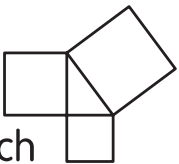


Josef Naas
Wolfgang Tutschke

Große Sätze und schöne Beweise der Mathematik

Identität des Schönen, Allgemeinen, Anwendbaren

Verlag
Harri
Deutsch



Dr. Josef Naas (†) war Professor für Mathematik an der Akademie der Wissenschaften in Berlin.

Dr. Wolfgang Tutschke ist emeritierter Professor für Mathematik an der Technischen Universität Graz.

Wissenschaftlicher Verlag Harri Deutsch GmbH
Gräfstraße 47
60486 Frankfurt am Main
verlag@harri-deutsch.de
www.harri-deutsch.de

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-8171-1822-9

Dieses Werk ist urheberrechtlich geschützt.

Alle Rechte, auch die der Übersetzung, des Nachdrucks und der Vervielfältigung des Buches – oder von Teilen daraus – sind vorbehalten. Kein Teil des Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder ein anderes Verfahren), auch nicht für Zwecke der Unterrichtsgestaltung, reproduziert oder unter Verwendung elektronischer Systeme verarbeitet werden.

Zu widerhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Der Inhalt des Werkes wurde sorgfältig erarbeitet. Dennoch übernehmen Autor und Verlag für die Richtigkeit von Angaben, Hinweisen und Ratschlägen sowie für eventuelle Druckfehler keine Haftung.

3., korrigierte und erweiterte Auflage 2009

©Wissenschaftlicher Verlag Harri Deutsch GmbH, Frankfurt am Main, 2009

Druck: betz-druck GmbH, Darmstadt <www.betz-druck.de>

Printed in Germany

Vorwort zur 3., korrigierten und erweiterten Auflage

Seit der ersten Auflage der *Großen Sätze und schönen Beweise der Mathematik* sind mehr als zwanzig Jahre vergangen, eine Zeitspanne, in der sich auch in der Mathematik vielerlei neue Entwicklungen ergeben haben. Insbesondere die Nutzung elektronischer Hilfsmittel hat auch die Mathematik zu einem noch bedeutsameren Wirtschaftsfaktor werden lassen. Diese Entwicklung hat neben anderen Auswirkungen auch dafür erforderliche Spezialisierungen in der Mathematik noch verstärkt. Damit ist das im Vorwort der ersten Auflage erwähnte Anliegen auch heute noch unverändert aktuell: Um die Wirksamkeit einer Grundlagenwissenschaft wie der Mathematik zu erhöhen, soll versucht werden, prinzipielle mathematische Denkweisen einem größeren Leserkreis aus allen Wissensbereichen zugänglich zu machen.

Das Buch wendet sich daher insbesondere auch an Leser (wie Naturwissenschaftler und Ingenieure), die spezielle mathematische Methoden in ihrem Beruf anwenden, und die darüber hinaus erfahren wollen, was die Mathematik sonst noch so zu bieten hat. Wer ein volles Mathematik-Studium absolviert, lernt natürlich ein relativ breites Spektrum mathematischer Denkweisen kennen, wenngleich die jetzige Aufspaltung des Diplom-Studiums in ein Bachelor- und ein Master-Studium die Tendenz einer frühzeitigen Spezialisierung ebenfalls verstärkt. Und so ist es nicht unwahrscheinlich, dass auch ein ausgebildeter Mathematiker den einen oder anderen für ihn neuen Gesichtspunkt findet.

Daher bin ich dem Verlag HARRI DEUTSCH dankbar, eine Neuauflage geplant zu haben, nachdem die bisherigen Auflagen und Nachdrucke im Akademie-Verlag Berlin und Verlag HARRI DEUTSCH vergriffen waren. Während das ursprüngliche Manuskript in der früher üblichen Weise hergestellt worden war (handschriftliches Manuskript, dann Abschrift mit Schreibmaschine und Eintragen der Formeln, händischer Drucksatz mit Fahnen- und Bogenkorrektur), hatte sich der Verlag HARRI DEUTSCH freundlicherweise zu einer Digitalisierung des Manuskriptes entschlossen. Diese digitalisierte Fassung wurde dann bei der Vorbereitung der hiermit vorliegenden dritten Auflage zur Korrektur und Erweiterung verwendet.

Mit Ausnahme von Kapitel 13 wurden alle bisherigen Beweisanordnungen beibehalten. Allerdings wird der Beweis des Satzes von CAUCHY-KOWALEWSKAJA jetzt anstelle durch sukzessive Approximationen durch das elegantere Kontraktions-Prinzip durchgeführt. In die vorliegende Neuauflage wurde daher (am Ende von Kapitel 9) auch ein Beweis des Kontraktions-Prinzips aufgenommen. In Kapitel 14 wird das WEYLsche Lemma nicht nur auf die bisherige Weise (über die Mittelwerteigenschaft) bewiesen, sondern es wird ein weiterer Beweis hinzugefügt, der auf der Approximation von stetigen Funktionen durch zweimal stetig differenzierbare beruht. Weitere Ergänzungen

betreffen beispielsweise einen Ausblick auf den Fixpunktsatz von LERAY-SCHAUDER sowie einen Überblick über das Lösen (nichtlinearer) impliziter Gleichungen. Teilweise sollen diese Ergänzungen auch an Methoden herañführen, die von der gegenwärtigen Forschung benutzt werden. Natürlich betrifft das in erster Linie Themen, die zu meinen Forschungs-Interessen gehören.

Beim Aufbau des Buches wurde versucht, die einzelnen Kapitel mehr oder weniger unabhängig voneinander zu gestalten. Rück- und Querverweise, wie sie in einem systematischen Lehrbuch unumgänglich sind, wurden weitestgehend vermieden. Im Gegenteil, um eine größtmögliche Selbständigkeit der Kapitel zu erreichen, wurden grundlegende Konstruktionen (wie das CANTORSche Diagonalverfahren in den Kapiteln 3, 5 und 11) sogar mehrfach erläutert. Andererseits gibt es aber auch Fälle, wo die Aussage eines Kapitels in einem anderen Kapitel gebraucht wird. Das ist beispielsweise beim Beweis des SCHAUDERSchen Fixpunktsatzes der Fall (Kapitel 11), da dieser auf den BROUWERSchen Fixpunktsatz (Kapitel 10) zurückgeführt werden kann.

Im Gegensatz zu den früheren Ausgaben werden alle Literaturzitate jetzt am Ende des Buches gesammelt. Natürlich ist es bei der großen Breite der behandelten Themen auch nicht annähernd möglich, einen vollständigen Überblick über relevante Literatur zu geben. In erster Linie werden benutzte Quellen dokumentiert (soweit es sich nicht um Kenntnisse handelt, die zum Standard-Wissen eines ausgebildeten Mathematikers gehören). Davon abgesehen, sollen Anregungen zu weiterführenden Literaturstudien gegeben werden.

Mein Dank an den Verlag HARRI DEUTSCH geht insbesondere auch an dessen für elektronische Medien zuständigen Abteilungsleiter, Herrn KLAUS HORN. Bei der Vorbereitung dieser Neuauflage konnte ich mich stets auf seine angenehme und konstruktive Zusammenarbeit verlassen, die weit über das rein Verlegerische hinausging: Seine Sachkenntnisse als Physiker haben zu mancher Verbesserung des Textes geführt. Auch das Satzbüro Tanovski & Partners möchte ich dankend erwähnen. Die digitale Fassung wurde sorgfältig erstellt, und sogar der eine oder andere Schreibfehler der Vorlage erkannt und korrigiert.

So bleibt mir nur zu wünschen, dass die vorliegende Neuauflage einen Beitrag zur Pflege mathematischer Denkweisen leistet.

Graz, im Januar 2009
WOLFGANG TUTSCHKE

Vorwort

In allen Wissenschaften ist die Tendenz einer zunehmenden Spezialisierung zu beobachten; so auch in der Mathematik. Diese zunehmende Spezialisierung, die mit immer breiter gefächerten Anwendungen verbunden ist, führt zu der Gefahr, dass die Mathematik in der Praxis nicht mehr als einheitliche Wissenschaft in Erscheinung tritt. Solche Tendenzen hat es auch in früheren Entwicklungsphasen der Mathematik gegeben; sie wurden stets von Versuchen begleitet, einer Zersplitterung entgegenzuwirken und die Einheit der Mathematik zur Geltung zu bringen. Beispiele dafür sind die »Elemente« von EUKLID oder die Algebraisierung der Geometrie durch DESCARTES. Der Nutzen, den eine solche auf die Herstellung der Einheit der Mathematik gerichtete Tätigkeit hat, ist ganz offensichtlich, vor allem schon auf Grund unserer historischen Erfahrungen.

In der heutigen Mathematik zeigt sich die erfreuliche Tatsache, dass sie außerordentlich reichhaltig geworden ist, dass zu vielen alten Anwendungsgebieten immer noch neue hinzugetreten sind. Daher ist es heute sogar wichtiger denn je, die Einheit der Mathematik zu fördern. Dazu ist es erforderlich, das Wesen der Mathematik, das besondere in ihren Beweismethoden zutage tritt, sichtbar zu machen. Der Grund für eine solch große Bedeutung mathematischer Beweismethoden liegt darin, dass sie nicht nur auf speziellen Teilgebieten genutzt werden, sondern allgemeinere Verwendung finden. In diesem Sinne ist ein wichtiger Schritt zum Verständnis des Wesens der Mathematik das Kennenlernen der Beweise für ihre fundamentalen Sätze. In dem vorliegenden Buch sind 15 solcher Sätze aus tragenden Teilgebieten der heutigen Mathematik ausgewählt worden. Dazu gehören Sätze über die Lösbarkeit von Gleichungen, wie die FREDHOLMSche Alternative oder der BROUWERSche und der SCHAUDERSche Fixpunktsatz, die sich vielseitig auswirken und deren Beweise gleichzeitig prinzipielle Bedeutung haben.

Derartige übergreifende mathematische Gesichtspunkte kommen auch beim Beweis von Aussagen zum Tragen, die ganz spezielle Anwendungen betreffen; ein Beispiel hierfür ist das EINSTEINSche Additionstheorem für Geschwindigkeiten, dessen tieferer Sinn erst durch den mathematischen Gruppenbegriff erkannt werden kann. Die mathematische Praxis führt zwangsläufig aber auch zu beweistheoretischen Fragen, wie sie in den Unterschieden zwischen Wahrheit und Beweisbarkeit in der Mathematik zum Ausdruck kommen. Desgleichen treten – auch bei konkreten analytischen Untersuchungen – mengentheoretische Problemstellungen auf, wie sie sich beispielsweise in der Äquivalenz von Auswahlaxiom und ZORNschem Lemma äußern. Die moderne Entwicklung der Mathematik erforderte auch, den klassischen Differenzierbarkeitsbegriff und im Zusammenhang damit auch den Lösungsbegriff für Differentialgleichungen zu erweitern. Damit entsteht natürlich auch die Frage, wann eine verallgemeinerte Lösung

doch klassische Lösung ist. Eine Antwort hierauf gibt beispielsweise das WEYLSche Lemma, dem Kapitel 14 dieses Buches gewidmet ist.

In den 15 Kapiteln des Buches, die weitestgehend selbständigen Charakter tragen, werden zentrale mathematische Problemstellungen aus verschiedenen Zweigen der Mathematik und ihrer Anwendungen behandelt. Jeder Abschnitt ist einem fundamentalen mathematischen Sachverhalt gewidmet, wobei die Allgemeingültigkeit der gewonnenen Ergebnisse eingeschätzt wird. Außerdem enthält das Buch einen Dialog über Mathematik, der Bemerkungen über das Wesen der heutigen Mathematik und daraus resultierende Konsequenzen zum Gegenstand hat. Darin sind insbesondere auch Hinweise auf die Identität des Schönen, Allgemeinen und Anwendbaren in der Mathematik eingeschlossen.

Zu den im Buchtitel auftretenden Begriffen »groß« und »schön« sei eine kurze Bemerkung gemacht. Ihrem Ursprung und ihrer hauptsächlichlichen Verwendung in der Umgangssprache entsprechend haben beide zunächst einen stark subjektiven Charakter. Andererseits werden im wissenschaftlichen Sprachgebrauch beide Begriffe aber durch die Wissenschaftsentwicklung selbst objektiviert, wobei als wirklich große Sätze der Mathematik diejenigen hervortreten, die allgemein sind und sich demzufolge auch vielfach anwenden lassen. Wirklich zentrale Sätze werden durch die Wissenschaftsentwicklung auch stets so tiefgreifend durchgearbeitet, dass ihre Beweise klar und durchsichtig und damit auch im Sinne des menschlichen Strebens nach Wahrheitsfindung natürlich und schön werden. Die inner- und außermathematische Ausstrahlungskraft der hier zu behandelnden Sätze ist sehr verschiedenartig. In diesem Sinne sind die ausgewählten Sätze keinesfalls gleichermaßen groß. Hier aber noch quantitative Abstufungen innerhalb des qualitativen Begriffes »groß« vornehmen zu wollen, hieße aber doch – zumindest im Rahmen unserer Zielstellungen – zu weit zu gehen. Das gleiche trifft auf die »Schönheit« der Beweise zu.

Die vorrangige Zielstellung des Buches ist es, Wesen und Bedeutung der Mathematik für einen breiten Leserkreis erkennbar zu machen und so das weitere Eindringen der Mathematik in die Praxis zu fördern.

Dieses Buch wendet sich an alle, die an der heutigen Mathematik interessiert sind. Die heute unvermeidbare Spezialisierung führt dazu, vornehmlich mit Teilgebieten unserer Wissenschaft bekannt zu sein. Das bewirkt oft, dass ein Überblick über die Möglichkeiten, die in der Mathematik insgesamt enthalten sind, fehlt.

Hierfür nun sollte die Lektüre des Buches nützlich sein und das Verständnis für Wesen und Anwendbarkeit der Mathematik vertiefen.

An vielen Schulen existieren derzeit Schülerarbeitsgemeinschaften, die unter der Anleitung erfahrener Mathematiklehrer die Begeisterung junger Talente für unser Fachgebiet wecken; dafür werden Materialien benötigt, die sich an prinzipiellen Entwicklungstendenzen der Mathematik orientieren; das vorliegende Buch enthält geeignete Beiträge dieser Art.

In den verschiedensten Berufsgruppen sind Kenntnisse klassischer mathematischer Methoden vorhanden, die beispielsweise in Studienrichtungen an Hoch- und Fachschulen mit Mathematik als Nebenfach unterrichtet werden. Da diese Methoden in der neueren Mathematik vielfach verbessert worden sind, ist es zweckmäßig, für die Verbreitung solcher Neuentwicklungen tätig zu sein. Deshalb behandelt das Buch auch klassische Fragestellungen mit modernen mathematischen Methoden, verdeutlicht deren Vorzüge, gibt dem Leser mancherlei Anregung für die mathematische Weiterbildung, um so geeignete Wege für eine breitere Nutzung mathematischer Methoden in der Praxis zu erschließen.

Einer der Autoren hat in Gestalt des Mathematischen Wörterbuches vor mehr als zwei Jahrzehnten ein umfangreiches Werk veröffentlicht, das eine Antwort auf die Frage gab, was Mathematik sei. Das jetzt vorliegende Buch strebt eine Antwort darauf an, wie Mathematik wird, wie sie neue Gestalt annimmt. Die vielen Auflagen des Mathematischen Wörterbuches sind ein deutlicher Hinweis auf seinen Nutzen. Die Autoren des vorliegenden Buches sind überzeugt davon, dass auch von ihm eine entsprechende Wirkung erwartet werden kann, weil das Interesse für die ständige Neugestaltung der Mathematik aktuell ist und bleibt.

Schließlich noch eine Bemerkung zum Lesen des Buches. Die einzelnen Kapitel sind in einer Weise voneinander unabhängig, dass sie in beliebiger Reihenfolge gelesen werden können. In manchen Kapiteln finden sich zwar gelegentliche Hinweise auf andere Kapitel. Diese sollen den interessierten Leser aber im allgemeinen nur auf Querverbindungen aufmerksam machen; sie sind für das Verständnis des jeweiligen Kapitels nicht erforderlich. Manche Definitionen und Schlussweisen werden sogar in mehreren Kapiteln wiederholt, falls sie für das Verständnis des jeweiligen Kapitels wesentlich sind. Der Aufbau der einzelnen Kapitel erfolgt im allgemeinen so, dass sie ohne Benutzung weitergehender Literatur verständlich sind.

Für Hinweise auf die Gestaltung einzelner Kapitel sprechen wir folgenden Fachkollegen, deren Ratschläge besonders hilfreich waren, unseren Dank aus: Herrn Dr. sc. G. DAUTCOURT vom Zentralinstitut für Astrophysik der Akademie der Wissenschaften der DDR für Hinweise zu Kapitel 4, Herrn Prof. G. F. MANDŽAVIDZE vom I. N. VEKUA-Institut für angewandte Mathematik der Universität Tbilissi zu Kapitel 9 und 10, Herrn Dr. sc. H. HERRE vom KARL-WEIERSTRASS-Institut für Mathematik der Akademie der Wissenschaften der DDR zu Kapitel 15.¹ Vorschläge hinsichtlich der Gestaltung des Gesamtmanuskriptes verdanken die Autoren Herrn Prof. P. H. MÜLLER von der Sektion Mathematik der Technischen Universität Dresden sowie Herrn Prof. S. PRÖSSDORF vom bereits oben erwähnten KARL-WEIERSTRASS-Institut für Mathematik. Den Mitarbeitern des Akademie-Verlages wird für

¹ Im vorliegenden Nachdruck wurde nur eine Beweislücke in Kapitel 7 beseitigt, auf die uns dankenswerterweise Herr Dr. A. POMP vom KARL-WEIERSTRASS-Institut für Mathematik der Akademie der Wissenschaften der DDR aufmerksam gemacht hat.

eine vielseitige, konstruktive Zusammenarbeit gedankt, insbesondere Fräulein Dipl.-Math. G. REIHER und Herrn Dr. R. HÖPPNER.

Berlin und Halle, im Juni 1985

JOSEF NAAS und WOLFGANG TUTSCHKE

Inhaltsverzeichnis

Vorwort zur 3., korrigierten und erweiterten Auflage	i
Vorwort	iii
1 Kleinste Quadrate	1
2 Dynamische Optimierung	11
3 Stochastische Prozesse	15
4 Einsteins Additionstheorem	25
5 Satz von Arzelà-Ascoli	37
6 Der Approximationssatz von Weierstraß und Stone	41
7 Auswahlaxiom	53
8 Satz von Hahn und Banach	61
9 Fredholmsche Alternative	67
10 Fixpunktsatz von Brouwer	99
11 Fixpunktsatz von Schauder	111
12 Monotone Operatoren	133
13 Anfangswertprobleme	145
14 Weylsches Lemma	161
15 Wahrheit und Beweisbarkeit	181
16 Ein Dialog über Mathematik	191
Index	201

1 Die Methode der kleinsten Quadrate

Bei der Bestimmung von Längen, Winkeln, Stromstärken oder anderer Größen aus Messungen sind Messfehler unvermeidlich: Einmal wird das Messgerät im allgemeinen nicht den genauen Wert der gemessenen Größe anzeigen, sondern es wird sich – beispielsweise durch nicht völlig auszuschließende und nicht berechenbare Störeinflüsse – ein davon verschiedener Wert einstellen; zum anderen sind auch Ablesefehler unumgänglich. Mit einem Wort, die gemessenen Werte werden mit einem zufälligen Fehler behaftet sein. Bei einigen Messungen wird sich ein zu kleiner, bei anderen ein zu großer Wert ergeben. Um die gesuchte Größe möglichst genau ermitteln zu können, wird man daher möglichst viele Messungen ausführen und aus ihren Werten dann einen Mittelwert bilden. Die einfachste Möglichkeit, dies zu tun, ist die folgende:

Angenommen, $\lambda_1, \dots, \lambda_n$ seien n unabhängig voneinander gewonnene Messwerte einer einzigen zu bestimmenden Größe λ . Setzt man voraus, dass zu kleine Messwerte ebenso wahrscheinlich wie zu große sind, dann wird der günstigste aus den vorliegenden Messdaten zu gewinnende Wert $\bar{\lambda}$ derjenige sein, für den die Summe der Abweichungen $\bar{\lambda} - \lambda_i$ von den Messwerten λ_i gleich Null ist:

$$\sum_{i=1}^n (\bar{\lambda} - \lambda_i) = 0.$$

Hieraus ergibt sich aber sofort

$$n\bar{\lambda} - \sum_{i=1}^n \lambda_i = 0,$$

so dass $\bar{\lambda}$ das arithmetische Mittel der λ_i ist,

$$\bar{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda_i.$$

Dieser Mittelwert $\bar{\lambda}$ lässt sich aber auch anders charakterisieren, nämlich dadurch, dass die Quadratsumme der Abweichungen von allen λ_i minimal ausfällt. Fordert man nämlich, dass

$$\sum_{i=1}^n (\lambda_i - \hat{\lambda})^2 \tag{*}$$

minimal ausfällt, so muss die Ableitung dieser von $\hat{\lambda}$ abhängenden Funktion gleich Null sein, so dass

$$2 \sum_{i=1}^n (\lambda_i - \hat{\lambda}) \cdot (-1) = 0$$

folgt. Das bedeutet aber, dass auch

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n \lambda_i$$

sein muss. Folglich ist, wie behauptet, $\bar{\lambda} = \hat{\lambda}$. Damit ist – zunächst unter der oben gemachten Voraussetzung über das Verschwinden der Fehlersumme – das folgende Prinzip zur Bestimmung des günstigsten Mittelwertes begründet:

Satz *Der günstigste Mittelwert $\bar{\lambda}$ ist derjenige, für den die Summe der Quadrate der Abweichungen von den (endlich vielen) Messwerten λ_i minimal ausfällt.*

Dieses Prinzip wird als *Methode der kleinsten Quadrate* bezeichnet.

Es ist klar, dass dieses Prinzip nicht immer anwendbar sein kann. Beispielsweise könnte es eine Eigenheit der Messapparatur sein, dass sich zu große Messwerte häufiger ergeben als zu kleine. In einem solchen Fall muss die Annahme, dass die Summe der Abweichungen des gesuchten Wertes von den Messwerten gleich Null sein soll, verworfen werden. Auf Grund solcher Tatsachen könnte man sich zu der Meinung verleiten lassen, dass eine allgemeingültige Theorie zur Ermittlung von Mittelwerten gar nicht möglich ist. Ein solcher Schluss wäre aber zu vorschnell, da die Erfahrung zeigt, dass die Verteilung von Messwerten im allgemeinen bestimmten Gesetzmäßigkeiten unterworfen ist. Eine solche Erfahrung kann zum Beispiel sein, dass unter den vorliegenden Messwerten sehr viele nur wenig von dem wahren Wert der gesuchten Größe abweichen werden, während es vergleichsweise nur sehr wenige Messwerte geben wird, die größere Abweichungen besitzen. Um solche Erfahrungen nicht nur qualitativ, sondern auch quantitativ zu erfassen, wollen wir abzählen, wie viele Messwerte unterhalb einer beliebig gewählten reellen Zahl λ liegen. Bei einer konkreten Messung, bei der n Messungen zur Bestimmung einer einzigen reellen Größe ausgeführt werden, könnte sich beispielsweise folgendes Bild ergeben:

Die erhaltenen Messwerte seien $\lambda_1, \dots, \lambda_n$, wobei wir zunächst voraussetzen wollen, dass diese Messwerte voneinander verschieden sind. Bei geeigneter Nummerierung können wir daher voraussetzen, dass die n Messwerte der Ungleichung

$$\lambda_1 < \lambda_2 < \dots < \lambda_n$$

genügen. Da es keine Messwerte unterhalb λ_1 gibt, kann auch gesagt werden, dass für alle $\lambda \leq \lambda_1$ die Wahrscheinlichkeit dafür, dass der Messwert kleiner als λ ist, gleich

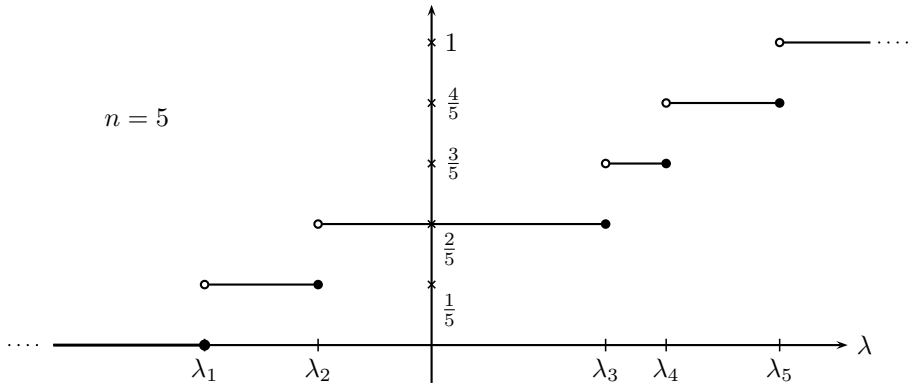


Abbildung 1.1 Empirische Verteilungsfunktion

Null ist.¹ Für alle λ mit $\lambda_1 < \lambda \leq \lambda_2$ gibt es genau einen Messwert, der kleiner als λ ist. Da es insgesamt n (gleichberechtigte) Messwerte gibt, hat jeder die gleiche (klassische) Wahrscheinlichkeit $\frac{1}{n}$. Also ist bei $\lambda_1 < \lambda \leq \lambda_2$ die Wahrscheinlichkeit dafür, dass der Messwert kleiner als λ_2 ist, gleich $\frac{1}{n}$. Ist $\lambda_2 < \lambda \leq \lambda_3$, so liegen die zwei Messwerte λ_1, λ_2 echt unterhalb λ_3 . Ein beliebiger Messwerte liegt daher mit der Wahrscheinlichkeit $\frac{2}{n}$ unterhalb λ . Liegt λ in dem durch $\lambda_i < \lambda \leq \lambda_{i+1}$ charakterisierten Intervall, so liegen also genau i Messwerte unterhalb λ , so dass dann die Wahrscheinlichkeit dafür, dass einer der n erhaltenen Messwerte kleiner als λ_{i+1} ist, durch $\frac{i}{n}$ gegeben wird. Ist schließlich λ größer als der größte gemessene Wert, also $\lambda > \lambda_n$, so liegen alle n gemessenen Werte unterhalb λ , so dass jetzt folgendes gilt: Für $\lambda > \lambda_n$ liegt jeder Messwert mit der Wahrscheinlichkeit 1 unterhalb λ .

Auf die angegebene Art ist jeder reellen Zahl λ eine reelle Zahl $F_n(\lambda)$ zugeordnet worden, die – allgemein gesprochen – folgendermaßen interpretiert werden kann:

$F_n(\lambda)$ ist die Wahrscheinlichkeit dafür, dass bei einer aus n Messungen bestehenden Messreihe einer der möglichen Messwerte kleiner ist als λ .

Die derartig definierte Funktion F_n ist stückweise konstant. Ihre Sprungstellen sind die gemessenen Werte $\lambda_1, \dots, \lambda_n$. An diesen Stellen springt die Funktion F_n um $\frac{1}{n}$. Ihrer Definition gemäß hat die Funktion F_n den in Abb. 1.1 dargestellten Verlauf.

¹ Dieser Aussage liegt die klassische (von LAPLACE stammende) Definition der Wahrscheinlichkeit zugrunde, nach der die Wahrscheinlichkeit der Quotient aus der Anzahl der einem Ereignis günstigen Fälle und der Anzahl aller möglichen Fälle ist. Nach dieser ist die Wahrscheinlichkeit, beispielsweise eine Eins zu würfeln, gleich $\frac{1}{6}$. Ähnlich ist $\frac{2}{6} = \frac{1}{3}$ die Wahrscheinlichkeit, eine Eins oder eine Zwei zu würfeln, weil es in diesem Fall bei insgesamt 6 möglichen Fällen sogar 2 für das beschriebene Ereignis »günstige« Fälle gibt. Sind bei dem im obigen Text beschriebenen Fall die wirklich vorkommenden (gleichberechtigten) Messwerte $\lambda_1, \dots, \lambda_n$, so gibt es insgesamt n mögliche Fälle. Das Ereignis, dass der Messwert kleiner als λ ist, tritt im Falle $\lambda < \lambda_1$ überhaupt nicht ein, es gibt dafür also nur null »günstige« Fälle. Daher kann die entsprechende Wahrscheinlichkeit gleich Null gesetzt werden.

Wir können uns jetzt auch noch von der Voraussetzung befreien, dass die Messwerte alle nur einmal auftreten. Tritt der Messwert λ_i beispielsweise k -fach auf, so besitzt F_n an der Stelle λ_i den Sprung $\frac{k}{n}$. Bei mehrfachem Auftreten von Messwerten fallen sozusagen einige Sprungstellen zusammen, wobei sich gleichzeitig aber die Sprunghöhe entsprechend der Anzahl der zusammenfallenden Messwerte vergrößert.

In allen Fällen – ob die Messwerte nur einfach oder auch mehrfach auftreten – charakterisiert die entsprechende Funktion F_n , wie die n Messwerte auf der λ -Achse verteilt sind. Aus diesem Grunde nennt man F_n eine zu einer aus n Messungen bestehenden Messreihe zugehörige (empirische) *Verteilungsfunktion*. Fasst man zusammen, so sieht man, dass F_n eine stückweise konstante Funktion ist, die die folgenden Eigenschaften besitzt:

- a) Es ist $0 \leq F_n(\lambda) \leq 1$ für alle λ .
- b) F_n ist monoton wachsend.
- c) F_n ist überall linksseitig stetig.
- d) Für hinreichend kleine λ -Werte (und zwar für solche, die kleiner oder gleich dem kleinsten Messwert sind) ist F_n identisch Null.
- e) Für hinreichend große λ -Werte (nämlich für alle λ -Werte, die größer als der größte Messwert sind) ist F_n identisch gleich 1.

Wir wiederholen: Die Sprungstellen der Funktion F_n sind die Messwerte λ_i , $i = 1, \dots, n$, wobei der Sprung von F_n an der Stelle λ_i gleich $\frac{k}{n}$ ist, wenn der Wert λ_i genau k -mal gemessen wird. Jedes F_n hat also maximal n Sprungstellen; genau n Sprungstellen liegen dann vor, wenn alle Messwerte nur einmal auftreten.

Lässt man n wachsen, führt man also Messreihen mit immer mehr Messungen durch, so werden die zugehörigen Verteilungsfunktionen F_n im allgemeinen also immer mehr Sprungstellen und dabei gleichzeitig immer kleinere Sprünge aufweisen. Betrachtet man beispielsweise neben einer Messreihe mit n Messungen und den Messwerten $\lambda_1, \dots, \lambda_n$ eine zweite Messreihe mit $2n$ (ebenfalls der Größe nach geordneten) Messwerten $\lambda'_1, \dots, \lambda'_{2n}$, so werden an die Stelle eines Messwertes λ_i der ersten Messreihe zwei Messwerte λ_{2i} und λ'_{2i+1} der zweiten Messreihe treten. Statt des einen Sprunges $\frac{1}{n}$ an der Stelle λ_i im Falle der ersten Verteilungsfunktion F_n wird die zu der zweiten Messreihe gehörende Verteilungsfunktion F_{2n} zwei Sprungstellen² λ_{2i} und λ'_{2i+1} mit den jeweiligen Sprüngen $\frac{1}{2n}$ besitzen. Durch Abb. 1.2 wird das Verhalten von F_n bzw. F_{2n} in der Nähe von λ_i bzw. λ'_{2i} wiedergegeben.

Der Verlauf der Funktion F_n hängt natürlich nicht nur von n ab, sondern auch davon, welche zufälligen Messwerte man bei einer aus n Messungen bestehenden Messreihe erhalten kann. Demzufolge hat es auch keinen Sinn, von einer eindeutig bestimmten Folge von Verteilungsfunktionen F_n für $n = 1, 2, \dots$ zu sprechen.

² die natürlich auch zusammenfallen können.

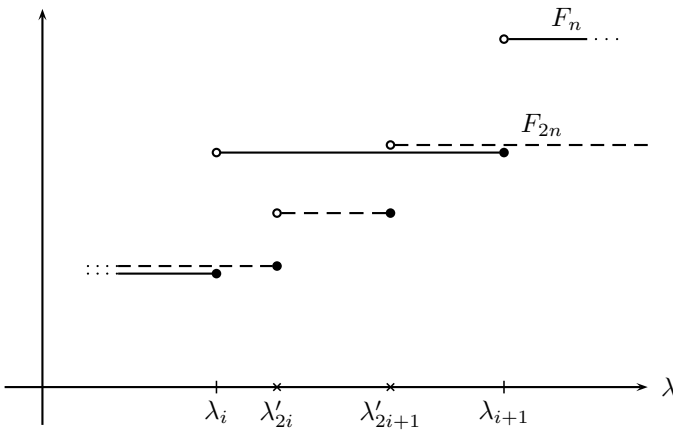


Abbildung 1.2 Messreihen mit n und $2n$ Messwerten

Andererseits ist aber für viele konkrete Messanordnungen aus der Erfahrung bekannt, dass für immer umfangreicher werdende Messreihen (d. h. für größer werdende n) die Verteilung F_n der Messwerte gegen eine Grenz-Verteilung F konvergiert. Präzise formuliert wollen wir die Existenz einer Grenz-Verteilungsfunktion $F = F(\lambda)$ annehmen, so dass die Verteilungsfunktionen F_n bei $n \rightarrow \infty$ punktweise gegen F konvergieren, d. h., dass für jedes λ bei $n \rightarrow \infty$ die Aussage

$$F_n(\lambda) \rightarrow F(\lambda)$$

gilt. Ausgehend von den fünf oben angegebenen Eigenschaften der Verteilungsfunktionen F_n wollen wir hierbei voraussetzen, dass die als Grenz-Verteilungsfunktionen in Betracht kommenden Funktionen F die folgenden Eigenschaften besitzen:

- a) Für alle λ gilt $0 \leq F(\lambda) \leq 1$.
- b) F ist monoton wachsend.
- c) F ist überall linksseitig stetig.
- d) $F(\lambda)$ geht gegen Null bei $\lambda \rightarrow -\infty$.
- e) Bei $\lambda \rightarrow +\infty$ strebt $F(\lambda)$ gegen 1.

Um nicht missverstanden zu werden: Es ist nicht möglich zu beweisen, dass bei einer konkret vorliegenden Messanordnung die F_n gegen eine bestimmte Grenz-Verteilungsfunktion F konvergieren. Vielmehr nimmt man – oftmals gestützt auf Erfahrungen – einfach an, dass die Messanordnung durch eine bestimmte Grenz-Verteilungsfunktion beschrieben werden kann. Ob eine ganz bestimmte Verteilungsfunktion F die Messanordnung wirklich (zumindest näherungsweise) beschreibt, wird durch einen sogenannten Signifikanz-Test entschieden.

Nun wollen wir uns nochmals daran erinnern, dass $F_n(\lambda)$ definiert war als Wahrscheinlichkeit dafür, dass bei einer aus n Messungen bestehenden Messreihe ein gemessener

Wert kleiner als λ ist. Analog kann man $F(\lambda)$ als Wahrscheinlichkeit dafür deuten, dass ein beliebiger Messwert kleiner als λ ist. Hierin kommt das Wesen des Übergangs von den Verteilungsfunktionen F_n zu der Grenz-Verteilungsfunktion F zum Ausdruck: Während bei den ersteren etwas ausgesagt wird über die Wahrscheinlichkeit eines einzelnen Messwertes im Rahmen einer aus genau n Messungen bestehenden Messreihe, gibt letztere die Wahrscheinlichkeit dafür an, dass irgendeine einzelne Messung einen Wert kleiner als λ liefert.

Sind λ_1 und λ_2 zwei beliebige λ -Werte, wobei $\lambda_1 < \lambda_2$ sei, so sind $F(\lambda_1)$ und $F(\lambda_2)$ also die Wahrscheinlichkeiten dafür, dass ein beliebiger Messwert kleiner als λ_1 bzw. kleiner als λ_2 ist. Damit wird durch

$$F(\lambda_2) - F(\lambda_1)$$

die Wahrscheinlichkeit dafür gegeben, dass ein willkürlich herausgegriffener Messwert λ der Ungleichung

$$\lambda_1 \leq \lambda < \lambda_2$$

genügt.

Was eigentlich eine Wahrscheinlichkeit ist, wollen wir hier nicht definieren. Bei der Definition von $F_n(\lambda)$ kam man mit der klassischen Wahrscheinlichkeitsdefinition aus, bei der die Wahrscheinlichkeit eines Ereignisses als der Quotient aus der Anzahl der hinsichtlich des ins Auge gefassten Ereignisses günstigen Fälle zu der Anzahl aller möglichen Fälle definiert wird.³ Wir wollen hier nur darauf hinweisen, dass man den Wert $F(\lambda)$ der Grenz-Verteilungsfunktion F an der Stelle λ wie folgt definieren kann: Sind bei n Messungen k_n Messwerte kleiner als λ , so ist

$$F(\lambda) = \lim_{n \rightarrow \infty} \frac{k_n}{n}.$$

Damit ist die Wahrscheinlichkeit $F(\lambda)$ als Grenzwert sogenannter relativer Häufigkeiten definiert.⁴

Wir wollen hier nur Verteilungsfunktionen F betrachten, die stetig differenzierbar sind. Bezeichnet man die Ableitung mit $p = p(\lambda)$, so kann $F(\lambda)$ bei Beachtung von Eigenschaft d) folglich in der Form

$$F(\lambda) = \int_{-\infty}^{\lambda} p(\lambda') d\lambda' \quad (1.1)$$

³ Sind bei einer aus n Messungen bestehenden Messreihe genau k_n Messwerte kleiner als λ , so ist demnach $F_n(\lambda) = k_n/n$.

⁴ Ein allgemeingültiger Wahrscheinlichkeitsbegriff ist der axiomatisch von A. N. KOLMOGOROW eingeführte (vgl. Kapitel 3, S. 15). Die durch eine Verteilungsfunktion F gegebene Wahrscheinlichkeit ordnet sich in dieses Konzept wie folgt ein: Ist A das Ereignis, dass der Messwert in dem Intervall $\lambda_1 \leq \lambda < \lambda_2$ liegt, so ist

$$PA = F(\lambda_2) - F(\lambda_1).$$

Die Menge Ω ist in diesem Fall die ganze λ -Achse, so dass beispielsweise $P\Omega = 1$ aus der Eigenschaft $\lim_{\lambda \rightarrow \infty} F(\lambda) = 1$ folgt.

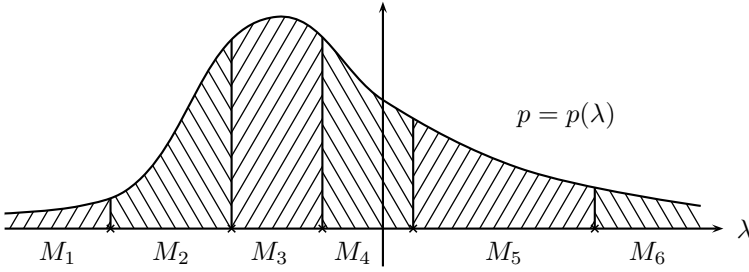


Abbildung 1.3 Wahrscheinlichkeitsdichten

geschrieben werden. Damit wird die Wahrscheinlichkeit dafür, dass der Messwert λ der Ungleichung $\lambda_1 \leq \lambda < \lambda_2$ genügt, durch

$$\int_{\lambda_1}^{\lambda_2} p(\lambda') d\lambda'$$

gegeben. Die Funktionen $p = p(\lambda)$ werden auch *Wahrscheinlichkeitsdichten* genannt.⁵ Wegen der Monotonie von F ist übrigens überall

$$p(\lambda) \geq 0.$$

Bei Beachtung von Eigenschaft e) von F folgt aus (1.1) die Relation

$$\int_{-\infty}^{+\infty} p(\lambda') d\lambda' = 1. \quad (1.2)$$

Nun wollen wir die λ -Achse so in n (endliche bzw. unendliche) Teilstücke $M_1, M_2, \dots, M_n$ zerlegen, dass

$$\int_{M_k} p(\lambda') d\lambda' = \frac{1}{n} \quad (1.3)$$

für jedes k ist⁶ (in Abb. 1.3 ist $n = 6$).

Es ist daher gleichwahrscheinlich, dass der betrachtete Messwert λ in einem der Intervalle M_k liegt. Diese Wahrscheinlichkeit ist für jedes dieser Intervalle gleich $\frac{1}{n}$.

⁵ Der Begriff Dichte stammt letztlich aus der Physik. Dort versteht man unter der Dichte einer räumlich möglicherweise ungleichmäßig verteilten Substanz die in der Volumeneinheit enthaltene Gesamtmenge. Ist die Dichtefunktion $c = c(x)$, so wird die im Gebiet G vorhandene Gesamtmenge durch $\iiint_G c(x) dx$ gegeben. Eine Anwendung des Begriffes der Dichte in der Theorie partieller Differentialgleichungen wird in Kapitel 14 auf Seite 175 erläutert.

⁶ Zerlegung der λ -Achse bedeutet, dass jeder Punkt λ in einem der M_k enthalten ist. Außerdem sollen benachbarte M_k höchstens einen Randpunkt (aber keine inneren Punkte) gemeinsam besitzen.

Über die Funktion $F = F(\lambda)$ wollen wir jetzt zusätzlich voraussetzen, dass für die Ableitung $p = p(\lambda)$ auch das Integral

$$\int_{-\infty}^{+\infty} \lambda' p(\lambda') d\lambda' \quad (1.4)$$

existiert.⁷ Sind $M_1, M_2, \dots, M_n$ eine Zerlegung der λ -Achse, so dass (1.3) gilt, so wird das Integral (1.4) zunächst in der Form

$$\sum_{k=1}^n \int_{M_k} \lambda' p(\lambda') d\lambda'$$

geschrieben. Auf jedes der n in dieser Summe auftretenden Integrale wird nun der zweite Mittelwert der Integralrechnung angewandt. Nach diesem existiert in jedem M_k (wenigstens) ein Zwischenpunkt $\tilde{\lambda}_k$, so dass

$$\int_{M_k} \lambda' p(\lambda') d\lambda' = \tilde{\lambda}_k \int_{M_k} p(\lambda') d\lambda'$$

ist. Beachtet man (1.3), so ergibt sich damit für das Integral (1.4) die Darstellung

$$\frac{1}{n}(\tilde{\lambda}_1 + \dots + \tilde{\lambda}_n). \quad (1.5)$$

Wegen (1.3) ist es gleichwahrscheinlich, dass ein Messwert von λ in einem der n Teilstücken M_k der λ -Achse liegt. Daher zeigt (1.5), dass das Integral (1.4) der Mittelwert von n gleichwahrscheinlichen Messwerten ist. Dies berechtigt uns, das Integral (1.4) als den zu der Verteilung $F = F(\lambda)$ gehörenden *Erwartungswert* zu interpretieren, also den Wert, den eine durch $F(\lambda)$ verteilte Größe eigentlich besitzt.

Bezeichnet man den Erwartungswert (1.4) mit $\bar{\lambda}$, so kann man wegen (1.2) den Erwartungswert auch durch

$$\int_{-\infty}^{+\infty} (\lambda' - \bar{\lambda}) p(\lambda') d\lambda' = 0 \quad (1.6)$$

charakterisieren. Beachtet man nochmals die Darstellung (1.5) des Integrals (1.4), so lässt sich (1.6) auch zu

$$\frac{1}{n}[(\tilde{\lambda}_1 - \bar{\lambda}) + \dots + (\tilde{\lambda}_n - \bar{\lambda})] = 0$$

⁷ Diese Voraussetzung braucht nicht erfüllt zu sein, auch wenn F alle anderen oben genannten Eigenschaften besitzt. Ist beispielsweise F die sogenannte CAUCHY-Verteilung

$$F(\lambda) = \frac{1}{\pi} \int_{-\infty}^{\lambda} \frac{1}{1 + \lambda'^2} d\lambda',$$

so existiert (1.4) nicht.

umformulieren. Daher lässt sich die Gleichung (1.6) auch so deuten: Der Erwartungswert ist dadurch charakterisiert, dass im Mittel alle denkbaren Messwerte die Abweichung Null von ihm besitzen.

Wir wollen nun zu der anfangs aufgeworfenen Frage zurückkehren, ob der Erwartungswert einer zufällig verteilten Größe zusätzlich auch durch die Methode der kleinsten Quadrate charakterisiert werden kann. Dazu nehmen wir über das gegebene $F = F(\lambda)$ an, dass nicht nur das Integral (1.4) existiert, sondern weiterhin auch noch

$$\int_{-\infty}^{+\infty} \lambda'^2 p(\lambda') d\lambda'. \quad (1.7)$$

Nun wollen wir $\hat{\lambda}$ so bestimmen, dass das von $\hat{\lambda}$ abhängende Integral

$$\int_{-\infty}^{+\infty} (\lambda' - \hat{\lambda})^2 p(\lambda') d\lambda' \quad (**)$$

einen minimalen Wert annimmt. Der Ausdruck (**) ersetzt hier die Summe (*). Rechnet man das Quadrat $(\lambda' - \hat{\lambda})^2$ aus, so ergibt sich für (**) die Darstellung

$$\int_{-\infty}^{+\infty} \lambda'^2 p(\lambda') d\lambda' - 2\hat{\lambda} \int_{-\infty}^{+\infty} \lambda' p(\lambda') d\lambda' + \hat{\lambda}^2, \quad (1.8)$$

wenn noch (1.2) beachtet wird. Die Form (1.8) von (**) zeigt außerdem noch, dass wegen der Existenz der Integrale (1.4) und (1.7) auch das Integral (**) bei jeder Wahl von $\hat{\lambda}$ existiert.

Nun stellt (1.8) ein Polynom in $\hat{\lambda}$ dar. Durch Differenzieren nach $\hat{\lambda}$ sieht man sofort, dass dieses Polynom dann minimal wird, wenn

$$\hat{\lambda} = \int_{-\infty}^{+\infty} \lambda' p(\lambda') d\lambda'$$

gewählt wird. Damit ist unter den oben zusammengestellten Voraussetzungen über die Verteilungsfunktion $F = F(\lambda)$ gezeigt:

Satz Die Anwendung der Methode der kleinsten Quadrate liefert den Erwartungswert auch im Falle von Messwerten, die eine (stetig differenzierbare) Verteilungsfunktion besitzen.

Anders formuliert: Unter den angegebenen Voraussetzungen, also bei verhältnismäßig allgemeinen Verteilungsfunktionen, ist die Methode der kleinsten Quadrate immer angepasst.

Ein Beispiel einer Verteilungsfunktion, für die alle erforderlichen Voraussetzungen zutreffen, ist

$$F(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{\lambda'^2}{2}} d\lambda'$$

(Normalverteilung mit 0 als Erwartungswert)⁸.

⁸ Die hier angegebene Normalverteilung hat außerdem die Varianz 1 (vgl. z. B. K. KRICKBERG [29]).

Abschließend noch eine Bemerkung über die rechnerische Anwendung der Methode der kleinsten Quadrate im Falle einer normalverteilten Variablen (oder bei einer Größe mit einer anderen zulässigen Verteilungsfunktion). In jedem konkreten Einzelfall liegen natürlich immer nur n Messwerte $\lambda_1, \dots, \lambda_n$ vor. Ist aber n sehr groß, so kann man davon ausgehen, dass in jedem der n gleichwahrscheinlichen Teilstücke M_k der λ -Achse je einer der Messwerte liegen wird, und man kann dann (**) näherungsweise durch $\frac{1}{n} \sum_{k=1}^n (\lambda_k - \hat{\lambda})^2$, also – bis auf den Faktor $\frac{1}{n}$ – wieder durch (*) ersetzen.

Die Methode der kleinsten Quadrate wurde hier beschrieben im Fall einer einzigen zu bestimmenden reellen Größe. Sie ist auch in komplizierteren Situationen anwendbar, z. B. bei der Parameterschätzung in linearen Modellen (vgl. z. B. V. NOLLAU [44]).