zeitpunkt t_i die mittlere Höhe \bar{h}_i mit zugehöriger Unsicherheit $u(\bar{h}_i)$ ermittelt. Abbildung 10 zeigt Messdaten aus [15], die für Erdinger Weißbier gewonnen wurden. Es wurden im Zeitraum zwischen 0 und 360 s 15 Messpunkte \bar{h}_i und deren zugehörige Unsicherheiten $u(\bar{h}_i)$ bestimmt.

Nun fragen wir uns, was als mathematisches Modell für den Bierschaumzerfall in Frage kommt. Wenn die Zerfallsrate proportional zum Volumen des Schaums und damit zur Höhe in einem zylindrischen Glas ist, also $dh/dt \propto h$, dann erwarten wir exponentiellen Zerfall gemäß:

$$h(t) = h_0 e^{-t/\tau} \quad (64)$$

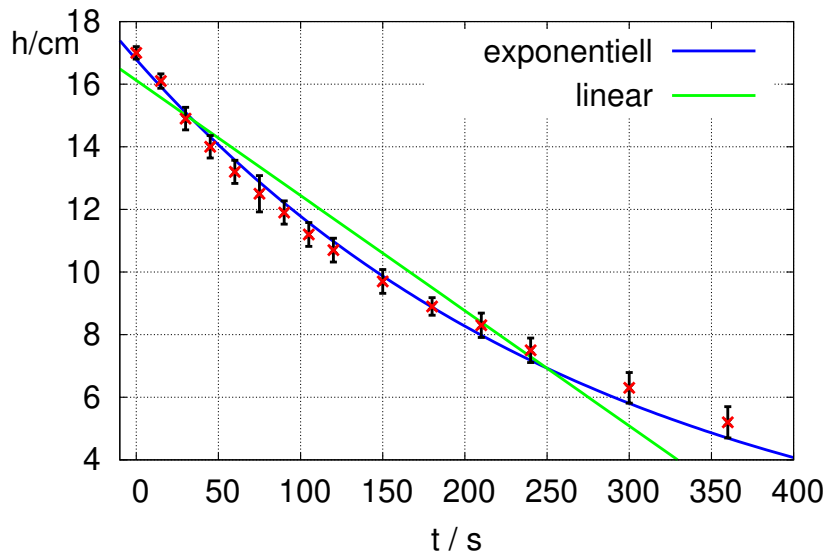


Abbildung 10: Der Zerfall von Bierschaum bei Erdinger Weißbier: Höhe des Schaums als Funktion der Zeit $h(t)$, die Messunsicherheiten sind als Fehlerbalken dargestellt. Die Daten sind aus [15] entnommen. Durchgezogene Linien: exponentielles und lineares Zerfallsmodell nach Gln. (64) und (65).

mit der Anfangshöhe h_0 und der Zeitkonstante τ , nach welcher die Höhe auf $1/e$ des Anfangswertes abgefallen ist. Als Alternativmodell testen wir einen linearen Zerfall gemäß:

$$h(t) = h_0 (1 - t/t_0), \quad t < t_0 \quad (65)$$

mit einer Anfangshöhe h_0 und einer charakteristischen Zeit t_0 , nach deren Verstreichen kein Schaum mehr vorhanden ist. Zunächst minimieren wir χ^2 mit Hilfe des Levenberg-Marquardt Algorithmus und erhalten dabei folgende Parameter:

	h_0/cm	t_0/s bzw. τ/s	χ_{\min}^2
linear	16.1	438.2	91.5
exponentiell	16.8	282.5	10.1

Nur, ob diese Parameter irgendeine Relevanz haben, erfahren wir erst, wenn klar ist ob die beobachteten Abweichungen der Messdaten von dem jeweiligen theoretischen Modell mit hoher Wahrscheinlichkeit auf statistische Messunsicherheiten zurückzuführen sind. Salopp gesagt, kann man auch „einen Elefanten mit einer Geraden anfitzen“ d.h. Parameter finden, die das χ^2 minimieren, die jedoch, da das Modell ungültig ist, völlig bedeutungslos sind. Auskunft darüber, ob dem so ist gibt uns nun der χ^2 -Test. Wir bestimmen das 95%-Quantil der χ^2 -Verteilung mit $n - s = 13$ Freiheitsgraden, d. h. wir wählen eine Vertrauensgrenze von 5% und erhalten aus der Tabelle im Anhang den Wert $\chi_{95\%}^2 = 22.4$. Das bedeutet, dass man im Falle der Gültigkeit des Modelles $h(t)$ mit einer Wahrscheinlichkeit von 95% Werte für χ^2 erhält, die kleiner oder gleich 22.4 sind. Damit zeigt sich, dass der exponentielle Zerfall ein akzeptables Modell darstellt, während wir das lineare

Zerfallsgesetz verwerfen müssen. Dies ist keine knappe Entscheidung, da sich zeigt, dass die Wahrscheinlichkeit ein $\chi^2 \geq 91.5$ zu erhalten bei 13 Freiheitsgraden lediglich $P(\chi^2 \geq 91.5) = 2.8 \cdot 10^{-14}$ beträgt. Wir haben also objektiven Grund die Hypothese „linearer Zerfall“ zu verwerfen, während sich für den exponentiellen Zerfall eine Wahrscheinlichkeit $P(\chi^2 \geq 10.1) = 68.6\%$ ergibt, d. h. in etwa zwei von drei Experimenten werden wir Werte von χ^2_{\min} erhalten, die größer sind, als wir sie im vorliegenden Experiment gefunden haben, in einem von drei Experimenten werden im Mittel die Werte kleiner sein. Dies ist realistisch und damit ist die Theorie akzeptiert.

4.3. Hypothesentests in der Bayes'schen Statistik

Wie bereits in Abschnitt 3.2 über Messunsicherheiten erwähnt, gewinnen die Methoden der Bayes'schen Statistik zunehmend an Bedeutung für die naturwissenschaftliche Datenanalyse [3, 6, 9, 10]. Eine Einführung in die Bayes'sche Statistik würde den Rahmen dieses Skriptes jedoch sprengen. Deshalb müssen wir den Leser auf die verfügbare Literatur verweisen (sehr gut für den Einstieg geeignet ist z. B. [10]) und werden uns im Folgenden auf ein paar wenige Hinweise und ein Beispiel beschränken.

Wie bereits mehrfach erwähnt, wird in der Bayes'schen Statistik ein anderer Wahrscheinlichkeitsbegriff zugrunde gelegt als in der konventionellen frequentistischen Sichtweise. Wahrscheinlichkeit wird nicht mehr nur als relative Häufigkeit verstanden, sondern gibt den Grad unserer Überzeugung an, mit dem wir auf der Basis der vorliegenden unvollständigen Information erwarten können, dass ein Ereignis eintritt. Damit beschreibt Wahrscheinlichkeit eher die Information, die wir über einen Sachverhalt zur Verfügung haben. Diese kann uns als (empirische) Verteilung relativer Häufigkeiten vorliegen, kann aber auch von anderer Form sein. Wie sich auch in diesen Fällen Wahrscheinlichkeitsverteilungen eindeutig zuordnen lassen, haben wir im Abschnitt 3.2 bereits diskutiert.

In der Bayes'schen Herangehensweise betrachtet man grundsätzlich *bedingte Wahrscheinlichkeiten*. In diesem Kontext bedeutet z. B. $P(A|B)$ die Wahrscheinlichkeit dafür, dass das Ereignis A eintritt *unter der Bedingung*, dass B vorliegt. Die Wahrscheinlichkeit, dass A und B tatsächlich zusammen auftreten ist dann $P(A, B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$. Daraus folgt das *Bayes'sche Theorem*:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}, \quad (66)$$

welches grundsätzlich den Ausgangspunkt für eine Datenanalyse im Rahmen der Bayes'schen Statistik bildet. Die Relevanz dieser zunächst abstrakten Überlegung erschließt sich, sobald man in Glg. (66) die Größen A und B durch *Hypothese* und *Messdaten* ersetzt. Dann liest sich obige Gleichung so:

$$P(\text{Hypothese} | \text{Daten}) = \frac{P(\text{Daten} | \text{Hypothese}) \cdot P(\text{Hypothese})}{P(\text{Daten})} \quad (67)$$

Dabei wird die linke Seite der Gleichung $P(\text{Hypothese} | \text{Daten})$ die *Posteriori-Wahrscheinlichkeit* genannt. Sie setzt sich zusammen aus der *Likelihood* $P(\text{Daten} | \text{Hypothese})$, die auch in der konventionellen Statistik verwendet wird, und der sogenannten *Priori-Wahrscheinlichkeit* $P(\text{Hypothese})$. Letztere stellt unseren Kenntnisstand *vor* der Messung dar und könnte so sinngemäß auch als „Vorurteil“ bezeichnet werden. Glg. (67) zeigt dann auf, wie neue Messdaten unsere Kenntnis über

einen Sachverhalt modifizieren. Der Ausdruck $P(\text{Daten})$ im Nenner von Glg. (67) stellt lediglich einen Normierungsfaktor dar.

Um die Begriffe zu verdeutlichen: Angenommen wir würfeln und fragen uns nach der Wahrscheinlichkeit, bei einem Wurf die „1“ zu erhalten. *A priori* würden wir einen ungezinkten Würfel erwarten, und deshalb davon ausgehen, dass sich die Anzahl erhaltener 1er bei N Würfeln mit einer Binomialverteilung mit $p_0 = 1/6$ beschreiben lässt. Finden wir nun in 100 Würfeln 25 mal die 1, dann werden wir jedoch geneigt sein, unsere Meinung zu revidieren. Genau dieser „Lernprozess“ ist in Glg. (67) formalisiert.

An Glg. (67) wird auch deutlich, warum die Bayes'sche Statistik sich eines anderen Wahrscheinlichkeitsbegriffes bedient, als es in der frequentistischen Herangehensweise üblich ist: Damit eine Interpretation des Bayes-Theorems in der Form Glg. (67) Sinn macht, muss man einer Hypothese eine Wahrscheinlichkeit zuordnen können. Dies ist mit einem Wahrscheinlichkeitsbegriff im Sinne „relativer Häufigkeiten des Auftretens von Ereignissen“ nicht möglich.

Noch ein anderer Punkt wird an dieser Stelle klar: In der bisherigen Behandlung von Hypothesentests war der Ausgangspunkt immer von der Art: „angenommen die Hypothese gilt, wie groß ist die Wahrscheinlichkeit, die tatsächlich gemessenen Daten zu erhalten?“. In der Sprache bedingter Wahrscheinlichkeit wäre dies also eine Aussage der Art $P(\text{Daten} | \text{Hypothese})$. Nur solche Aussagen können im Rahmen der konventionellen Statistik getroffen werden. Häufig wird dies unbewusst jedoch mit der umgekehrten Aussage gleichgesetzt, zum Beispiel, wenn man unkorrekter Weise im Rahmen der frequentistischen Statistik davon spricht „der wahre Wert der Messgröße liegt mit einer Wahrscheinlichkeit vom 68% im ermittelten Intervall $\bar{x} \pm s_{\bar{x}}$ “. Letztere Aussage ist von der Art $P(\text{Hypothese} | \text{Daten})$. Glg. (67) zeigt nun, dass beide Arten von Aussagen im Allgemeinen *nicht* äquivalent sind, aufgrund des fehlenden Faktors $P(\text{Hypothese})/P(\text{Daten})$.

Ein Beispiel dafür: Bereits wiederholt wurden an großen Teilchenbeschleunigern Ereignisse beobachtet, die im Rahmen des Standardmodells der Teilchenphysik, das ansonsten hervorragend experimentell bestätigt ist, sehr unwahrscheinlich sind, sagen wir im Rahmen des Standardmodells ein Wahrscheinlichkeit von unter 1% aufweisen. Diese Aussage ist wieder von der Art $P(\text{Daten} | \text{Hypothese})$. Damit wurde jedoch *nicht* das Standardmodell zu 99% widerlegt! Glg. (67) sagt warum: die *Posteriori-Wahrscheinlichkeit* ist *Likelihood* mal *Prior*. Alleine aus der Likelihood lässt sich also keine gültige Aussage über das Modell treffen. Und die *Priori-Wahrscheinlichkeit* $P(\text{Hypothese})$ muss im Falle des bisher hervorragend bestätigten Standardmodells als sehr hoch eingestuft werden [11].

In vielen Fällen werden sowohl die Messdaten als auch die Hypothesen durch kontinuierliche Größen beschrieben. Beispielsweise betrachten wir den kontinuierlichen Parameter μ , den wir durch den Mittelwert \bar{x} einer Messreihe abschätzen. Analog zu Glg. (67) lässt sich hierfür das Bayes-Theorem auch mit kontinuierliche Wahrscheinlichkeitsdichten formulieren:

$$f(\mu | \bar{x}) = \frac{f(\bar{x} | \mu) \cdot f_0(\mu)}{\int f(\bar{x} | \mu) \cdot f_0(\mu) d\mu} \quad (68)$$

Der *Prior* wird nun anstelle von $P(\text{Hypothese})$ durch die Wahrscheinlichkeitsdichte $f_0(\mu)$ beschrieben, und die Normierung der Posteriori-Dichte $f(\mu | \bar{x})$ wird durch $f(\bar{x}) = \int f(\bar{x} | \mu) \cdot f_0(\mu) d\mu$ im Nenner sichergestellt, also dem Integral über Likelihood mal Prior über alle prinzipiell möglichen Werte von μ . Was an dieser Stelle zu beachten ist: In Glg. (68) ist stets μ der variable Parameter, \bar{x} ist

konstant, nämlich genau der Wert, der aus der Messung resultierte. Anders als in der konventionellen Statistik, die neben dem gemessenen Wert auch alle anderen Werte \bar{x} einbezieht, die eine Messung prinzipiell hätte liefern können (die aber gar nicht gemessen wurden) und den „wahren Wert“ μ als (unbekannte) Konstante betrachtet, ist in Glg. (68) der Parameter μ verteilt. Über ihn wird letztlich eine Wahrscheinlichkeitsaussage getroffen.

Mit der Posteriori-Dichte $f(\mu | \bar{x})$ lassen sich nun auch Hypothesentests durchführen, ganz analog zur Art und Weise, wie sie in den vorangegangenen Abschnitten dargestellt wurden. Nur dass ein Hypothesentest sich diesmal nicht auf die Verteilung der möglichen \bar{x} bei gegebenem μ bezieht, sondern auf die Posteriori-Dichte der verteilten μ bei vorliegendem \bar{x} . Diese ist nun viel intuitiver zu interpretieren: Jetzt kann z. B. tatsächlich die Wahrscheinlichkeit dafür angegeben werden, dass der Wert μ in einem vorgefundenen Intervall liegt:

$$P(\bar{x} - \sigma \leq \mu \leq \bar{x} + \sigma | \bar{x}) = \int_{\bar{x} - \sigma}^{\bar{x} + \sigma} f(\mu | \bar{x}) d\mu \quad (69)$$

Ein Vertrauensintervall, das sich gemäß Glg. (69) auf eine solche Posteriori-Verteilung bezieht, wird (*Bayes'sches*) *Kredibilitätsintervall* genannt.

Zusammengefasst lässt sich sagen: In der konventionellen Statistik ist μ der konstante, wenn auch unbekannte, „wahre Wert“ der Messgröße und die Verteilung aller im Experiment möglichen Ergebnisse \bar{x} wird betrachtet. Überspitzt gesagt, bezieht sich die Analyse also auch auf Messdaten, die gar nicht vorliegen. Der Bayes'sche Ansatz vermeidet dies. Hier wird dem Parameter μ eine Verteilung zugeordnet und \bar{x} ist der konstante Wert des Messergebnisses. Die Posteriori-Dichte $f(\mu | \bar{x})$ stellt dann die gesamte Information dar, die uns über die Messgröße (den Parameter μ) zur Verfügung steht.

Wie nun praktisch ein solcher Hypothesentest im Rahmen der Bayes'schen Statistik durchgeführt wird, soll an einem abschließenden Beispiel erläutert werden.

Beispiel: Das Stromnormal – Fortsetzung

Zur Illustration des Verfahrens greifen wir das einfachste, bereits in Abschnitt 4.1.1 auf S. 32 im Kontext der konventionellen Statistik diskutierte Beispiel erneut auf: Wir überprüfen ein Stromnormal, das laut Herstellerangabe einen Strom von $15 \mu\text{A}$ liefern soll. Wir führen dazu eine Messreihe von $N = 50$ Strommessungen durch und erhalten aus dem Datensatz einen mittleren Strom von $\bar{I} = 14.85 \mu\text{A}$ und eine empirische Standardabweichung von $s = 0.3 \mu\text{A}$.

Wie schon vorher testen wir die Nullhypothese, nach der das Stromnormal wie angegeben einen Strom von $I_0 = 15 \mu\text{A}$ liefert. Grundsätzlich können wir nun aus der Bayes'schen Perspektive dem Parameter I_0 eine Wahrscheinlichkeit zuweisen. Da der Strom jedoch einen kontinuierlichen Parameter darstellt, hat jeder einzelne Wert die Wahrscheinlichkeit Null. Deshalb benutzen wir das Kredibilitätsintervall, das wir auf der Basis der Posteriori-Dichte $f(\mu | \bar{x})$ nach Glg. (68) ermitteln. Letztere müssen wir zunächst bestimmen. Dann wählen wir ein Kredibilitätsintervall von, sagen wir, 99% (entsprechend der Wahl von $\alpha = 1\%$ in Abschnitt 4.1.1). Liegt dann der Wert der Nullhypothese $15 \mu\text{A}$ außerhalb des Intervalls, lehnen wir die Hypothese ab.

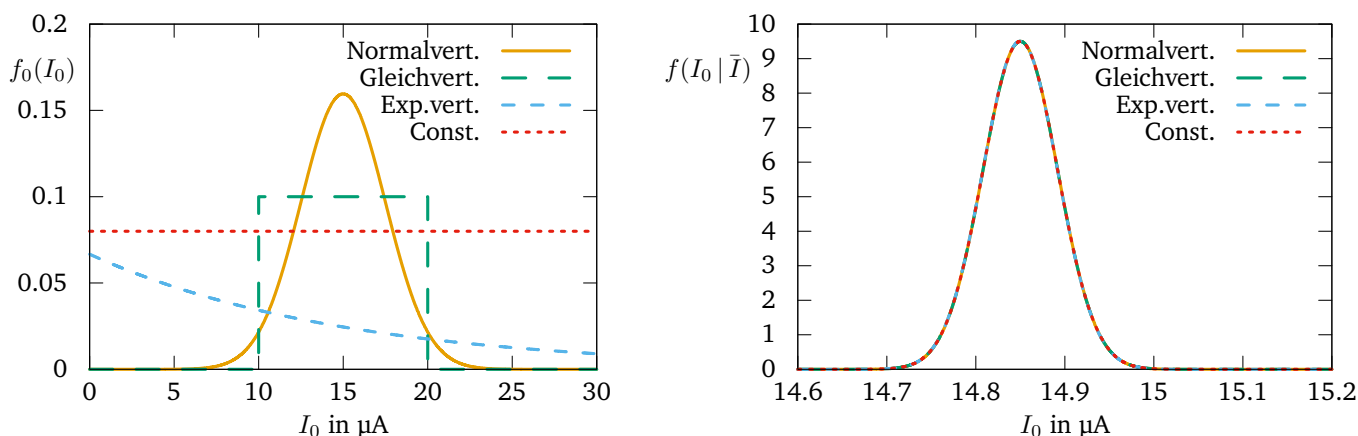


Abbildung 11: Unterschiedliche Prior-Verteilungen, die für ein Stromnormal, das nominell $15 \mu\text{A}$ liefern soll, angenommen werden könnten, wenn keine weitere Information vorliegt (links) sowie die daraus zusammen mit der Likelihood-Funktion resultierenden Posteriori-Verteilungen $f(I_0 | \bar{I})$ (rechts).

Um zur Posteriori-Dichte zu gelangen, sind zunächst die Priori-Dichte und die Likelihood-Funktion zu bestimmen. Die Likelihood ist die Wahrscheinlichkeit, den aus $N = 50$ Messungen bestimmten Mittelwert der Stromstärke \bar{I} zu messen unter der Bedingung, dass die Verteilung den Mittelwert I_0 aufweist. Da wir für \bar{I} eine Normalverteilung annehmen dürfen, deren Standardabweichung wir mit s/\sqrt{N} abschätzen,⁴ erhalten wir:

$$f(\bar{I} | I_0) = N(I_0, s/\sqrt{N}) = \frac{\sqrt{N}}{\sqrt{2\pi}s} e^{-\frac{N(\bar{I}-I_0)^2}{2s^2}} \quad (70)$$

Anders als in der frequentistischen Herangehensweise, bei der der Parameter I_0 unbekannt aber konstant war, und man die Verteilung möglicher Messwerte \bar{I} betrachtet hat, ist in dieser Betrachtungsweise \bar{I} eine Konstante, nämlich der gemessene Mittelwert, während der Parameter I_0 variiert.

Nun müssen wir noch eine Priori-Verteilung wählen. Hier ergeben sich u. U. mehrere Möglichkeiten, jedoch gilt immer, dass die Priori-Verteilung vollständig unabhängig von den vorliegenden Messdaten gewählt werden muss: Wenn wir z. B. den Standpunkt einnehmen, dass wir vor der Messung über keinerlei Information über die Messgröße verfügen, wählen wir eine *nicht-informative* Priori-Verteilung. Für den vorliegenden Fall wäre das eine Konstante, z. B. $f_0(I_0) = 1$. Problem an dieser „uneigentlichen“ Verteilung ist: sie ist nicht normierbar. Dies ist aber grundsätzlich zu verkräften, solange das Produkt, das im Zähler und Nenner von Glg. (68) auftaucht, normierbar bleibt.

Alternativ könnten wir auch eine moderat informative Verteilung als Prior ansetzen: Zum Beispiel eine Gleichverteilung auf einem festgelegten Intervall mit einem Mittelwert von $I_0 = 15 \mu\text{A}$. Wenn wir davon ausgehen, dass das Normal auf jeden Fall einen Strom im Bereich von $I_{\min} \dots I_{\max}$, also

⁴In einer vollen Bayes'schen Analyse würde man auch die Standardabweichung als verteilten Parameter s_0 ansetzen, also $f(\bar{I} | I_0, s_0)$ bestimmen und dann diesen Parameter *marginalisieren*: $f(\bar{I} | I_0) = \int f(\bar{I} | I_0, s_0) ds_0$. Die volle Rechnung würde den Rahmen dieser Darstellung jedoch sprengen.

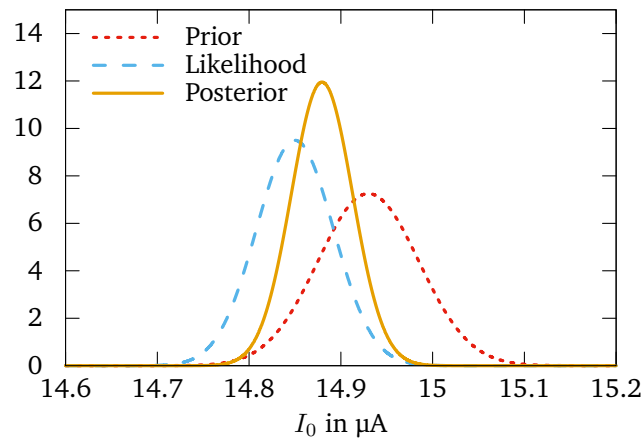


Abbildung 12: Die gleiche Analyse wie in Abb. 11, nur mit einem Prior, der sich aus einer unabhängigen Vormessung ergeben hat. Bei gleicher Likelihood-Funktion ist aufgrund des hohen Informationsgehaltes des Priors der Effekt auf die Posteriori-Verteilung sehr deutlich.

z. B. $10 \dots 20 \mu\text{A}$ liefert, dann ergibt sich:

$$f_0^{\text{gleich}}(I_0) = \begin{cases} \frac{1}{I_{\max} - I_{\min}} & \text{falls } I_{\min} \leq I_0 \leq I_{\max} \\ 0 & \text{sonst} \end{cases} \quad (71)$$

Eine weitere Möglichkeit wäre eine Normalverteilung mit dem Mittelwert $15 \mu\text{A}$ und einer ausreichend großen Standardabweichung: $f_0(I_0) = N(15 \mu\text{A}, 5 \mu\text{A})$. Oder aber wir sehen einen Mittelwert von $15 \mu\text{A}$ als gegeben an sowie die Randbedingung $I_0 > 0$, dann ergibt sich nach Abschnitt 3.3 eine Exponentialverteilung. Abb. 11(links) zeigt die sehr unterschiedlichen Priori-Verteilungen, die so aus verschiedenen Vorüberlegungen erwachsen könnten.

Nun ist lediglich noch das Integral im Nenner von Glg. (68) numerisch zu berechnen und die resultierenden Posteriori-Verteilungen $f(I_0 | \bar{I})$ können angegeben werden. Abb. 11(rechts) zeigt das Ergebnis: Im dem Falle, dass kaum Information in der Priori-Verteilung steckt, wie in unserem einfachen Beispiel, ist die Posteriori-Verteilung praktisch unabhängig von der speziellen Wahl des Priors $f_0(I_0)$ und das Ergebnis ist identisch, mit dem, das wir bereits vorher im Rahmen der frequentistischen Statistik erhalten hatten: Das 99% Kreditabilitätsintervall liegt für $f(I_0 | \bar{I})$ bei $(14.85 \pm 0.11) \mu\text{A}$. Der angegebene Wert von $15 \mu\text{A}$ liegt deutlich außerhalb und so lehnen wir die Nullhypothese ab.

Die eigentliche Stärke des Bayes'schen Ansatzes zeigt sich also erst dann, wenn substantielle *a priori* Information vorliegt. Diese führt dann zu einer deutlich schmaleren Priori-Verteilung, als die geringfügig oder gar nicht informativen Verteilungen, die wir bisher diskutiert haben. Nehmen wir z. B. an, wir hätten belastbare Information aus einer Vormessung am selben Stromnormal vorliegen, die uns einen Mittelwert $\bar{I}_2 = 14.93 \mu\text{A}$ mit einer Standardabweichung von $s_2/\sqrt{N_2} = 0.055 \mu\text{A}$ liefert. Dann könnten wir die aus dieser Messung gewonnene Posteriori-Verteilung als Prior für die obige Analyse verwenden. Die Likelihood-Funktion wäre die gleiche wie vorher und identisch mit den in Abb. 11(rechts) gezeigten Posteriori-Verteilungen. Die neue Posteriori-Verteilung weicht nun aber

deutlich sowohl vom Prior als auch von der Likelihood ab, wie in Abb. 12 gezeigt. Der Mittelwert der neuen Posteriori-Verteilung ist nun $\bar{I}_0 = 14.88 \mu\text{A}$ und das 99%-Kredibilitätsintervall ist gegeben durch $(14.88 \pm 0.086) \mu\text{A}$. Auch hier werden wir also die Nullhypothese ablehnen, aufgrund der zusätzlichen Information fällt die Entscheidung jedoch deutlich knapper aus.

Zusammengefasst lässt sich sagen: Die Stärke des Bayes'schen Ansatzes liegt darin, dass grundsätzlich *alle* verfügbare Information, nicht nur die statistische, in die Datenanalyse einfließen und dort auf einheitliche Weise behandelt werden kann. Aus diesem Grunde sind Bayes'sche Verfahren oft leistungsfähiger als die der konventionellen Statistik und deshalb setzen sie sich in verschiedenen Bereichen zunehmend durch und bilden mittlerweile bereits die Grundlage für international einheitliche Vorgaben zur Bestimmung und Angabe von Messunsicherheiten. Andererseits ist, wie am letzten Beispiel zu erkennen war, auch in einfachen Fällen der Rechenaufwand deutlich größer, so dass sich häufig eine computergestützte Datenanalyse anbietet. Außerdem wird kontrovers diskutiert, in wie weit ein gewisses Maß an Subjektivität, das in der Wahl der geeigneten Prior-Verteilung einfließt, in einer wissenschaftlichen Datenanalyse zulässig ist. Auch wenn die Werkzeuge der Bayes'schen Statistik in verschiedenen Bereichen so erfolgreich eingesetzt werden, dass dort schon von einer „Bayes'schen Revolution“ in der Datenanalyse die Rede ist [9], ist die damit verknüpfte erkenntnistheoretische Diskussion bislang noch keineswegs abgeschlossen.

Literatur

- [1] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, *Guide to the Expression of Uncertainty in Measurement*, Joint Committee for Guides in Metrology, JCGM 100, 2008; online unter: <http://www.bipm.org/en/publications/guides/gum.html>
- [2] M. Grabe, *Technisches Messen*, **67**, 6 (2000); auch online unter www.uncertainty.de.
- [3] K. Wiese, W. Wöger, *Messunsicherheit und Messdatenauswertung*, Wiley, New York, 1999.
- [4] T. Bornath, G. Walter, *Messunsicherheiten – Grundlagen*, Springer Spektrum, Wiesbaden, 2020; auch als EBook!
- [5] T. Bornath, G. Walter, *Messunsicherheiten – Anwendungen*, Springer Spektrum, Wiesbaden, 2020; auch als EBook!
- [6] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, *Evaluation of Measurement Data: Supplement 1 to the “Guide to the Expression of Uncertainty in Measurement”*, Joint Committee for Guides in Metrology, JCGM 101, 2008.
- [7] J. Lehn, H. Wegmann, *Einführung in die Statistik*, Teubner, Stuttgart, 2000.
- [8] U. Krenzel, *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 8. Auflage, Vieweg, Wiesbaden, 2005.
- [9] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press, Cambridge, 2005.
- [10] W. M. Bolstad, *Introduction to Bayesian Statistics*, Wiley, Hoboken N.J., 2004.
- [11] O. Passon, *Bayes’sche Statistik für Fußgänger*, online unter: <http://psiquadrat.de/downloads/bayes.pdf>.
- [12] N. Gerschenfeld, *The Nature of Mathematical Modelling*, Cambridge University Press, Cambridge, 1999.
- [13] W. Kessel, PTB-Broschüre, online unter: <http://www.ptb.de/de/publikationen/download/dl00001.html>
- [14] J. C. Earnshaw, *Eur. J. Phys.*, **11**, 338-342 (1990)
- [15] A. Leike, *Eur. J. Phys.*, **23**, 21-26 (2002)
- [16] H.-O. Georgii, *Stochastik*, 3. Auflage, de Gruyter, Berlin 2007.
- [17] W. H. Press, S. A. Teukolsky, W. T. Vetterling und B. P. Flannery, *Numerical Recipes – The Art of Scientific Computing*, 3. Auflage, Cambridge University Press, Cambridge, 2007. Auch online unter: <http://www.nr.com>.

A. Die p -Quantile der χ^2 -Verteilung

Angegeben sind die p -Quantile χ_p^2 der χ^2 -Verteilung für k Freiheitsgrade. Es gilt dabei:

$$p = P(\chi^2 \leq \chi_p^2) = \int_0^{\chi_p^2} f_{\chi^2}(z, k) dz.$$

k	$p = 0.99$	$p = 0.975$	$p = 0.95$	$p = 0.90$	$p = 0.80$	$p = 0.70$	$p = 0.50$	$p = 0.30$	$p = 0.20$	$p = 0.10$	$p = 0.05$
1	6.63	5.02	3.84	2.71	1.64	1.07	0.455	0.148	0.064	0.016	0.0039
2	9.21	7.38	5.99	4.61	3.22	2.41	1.386	0.713	0.446	0.211	0.103
3	11.3	9.35	7.81	6.25	6.64	3.67	2.366	1.424	1.005	0.584	0.352
4	13.3	11.1	9.49	7.78	6.0	4.9	3.36	2.19	1.65	1.06	0.71
5	15.1	12.8	12.1	9.24	7.3	6.1	4.35	3.00	2.34	1.61	1.14
6	16.8	14.4	12.6	10.6	8.6	7.2	5.35	3.83	3.07	2.20	1.63
7	18.5	16.0	14.1	12.0	9.8	8.4	6.35	4.67	3.82	2.83	2.17
8	20.1	17.5	15.5	13.4	11.0	9.5	7.34	5.53	4.59	3.49	2.73
9	21.7	19.0	16.9	14.7	12.2	10.7	8.34	6.39	5.38	4.17	3.32
10	23.2	20.5	18.3	16.0	13.4	11.8	9.34	7.27	6.18	4.86	3.94
11	24.7	21.9	19.7	17.3	14.6	12.9	10.3	8.1	7.0	5.6	4.6
12	26.2	23.3	21.0	18.5	15.8	14.0	11.3	9.0	7.8	6.3	5.2
13	27.7	24.7	22.4	19.8	17.0	15.1	12.3	9.9	8.6	7.0	5.9
14	29.1	26.1	23.7	21.1	18.2	16.2	13.3	10.8	9.5	7.8	6.6
15	30.6	27.5	25.0	22.3	19.3	17.3	14.3	11.7	10.3	8.5	7.3
16	32.0	28.8	26.3	23.5	20.5	18.4	15.3	12.6	11.2	9.3	8.0
17	33.4	30.2	27.6	24.8	21.6	19.5	16.3	13.5	12.0	10.1	8.7
18	34.8	31.5	28.9	26.0	22.8	20.6	17.3	14.4	12.9	10.9	9.4
19	36.2	32.9	30.1	27.2	23.9	21.7	18.3	15.4	13.7	11.7	10.1
20	37.6	34.2	31.4	28.4	25.0	22.8	19.3	16.3	14.6	12.4	10.9
21	38.9	35.5	32.7	29.6	26.2	23.9	20.3	17.2	15.4	13.2	11.6
22	40.3	36.8	33.9	30.8	27.3	24.9	21.3	18.1	16.3	14.0	12.3
23	41.6	38.1	35.2	32.0	28.4	26.0	22.3	19.0	17.2	14.8	13.1
24	43.0	39.4	36.4	33.2	29.6	27.1	23.3	19.9	18.1	15.7	13.8
25	44.3	40.6	37.7	34.4	30.7	28.2	24.3	20.9	18.9	16.5	14.6
26	45.6	41.9	38.9	35.6	31.8	29.2	25.3	21.9	19.8	17.3	15.4
27	47.0	43.2	40.1	36.7	32.9	30.3	26.3	22.7	20.7	18.1	16.2
28	48.3	44.5	41.3	37.9	34.0	31.4	27.3	23.6	21.6	18.9	16.9
29	49.6	45.7	42.6	39.1	35.1	32.5	28.3	24.6	22.5	19.8	17.7
30	50.9	47.0	43.8	40.3	36.3	33.5	29.3	25.5	23.4	20.6	18.5