Mathematische Semesterberichte manuscript No.
(will be inserted by the editor)

Ulrich Krengel

Von der Bestimmung von Planetenbahnen zur modernen Statistik

Carl Friedrich Gauß — Werk und Wirkung

Eingegangen: date / Angenommen: date

Zusammenfassung Die Bedeutung von Carl Friedrich Gauß bei der Entdeckung wichtiger statistischer Begriffe und Methoden wird dargestellt. Dabei gehen wir auch Prioritätsfragen nach. Ein Ausblick auf den Einfluss von Gauß auf Entwicklungen im 20. Jahrhundert wird gegeben.

Schlüsselwörter Gauß · Statistik · Kleinste Quadrate

Mathematics Subject Classification (2000) 01-99 · 01A55

Inhaltsverzeichnis

Vortrag im Rahmen der von der Göttinger Akademie der Wissenschaften im Sommersemester 2005 veranstalteten öffentlichen Ringvorlesung „Carl Friedrich Gauß — Werk und Wirkung“

Ulrich Krengel
Institut für Mathematische Stochastik
Universität Göttingen
Maschmühlenweg 8–10
D-37073 Göttingen
Germany
E-Mail: krengel@math.uni-goettingen.de

1 Vorbemerkung

Im täglichen Sprachgebrauch wird das Wort „Statistik“ häufig in ganz anderer Bedeutung verwendet als in der Wissenschaft. Man spricht von Statistik z.B., wenn berichtet wird, wer das erste Tor in der Saison 2004/05 der Fußball-Bundesliga geschossen hat und wie viele Minuten nach Spielbeginn es fiel. Näher kommt man dem, was wir meinen, schon mit Daten wie Einwohnerzahlen, aber den Mathematiker und Anwender interessiert vor allem, was man „schließende Statistik“ nennt: Welche Schlüsse können wir aus beobachteten Daten ziehen?

Der bedeutende britische Physiker Sir William Thomson (1824–1907), genannt Lord Kelvin, nach dem die beim absoluten Nullpunkt beginnende Kelvin-Temperatur-Skala benannt ist, hat einmal gesagt: “When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.”

Ähnliche Aussagen haben z.B. auch Leonardo da Vinci, Kant, Roger und Francis Bacon gemacht. Dies unterstreicht die überragende Rolle von Messungen und Daten für die Wissenschaft. Vielleicht ist das obige Zitat etwas überspitzt formuliert — schließlich waren auch die Errungenschaften von Alexander von Humboldt große Wissenschaft, obwohl vieles davon nicht in Zahlen beschreibbar war — aber es trifft doch einen entscheidenden Punkt. Damit wird auch die Bedeutung von „Mathematischer Statistik“ für alle messende Wissenschaft sichtbar:

> Mathematische Statistik ist die Wissenschaft von der Planung und Auswertung von Beobachtungen und Messungen. Auf Grund eines möglichst realistischen Modells zufälliger Einflüsse werden Abschätzungen der Unsicherheit und des zu erwartenden Verlustes bei Fehlbeurteilungen angegeben und damit die Qualität von Ergebnissen und Verfahren beurteilt.

Wenn wir diese Definition akzeptieren, war, wie wir sehen werden, Carl Friedrich Gauß einer der bedeutendsten Gründungsväter dieser Wissenschaft.

2 Die Bestimmung der Ceres-Bahn

Am 1. Januar 1801 entdeckte der italienische Astronom Guiseppe Piazzi (1746–1826) den Planetoiden Ceres, einen Kleinplaneten, dessen Durchmesser mit ca. 780 km nur ein Sechzehntel des Erddurchmessers und weniger als ein Viertel des Monddurchmessers beträgt. Piazzi konnte ihn nur 41 Tage beobachten, dann gelangte der Planetoid in eine Region des Himmels, in der die hellen Sonnenstrahlen eine Beobachtung unmöglich machten. Gauß erhielt im Oktober 1801 Zugang zu den Beobachtungsdaten von Piazzi. Er hatte sich schon in den Jahren davor mit astronomischen Berechnungen, insbesondere mit Methoden zur Bestimmung von Planetenbahnen beschäftigt. Verschiedene Astronomen versuchten aus den wenigen Messungen Piazzis die Bahn von Ceres zu bestimmen, um ihn wieder zu finden, aber nur die Ergebnisse von Gauß waren genau genug. Auf Grund seiner Angaben fand am 7. Dezember 1801 der Astronom Franz Xaver von Zach (1754–1832) auf dem Seeberg bei Gotha Ceres fast genau in der von Gauß angegebenen Position, unabhängig davon wenig später auch der Bremer Astronom Wilhelm Olbers (1758–1840), der später einer der besten Freunde von Gauß wurde.

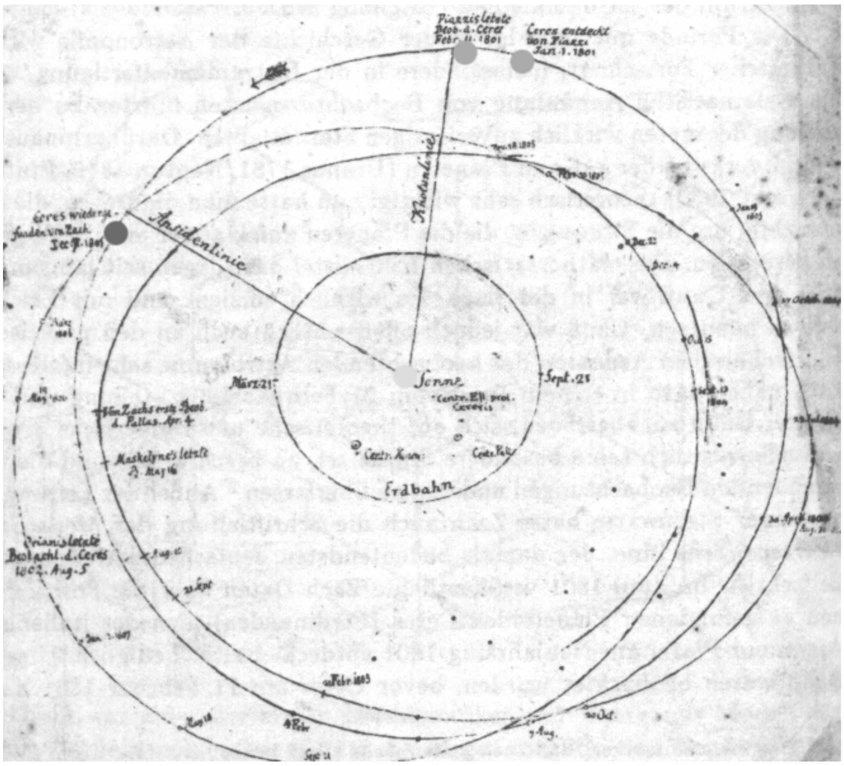

Abb. 1 Skizze der Ceres- und Pallasbahn (Nachlass Gauß, Handbuch 4). Niedersächsische Staats- und Universitätsbibliothek Göttingen. Die dicht nebeneinander liegenden dicken Punkte ganz oben markieren (entgegen dem Uhrzeigersinn) die Positionen von Ceres bei der ersten und letzten Beobachtung Piazzis. Im dritten Punkt wurde der Planetoid durch von Zach wiederaufgefunden.

Diese Leistung von Gauß machte ihn auf einen Schlag international berühmt und war letztlich ausschlaggebend für seine Berufung nach Göttingen. Gauß hatte schon vorher in der Mathematik Spitzenleistungen erzielt, die Geschichte gemacht haben, aber für die Öffentlichkeit und für die Herrschenden war die Astronomie eine Königsdisziplin.

Nach eigenen Angaben hatte Gauß schon bei der Bahnbestimmung der Ceres und davor eine Methode angewandt, die heute unter dem Namen „Methode der kleinsten Quadrate" in der Statistik eine herausragende Rolle spielt.

3 Die Methode der kleinsten Quadrate

Ich möchte die Methode zunächst an einem einfachen Beispiel erklären und hoffe, durch die wenigen einfachen Formeln niemand zu verschrecken.

Oft ist man daran interessiert, wie mehreren Messpunkten eine Gerade oder eine andere einfache Kurve am besten angepasst werden kann. Gesucht ist also

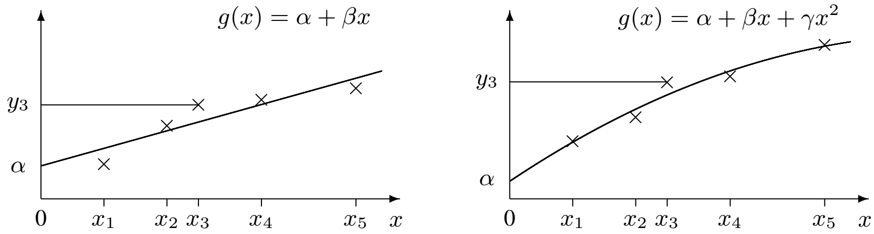

Abb. 2 Zur Methode der kleinsten Quadrate

z.B. eine Gerade $g(x) = \alpha + \beta x$, oder eine Parabel $g(x) = \alpha + \beta x + \gamma x^2$, die möglichst gut zu mehreren gemessenen Punkten passt.

In Abhängigkeit von gegebenen Werten $x_1, \ldots, x_5$ einer Größe werden zugehörige Werte einer anderen Größe gemessen. Wie sind die Unbekannten „Parameter" (Zahlen) $\alpha, \beta, \ldots$ zu wählen, damit die Gerade (oder Parabel) möglichst gut zu den Messwerten passt? Die Methode der kleinsten Quadrate besagt: Die Gerade (Parabel) ist so zu wählen, dass die Summe der Quadrate der Fehler minimal ist:

$$\big((y_1 - g(x_1)\big)^2 + \big((y_2 - g(x_2)\big)^2 + \cdots + \big((y_5 - g(x_5)^2\big) \qquad \text{Minimal!}$$

Diese Methode wurde — ebenfalls im Zusammenhang mit astronomischen Untersuchungen — im Jahr 1805 von Adrien-Marie Legendre (1752–1833) publiziert. Schon vorher war eine ähnliche Methode von Roger Joseph Boscovich (1711–1787) benutzt worden, der aber die Summe der Fehler (und nicht der Quadrate) minimiert hatte. Dies war rechnerisch viel aufwändiger.

Was war nun mit der Methode der kleinsten Quadrate im Fall von Planetenbahnen zu tun? Ich will ein wenig vereinfachen und statt einer Ellipse im dreidimensionalen Raum das Problem nur für einen Kreis in der Ebene beschreiben. Bekannt sind nun Messpunkte $M_1, M_2, \ldots, M_5$ für die Lage des Planetoiden und die Zeitpunkte $t_1, t_2, \ldots, t_5$ der Messungen (s. Abb. 3). Es ist eine Kreisbahn anzugeben, die möglichst gut angepasst ist, aber auch Bahnpunkte $V_1, V_2, \ldots, V_5$ auf dem Kreis. Seien p, q die Koordinaten des Kreismittelpunktes und sei r der Radius des Kreises. Die Abstände der Bahnpunkte sind dabei so zu wählen, dass sie zu der durch den Radius eindeutig bestimmten Geschwindigkeit des Planetoiden passen. Und alles muss so festgelegt werden, dass die Summe der Quadrate der Abstände der Beobachtungspunkte von den Bahnpunkten minimal ist:

$$|M_1 - V_1|^2 + |M_2 - V_2|^2 + \cdots + |M_5 - V_5|^2 \qquad \text{Minimal!}$$

Die so gefundene Kreisbahn ist die nach der Methode der kleinsten Quadrate bestimmte. Verzichtet man auf die Berücksichtigung der Zeitpunkte, so wird alles etwas einfacher aber auch unsicherer.

4 Zur Prioritätenfrage

Legendre publizierte die Methode der kleinsten Quadrate 1805 in seiner Schrift „Nouvelles méthodes pour la détermination des orbites des comètes". Er betrach-

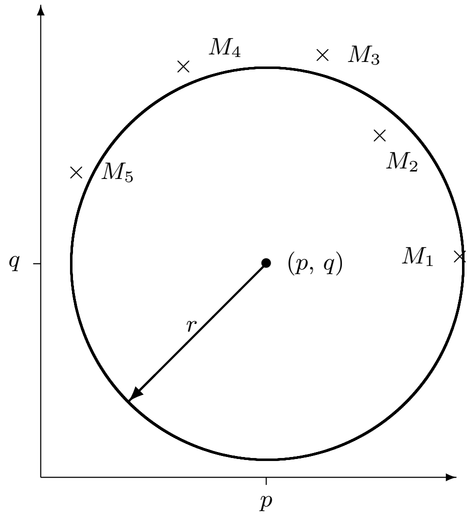

Abb. 3 Methode der kleinsten Quadrate zur Berechnung einer Kreisbahn

tete nur den linearen Fall, d.h. die gesuchten Parameter α, β, ... waren mit bekannten Größen, z.B. x_1, x_2^3, ... zu multiplizieren und die so erhaltenen Terme zu addieren. Er gab nur Plausibilitätsargumente als Begründung an, z.B. dass die damals schon bekannte Verwendung von Mittelwerten ein Spezialfall sei.

Gauß beschrieb und begründete die Methode 1809 in seinem astronomischen Hauptwerk „Theoria motus corporum coelestium in sectionibus conicis solem ambientium“ (im Folgenden kurz: „Theoria motus“) auch im allgemeineren nichtlinearen Fall. (Von den oben beschriebenen drei Beispielen sind die Gerade und Parabel mit dem Ansatz von Legendre beschreibbar, aber für die Kreisbahn benötigt man den allgemeineren Fall.) Gauß sprach von „seiner Methode“ (principium nostrum) und davon, dass er die Methode seit 1795 benutzt habe. Dies löste einen Prioritätsstreit aus, der noch heute die Gemüter bewegt.

Die Priorität, was den Publikationszeitpunkt anbelangt, gehört ganz offensichtlich Legendre. Vielleicht wäre die Kontroverse nie so scharf ausgetragen worden, wenn Gauß ein wenig Einfühlungsvermögen bewiesen hätte. Man kann nur spekulieren, ob Legendre, der zuerst in sehr freundlicher Form darauf hingewiesen hatte, dass ihm an der Priorität viel lag, sich überhaupt beschwert hätte, wenn Gauß schon 1809 geschrieben hätte, die Methode sei offenbar unabhängig von Legendre und ihm gefunden worden, ohne darauf zu bestehen, dass er sie zuerst verwendet habe.

Dass Gauß die Methode der kleinsten Quadrate schon vor 1800 gekannt und benutzt hat, ist heute kaum noch zu bezweifeln, wenn es auch keinen absolut lückenlosen Beweis gibt.

Zunächst einmal wird die Behauptung von Gauß durch zwei Einträge in sein Tagebuch gestützt, in das er in Kurzform beginnend mit der Eintragung der Konstruierbarkeit des regelmäßigen 17-Ecks mit Zirkel und Lineal (1796) seine wichtigsten Entdeckungen eintrug. Ein Eintrag von 17.6.1798 lautet: „Calculus probabilitatis contra LaPlace defensus“ (Wahrscheinlichkeitsrechnung gegen Laplace verteidigt). Dies lässt sich wohl kaum anders deuten, als dass Gauß festgestellt hatte, dass die Methode der kleinsten Quadrate sich mit einem Maximum-Likelihood-Argument wahrscheinlichkeitstheoretisch rechfertigen ließ, während dies für die von Laplace verwendete (wie oben geschildert auf Boscovich zurückgehende) Methode nicht galt. Diese Interpretation hat Gauß auch später in einem Brief an Olbers gegeben.

Ein Eintrag im Juni 1798 lautet: „Problema eliminationis ita solutum ut nihil amplius desiderati possit." (Das Eliminationsproblem so gelöst, dass es nichts zu wünschen übrig lässt.) Die berühmte gaußsche Elimination hat Gauß in Zusammenhang mit der Methode der kleinsten Quadrate publiziert, die wegen der sogenannten Normalgleichungen ganz natürlich zu dieser Fragestellung führen.

Ferner gibt es hoch angesehene Zeugen. Nachdem der Streit zwischen Gauß und Legendre ausgebrochen war, hat Gauß 1812 seinen Kollegen Olbers gebeten, öffentlich zu bezeugen, dass Gauß ihm Anwendungen der Methode schon 1803 gezeigt habe. Olbers schrieb sofort zurück, er wolle dies gern und willig bezeugen. Dass er eine entsprechende Fußnote erst 1816 publizierte, muss nicht bedeuten, dass er sich unsicher war. Es könnte ihm vielleicht auch unangenehm gewesen sein, in die Sache hineingezogen zu werden.

In einem 1844 geschriebenen Brief des Astronomen Friedrich Wilhelm Bessel (1784–1846) an Alexander von Humboldt steht: „Legendre ist der Erfinder der Methode der kleinsten Quadrate, wenn ich auch Zeugnis ablegen kann, dass Gauß mir, ehe Legendre etwas darüber drucken ließ, etwas davon erzählt hat." Für Bessel ist also der Publikationstermin das ausschlaggebende Kriterium, und er stellt sich auf die Seite von Legendre, aber dennoch bemüht er sich um Objektivität.

Es ist nahe liegend, auch in den von Gauß vor 1805 publizierten Resultaten nach Spuren der Methode der kleinsten Quadrate zu suchen, obwohl Gauß bedauert hat, dass die Blätter, auf denen er die Resultate ausgerechnet hat, nicht von ihm aufgehoben wurden. Für die Bahnbestimmung der Ceres scheint dies unmöglich, weil Gauß in der Einleitung zur „Theoria motus" schreibt, er habe seit 1801 die Methoden mehrfach verändert und verbessert.

Ein anderer Zugang ist 1981 von S. Stigler versucht worden. Gauß schrieb 1799 an von Zach: „Es gibt einen Druckfehler in dem Artikel über die Messungen des Längengrades. Ich fand diesen Fehler als ich meine Methode anwandte, die ich Ihnen gezeigt habe." Diese Bemerkung ist mit den von Gauß genannten korrigierten Daten im Jahr 1799 durch von Zach publiziert worden. Sie bezog sich auf Entfernungsmessungen des durch den Pantheon in Paris führenden Meridians, die zur Bestimmung der Länge des Meters dienten. Gauß' Freund, der Geodät Schumacher schlug Gauß 1831 brieflich vor, mit Hilfe dieser Zahlen zu beweisen, dass er die Wahrheit gesagt habe. Gauß lehnte dies ab und bestand darauf, man solle ihm vertrauen, ohne solche Methoden anzuwenden. Die Untersuchungen von Stigler und anderen (Celmiņš, Brosche, Odenkirchen) haben letztlich auch nicht zu dem gewünschten Nachweis geführt. Entweder hat Gauß sich verrechnet oder heute nicht mehr rekonstruierbare Ansätze gemacht.

Vor wenigen Jahren sind aber nun doch noch Spuren der Anwendung der Methode der kleinsten Quadrate in einer anderen Arbeit von Gauß aus dem Jahr 1799 nachgewiesen worden.

Abb. 4 Ulugh Beg (1394–1449) und Adrien-Marie Legendre (1752–1833)

In dem schon erwähnten Brief an Olbers schreibt Gauß: „Die Papiere, worin ich in früheren Jahren, z.B. 1799, die Methode auf die Zeitgleichungstafel von Ulugh Beg[1] angewandt habe, sind verloren gegangen . . . “

Von dem gaußschen Manuskript war 1799 nur ein Auszug publiziert worden. Das Originalmanuskript wurde 1924 von dem Göttinger Astronomen Brendel in den Gauß-Werken herausgegeben, wobei Brendel eigene Deutungen einbrachte. Auf Grund dieser Methoden und Daten kam Jacques Dutka 1996 zu dem Schluss:

„Although some of the details of Gauss' calculations have been lost, there can be little doubt after comparison with the computer generated values above that he actually used the method of least squares in the spring of 1799 and communicated the results as stated in his letter to von Zach."

Gauß hat sich in Prioritätsfragen sicher nicht so verhalten, wie man es von einem Menschen von Format erwarten sollte. Zum Beispiel hat er die Existenz einer nichteuklidischen Geometrie früh entdeckt und nur in seinem Tagebuch vermerkt, und als Janos Bolyai unabhängig von Lobatschewski und Gauß die gleiche Entdeckung machte, verletzte er diesen schwer durch die Bemerkung, dies zu loben hieße sich selbst loben. Es gab weitere Beispiele. G. W. Stewart weist im „Afterword" zu seiner schönen englischen Übersetzung von Gauß' „Theoria combinationis observationum erroribus minimis obnoxiae" (auf die wir später zu sprechen kommen) darauf hin, dass der Konflikt mit Legendre anders gelagert ist. Um andere Prioritäten hat Gauß nicht gekämpft. Er wird in Zusammenhang mit der nichteuklidischen Geometrie zwar erwähnt, aber Bolyai und Lobatschewski werden als die Entdecker gefeiert. Dass Gauß um die Priorität im Fall der Methode der kleinsten Quadrate kämpfte, lag vielleicht daran, dass die „Theoria motus" zur Zeit der Publikation von Legendres Arbeit schon weit gediehen war und die Fertigstellung durch verschiedene Umstände verzögert wurde [siehe die Korrespondenz mit seinem Verleger Perthes in Reich (2001)]. Jedenfalls hat er

[1] Ulugh Beg (1394–1449) war ein Königssohn in Samarkand, der als Astronom gearbeitet hat. Es ging um Daten über die Exzentrizität der Erdbahn.

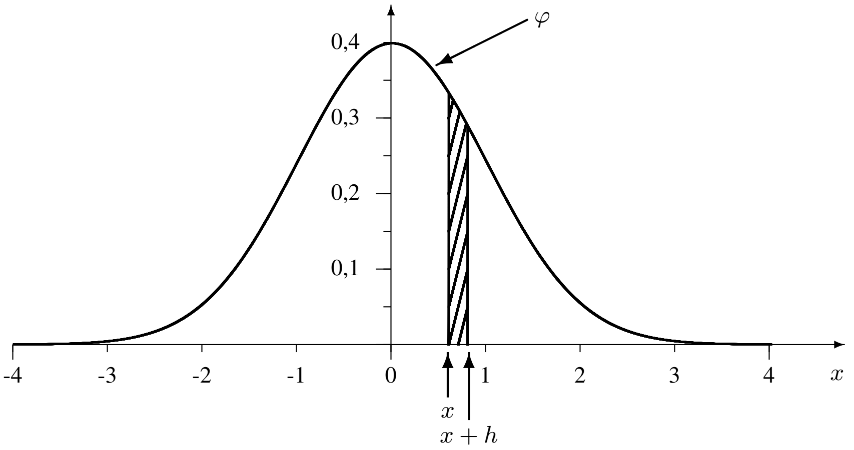

Abb. 5 Dichte einer Normalverteilung

viel Arbeit in diese Untersuchung gesteckt. Stewart schließt: „Legendre and Gauß should be considered as independent codiscoverers. Legendre caused the method of least squares to become widely known. Gauß and Laplace went on to give it a sound theoretical basis.“

Ich würde mich nicht denen anschließen, die den Publikationszeitpunkt als ausschließliches Kriterium für Priorität betrachten wollen. Aus eigener Erfahrung weiß ich, dass dies zu Fehlbeurteilungen führen kann. Wenn man allerdings längere Zeit eine Entdeckung in seinen Papieren vergräbt, nicht daran weiterarbeitet und niemand davon berichtet, kann man ganz sicher nicht später auf der Priorität bestehen.

5 Der Gedankengang von Gauß

Ich möchte nun skizzieren, wie Gauß 1809 die Methode der kleinsten Quadrate begründete und wie ihn dies zur Benutzung der Normalverteilung führte.

Es seien unbekannte Zahlen zu bestimmen, etwa die Koordinaten p, q eines Kreismittelpunktes und der Radius r. Dazu beobachten wir „Bahnpunkte“ $V_1, V_2, \ldots, V_n$. Da es immer (hoffentlich kleine) Messfehler gibt, weichen die registrierten Messpunkte $M_1, \ldots, M_n$ von den wahren Werten $V_1, \ldots, V_n$ (die wir nicht kennen) ab. Ich vereinfache jetzt ein wenig und betrachte die $V_1, \ldots, V_n$ und $M_1, \ldots, M_n$ als eindimensionale Größen. (Sonst müsste man die beiden Koordinaten getrennt messen.)

$M_1 - V_1$ ist der Beobachtungsfehler bei der Beobachtung von V_1, $M_2 - V_2$ der bei der Beobachtung von V_2, usw. Die Größe des Fehlers ist zufallsabhängig. Für eine uns zunächst unbekannte Funktion φ ist für kleines $h > 0$ die Wahrscheinlichkeit, dass der Fehler zwischen zwei Werten x und $x+h$ liegt, ungefähr $\varphi(x) \cdot h$ (genauer: er ist gleich der Fläche unter der Kurve φ zwischen den Werten x und $x+h$).

Gauß nimmt an, dass φ symmetrisch zu 0 ist, stetig und sehr große Fehler sehr unwahrscheinlich sind. Er setzt außerdem voraus, dass die Fehler bei den verschiedenen Beobachtungen unabhängig sind. Die Wahrscheinlichkeit, dass die Werte $M_1, \ldots, M_n$ beobachtet werden, wird also durch eine „Dichte“

$$\Omega = \varphi(M_1 - V_1) \cdot \varphi(M_2 - V_2) \cdot \cdots \cdot \varphi(M_n - V_n)$$

beschrieben. Am „plausibelsten“ sind die Werte von p, q, r, die den tatsächlichen Beobachtungen die größte Wahrscheinlichkeit verleihen, für die also Ω am größten wird. Diese Werte nennt man Maximum-Likelihood-Schätzer.

[Genauer geht Gauß von einer Gleichverteilung in dem Raum der Werte (p, q, r) aus (die es nicht gibt). Dies müsste man durch ein Grenzwertargument präzisieren. Er spricht dann von den wahrscheinlichsten Werten p, q, r.]

Wir können nun aber die plausibelsten Parameterwerte nicht bestimmen, weil wir φ nicht kennen. Um φ zu bestimmen, stützt sich Gauß auf die folgende Beobachtung: Wenn eine einzige unbekannte Größe zu bestimmen ist, etwa eine Länge l, wird diese oft mehrmals gemessen und der Mittelwert der Messergebnisse $M_1, \ldots, M_n$, also

$$\frac{1}{n}\bigl(M_1 + M_2 + \cdots + M_n\bigr)$$

gilt als der „wahrscheinlichste“ Wert von l. Wir haben hier den Spezialfall

$$V_1 = l,\ V_2 = l,\ \ldots,\ V_n = l\,.$$

Gauß zeigt nun mit einem kurzen aber höchst originellen Argument: Der Mittelwert ist genau dann für jedes n der Maximum-Likelihood-Schätzer von l, wenn φ die folgende Gestalt hat:

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\!\left(-\frac{x^2}{2\,\sigma^2}\right).$$

Dabei ist σ Maß für die Genauigkeit. Ist σ groß, so streuen die Beobachtungen stark; sie sind dann ungenauer. Gauß zeigt ferner: Mit dieser Wahl von φ ergibt auch für allgemeine Messungen mehrerer verschiedener Punkte $V_1, V_2, \ldots, V_n$ die Methode der kleinsten Quadrate gerade den Maximum-Likelihood-Schätzer.

> **Gauß’ Folgerung:** Die Methode der kleinsten Quadrate ist anzuwenden, wenn bei allen Beobachtungen der gleiche Grad der Genauigkeit zu präsumieren ist.
> Dieser Grundsatz … muss allenthalben an Stelle eines Axioms mit demselben Rechte gelten, mit welchem das arithmetische Mittel unter mehreren beobachteten Werten derselben Größe als der wahrscheinlichste Wert angenommen wird.

1986 hat der angesehene Statistiker Lucien Le Cam in der viel gelesenen Zeitschrift „Statistical Science“ die gaußsche Überlegung heftig kritisiert. Er schrieb in Anführungszeichen (also scheinbar als Zitat?) Gauß habe argumentiert: „Everyone knows that the average of observations is the best estimate of the expectation … Therefore the method of least squares is best.“ Dies sei ein Zirkelschluss, da die Mittelwertbildung ein Spezialfall der Methode der kleinsten Quadrate ist. Wie man oben sieht, schließt Gauß nicht so. Er behauptet nicht, dass die Methode optimal sei, sondern nur, dass sie im allgemeinen Fall nicht weniger berechtigt sei

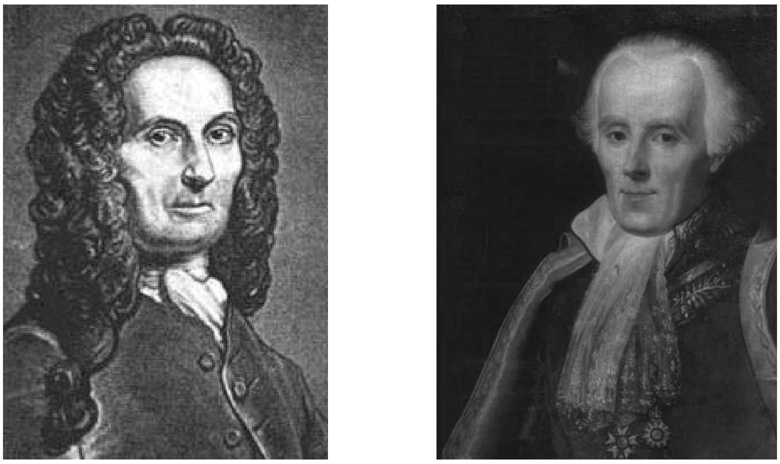

Abb. 6 Abraham de Moivre (1667–1754) und Pierre Simon Laplace (1749–1827)

als im Spezialfall. Es ist richtig, dass die gaußschen Schlüsse nicht so sauber sind wie z.B. seine Arbeiten zur Zahlentheorie. Aber die Grundlagen der Wahrscheinlichkeitstheorie wurden erst im 20. Jahrhundert systematisch entwickelt. Mit den heutigen Begriffsbildungen lässt sich der Gedankengang von Gauß völlig rigoros fassen.

Die von Gauß angegebene Funktion φ wird auch gaußsche Glockenkurve genannt. Sie wurde zu Ehren von Gauß auch auf dem alten Zehn-DM-Schein abgebildet. Die zugehörige Wahrscheinlichkeitsverteilung hieß früher Gauß-Verteilung und heute meist Normalverteilung. Sie hat auch einer wichtigen Klasse von stochastischen Prozessen, den Gauß-Prozessen, den Namen gegeben.

Es stellt sich daher die Frage, wieweit dies berechtigt ist, denn diese Wahrscheinlichkeitsverteilung hat der bedeutende französische Mathematiker Abraham de Moivre (1667–1754), der als Protestant nach England emigrieren musste, als erster abgeleitet. Er hat im Jahr 1733 diese Verteilung als Grenzwert der Binomialverteilung (bei passender Normierung) erhalten. Dies war die Urform des neben dem Gesetz der großen Zahl berühmtesten wahrscheinlichkeitstheoretischen Satzes, des so genannten Zentralen Grenzwertsatzes. Gauß dürfte dieses Resultat von de Moivre gekannt haben. Der Name De-Moivre-Verteilung wäre daher der historisch korrektere.

Aber der Satz von de Moivre ermöglichte vor allem schnelle approximative Berechnungen von Wahrscheinlichkeiten. Gauß hat die Verteilung mit der realen Welt verbunden und in die Statistik eingeführt. Damit hat er einen außerordentlich bedeutenden Einfluss auf die Entwicklung der Statistik im 19. und 20. Jahrhundert ausgeübt. Die an astronomischen Anwendungen arbeitenden Wissenschaftler waren an den neuen Anwendungen interessiert und weniger an einem Grenzwertsatz.

6 Einwände von Pierre Simon Laplace (1810)

Die gaußsche Arbeit hat sofort nach ihrem Erscheinen den berühmten französischen Mathematiker Pierre Simon Laplace (1749–1827) inspiriert. Laplace hatte gerade an Verallgemeinerungen des Satzes von de Moivre gearbeitet und wichtige erste Verallgemeinerungen des Zentralen Grenzwertsatzes bewiesen. Natürlich erkannte er die Funktion φ sofort wieder.

Er brachte zwei berechtigte Einwände gegen die gaußsche Rechtfertigung der Methode der kleinsten Quadrate vor. Wir beschränken uns hier auf den Spezialfall von mehreren unabhängigen Messungen eines Wertes μ:

1. Es kommt nicht darauf an, als Schätzwert $\hat{\mu}$ von μ den Wert zu wählen, der den gemachten Beobachtungen die größte Wahrscheinlichkeit verleiht, sondern den, für den der Schätzfehler im Mittel am kleinsten ist.
2. Laplace glaubte, dass man nicht allgemein von normalverteilten Fehlern ausgehen könne.

Gauß hat diese Kritikpunkte akzeptiert. Die weitere Entwicklung des Themas wurde nun aber erstmal von Laplace vorangetrieben, denn Gauß war inzwischen mit anderen Arbeiten beschäftigt. Laplace betrachtete wieder die von Legendre formalisierte lineare Situation. Im Fall von zwei Parametern α, β, die zu schätzen sind, macht man also einen Ansatz

$$\begin{aligned} Y_1 &= \alpha\, A_1 + \beta\, B_1 + E_1\,, \\ Y_2 &= \alpha\, A_2 + \beta\, B_2 + E_2\,, \\ &\vdots \\ Y_n &= \alpha\, A_n + \beta\, B_n + E_n\,. \end{aligned}$$

Dabei sind $Y_1, \ldots, Y_n$ die beobachteten zufälligen Werte, und die Zahlen A_1, $B_1, \ldots, A_n$, B_n sollen bekannt sein. Die Messfehler E_1, $E_2, \ldots$ seien unabhängig voneinander, sie sollen eine zum Punkt 0 symmetrische Verteilung haben und beschränkt sein. (Fehler in einer Richtung seien also ebenso wahrscheinlich wie in der entgegengesetzten Richtung.) Man muss natürlich annehmen, dass sich α und β überhaupt schätzen lassen. Wenn also alle Messfehler $= 0$ sind, müssen sich α und β ausrechnen lassen.

Wir nennen der Einfachheit halber das Ergebnis von Laplace nur in dem Fall, in dem nur α zu schätzen ist und $\beta = 0$ ist. Laplace schränkt sich auf die Betrachtung von Schätzern $\hat{\alpha}$ von α ein, die Linearkombinationen der Beobachtungswerte sind. Beispiele sind

$$\hat{\alpha} = \frac{1}{7}\,(Y_1 + Y_2 + \cdots + Y_7)$$

oder

$$\hat{\alpha} = \frac{4}{3}\,Y_1 + \frac{1}{8}\,Y_2 - 2\,Y_3\,.$$

Als Begründung für diese Einschränkung gibt Laplace an, es sei unpraktisch sich bei vielen Gleichungen auf allgemeinere Ansätze einzulassen. Laplace zeigt dann:

Wenn der Schätzer $\hat{\alpha}$ nahezu normalverteilt und erwartungstreu ist, was man bei sehr hohen Beobachtungszahlen erreichen kann, ist die Methode der kleinsten Quadrate (nahezu) die Methode, für die der Erwartungswert $E\big(|\hat{\alpha} - \alpha|\big)$ des Schätzfehlers am kleinsten ist. Auch der Erwartungswert $E\big((\hat{\alpha} - \alpha)^2\big)$ des Fehlerquadrats wird minimiert.

7 Der Satz von Gauß

Gauß kehrte mit einem Paukenschlag in die Diskussion zurück. Seine Arbeit „Theoria combinationis observationum erroribus minimis obnoxiae“ (im Folgenden kurz: „Theoria combinationis observationum“), von der der erste Teil 1821, der zweite Teil 1823 und ein Supplement 1828 erschien, war ein Meilenstein für die Mathematische Statistik. Für den Satz von de Moivre gilt, dass man oft sehr hohe Beobachtungszahlen (z.B. bei Münzwürfen) braucht, um eine gute Approximation durch die Normalverteilung zu erhalten. Gauß zeigte, dass nicht nur für große Beobachtungszahlen sondern für jede Beobachtungszahl und nicht nur approximativ sondern präzise Folgendes gilt:

> Wenn es überhaupt einen linearen erwartungstreuen Schätzer von α (oder allgemeiner von einer Linearkombination der Parameter) gibt, ist der „Kleinste-Quadrate-Schätzer“ der mit minimaler Varianz.

Gauß nahm nur an, dass die Varianz der Beobachtungsfehler gleich sei, er kam also obendrein noch mit viel schwächeren Annahmen aus als Laplace. Sein Satz ist noch heute der wichtigste Satz in der Theorie des Schätzens in linearen Modellen. Gauß selbst hielt ihn für einen der wichtigsten Sätze in der Anwendung der Mathematik auf die Naturwissenschaften.

Heute ist der Satz von Gauß relativ leicht zu beweisen. Wir haben die ganze Hilbertraumtheorie zur Verfügung, verstehen Zufallsvariable als Elemente eines unendlich-dimensionalen Raums, wissen, dass Unabhängigkeit Orthogonalität zur Folge hat, usw. Es ist als ob Gauß diese erst mehr als hundert Jahre später entstandenen Begriffe intuitiv vorausgeahnt hat.

In vielen Lehrbüchern wird der Satz von Gauß als Satz von Gauß-Markow bezeichnet. Dies beruht auf einem Missverständnis von J. Neyman, der 1934 als Quelle ein Lehrbuch von Markow aus dem Jahr 1912 angab. Markow hat dort zwar im Text Gauß nicht zitiert, aber aus seinem Literaturverzeichnis geht hervor, dass er die damals schon gut bekannte Literatur zur Methode der kleinsten Quadrate kannte. 1934 war Markow aber schon 12 Jahre tot, und konnte das Missverständnis nicht mehr ausräumen. (Vgl. Hald 1998.)

Die Arbeit „Theoria combinationis observationum“ ist aber nicht nur wegen des obigen Satzes von Gauß ein Meilenstein, sondern sie enthält weitere historisch wichtige Beiträge:

- Diskussion allgemeiner Verlustfunktionen, speziell der quadratischen Verlustfunktion
- Einführung der Varianz (die Gauß noch mittleren quadratischen Fehler nannte)
- Schätzung der Genauigkeit von Schätzungen, speziell Schätzung der Varianz durch
  $$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 ,$$
  wobei $\bar{x}$ der Mittelwert der Beobachtungswerte x_i ist.
- Schätzung von Parametern unter Nebenbedingungen
- Rekursives Updating von Schätzungen
- Fehlerfortpflanzungsgesetz

Die gaußsche Fehlertheorie wurde weltberühmt. Ich denke, dass Gauß der Erste war, der die eingangs gegebene Begriffsbestimmung des Mathematischen Statistikers voll erfüllte. Man kann natürlich auch Laplace nennen, der in die gleiche Richtung zielte, der aber heute eher wegen seiner Beiträge zum Zentralen Grenzwertsatz in Erinnerung bleibt.

8 Wie wurden die gaußschen Ideen und Methoden weiterentwickelt?

Die gaußschen Arbeiten beziehen sich auf Messfehler. Der Anwendungsbereich der Normalverteilung wurde im 19. Jahrhundert stark ausgeweitet, indem die Normalverteilung als nützliche und oft passende Beschreibung für biologische und soziale Größen eingeführt wurde. Nun lag die Zufälligkeit nicht mehr in den Schwankungen der verschiedenen Messergebnisse bei der Messung einer festen Größe sondern in den natürlichen Schwankungen der Größe verschiedener Objekte, etwa Körpergrößen, Ernteerträge usw. Diese Ausdehnung der Anwendbarkeit verdanken wir vor allem Adolphe Quetelet (1796–1874) und Francis Galton (1822–1911).

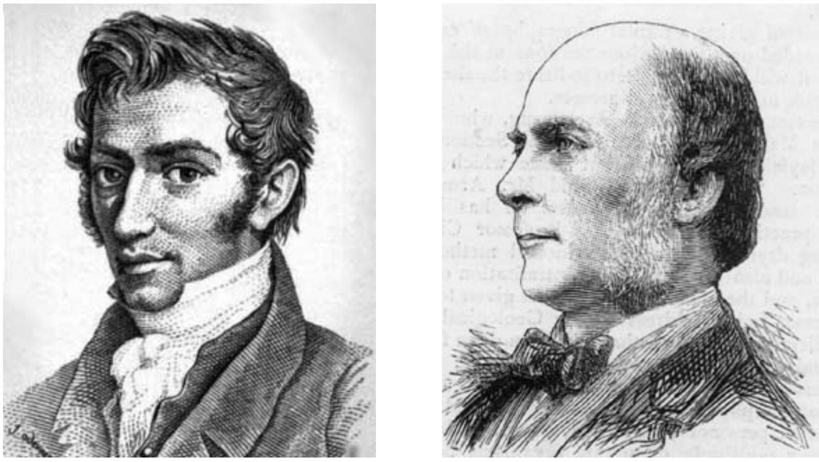

Abb. 7 Adolphe Quetelet (1796–1874) und Francis Galton (1822–1911)

Zwischen 1920 und 1930 wurde von Ronald Aylmer Fisher (1890–1962) in der landwirtschaftlichen Versuchsanstalt Rothamsted die Varianzanalyse entwickelt. Diese verwendet das gleiche lineare Modell wie im Satz von Gauß, aber da es nun um Tests geht, braucht man eine spezielle Verteilungsannahme. Fisher setzte die „Fehler" bzw. die „natürlichen Schwankungen" als normalverteilt voraus. Genau genommen handelt es sich sogar um spezielle lineare Modelle, nämlich solche, bei denen die bekannten Konstanten 0 oder 1 sind. Ich will mit zwei Beispielen andeuten, von welcher Art die Fragestellungen sind.

Einfache Varianzanalyse: Im einfachsten Fall variiert nur ein Faktor. Man misst also z.B. Ernteerträge bei der Anwendung von zwei verschiedenen Dünge-

Abb. 8 Ronald Aylmer Fisher (1890–1962)

mitteln (Klasse 1 und 2) und ohne Düngung (Klasse 3). Die Mittelwerte wären, selbst wenn es keine Unterschiede zwischen den drei Methoden gäbe, schon wegen der natürlichen Schwankungen verschieden. Wie groß müssen Unterschiede sein, damit sie nicht durch den Zufall erklärbar sind, sondern „signifikant"?

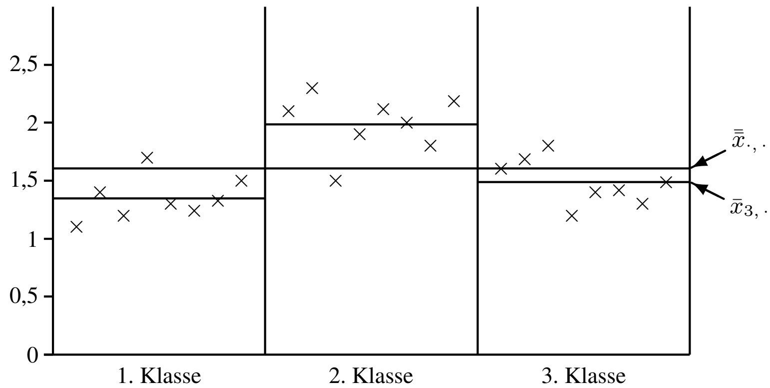

Abb. 9 Mittelwerte der Klassen und Gesamtmittelwert

Tatsächlich untersucht man auch Experimente mit Differenzierung nach mehreren verschiedenen Faktoren. Studiert man etwa die Wirkung einer Sonnencreme, so könnte man die Faktoren Hauttyp (A, B, C), Dauer der Bestrahlung, Geschlecht unterscheiden.

In jedem der leeren Kästchen wären die Ergebnisse der Messungen an mehreren Personen zu registrieren. Untersucht man mehrere Cremes, so ist die Sorte ein vierter Faktor.

Tabelle 1 Beispiel einer Varianzanalyse mit drei Faktoren

	20 Minuten			1 Stunde			2 Stunden		
	A	B	C	A	B	C	A	B	C
Männlich									
Weiblich									

9 Verallgemeinerte lineare Modelle

Aber auch ganz aktuelle Forschung setzt die Untersuchungen von Gauß fort. Seit ca. 1970 werden so genannte verallgemeinerte lineare Modelle intensiv studiert. Um den Unterschied zu erklären, erinnere ich noch einmal an die linearen Modelle. Hier werden Zufallsvariable der Form $Y_1 = \mu_1 + E_1$, $Y_2 = \mu_2 + E_2, \ldots$ mit

$$\mu_1 = \alpha\, x_{1,\,1} + \beta\, x_{1,\,2}\,,$$
$$\mu_2 = \ldots$$

beobachtet, wobei die Erwartungswerte der Messfehler $E_1,\, E_2, \ldots$ alle $= 0$ sind. Bei verallgemeinerten linearen Modellen wird dagegen nur noch angenommen, dass $Y_1 = g(\mu_1) + E_1$, $Y_2 = g(Y_2) + E_2, \ldots$ sei. Der Erwartungswert der beobachteten Größen $Y_1,\, Y_2, \ldots$ muss also nicht mehr linear sein sondern nur eine Funktion einer linearen Größe. g heißt Linkfunktion. Neben der Normalverteilung lässt man auch mehrere andere Verteilungen zu.

Das interessante daran ist, dass dieser allgemeinere Ansatz es auch erlaubt eine ganze Reihe moderner statistischer Verfahren zu beschreiben und zu untersuchen. Darunter sind auch ganz neue, aber ich möchte ein relativ einfaches Beispiel wählen, obwohl es schon älteren Datums ist.

Bei Giften interessiert man sich für die mittlere tödliche Dosis oder allgemeiner für die Dosis-Wirkungs-Kurve. 1934 hat Bliss untersucht, wie eine Nikotin-Lösung auf Blattläuse wirkt. Auf der Abzisse, also der waagerechten Koordinatenachse, wird die Konzentration im so genannten logarithmischen Maßstab aufgetragen, d.h. $x = 0$ entspricht einer festgelegten Standard-Konzentration, z.B. 1 g Nikotin gelöst in 1 Liter Wasser. $x = 1$ entspricht der doppelten Konzentration, $x = 2$ der vierfachen Konzentration (usw.) und $x = -1$ der halben Konzentration. Sprüht man die Lösung auf eine mit Blattläusen bevölkerte Pflanze, so sterben umso mehr Blattläuse je höher die Konzentration ist. Die Konzentration, bei der gerade die Hälfte umkommt, heißt mittlere tödliche Dosis. Sehr häufig haben die „Dosis-Wirkungs-Kurven“, die angeben, wie groß die Sterbewahrscheinlichkeit p in Abhängigkeit von der Konzentration x ist, die Form

$$p = \Phi(\alpha + \beta\, x)\,,$$

wobei Φ die so genannte Verteilungsfunktion der Standard-Normalverteilung ist. Man möchte die mittlere tödliche Dosis und den „Anstieg“ β bestimmen. Ist $x_{1/2}$ die mittlere tödliche Dosis, so gilt also

$$\Phi(\alpha + \beta x_{1/2}) = \tfrac{1}{2}\,.$$

Hat man α und β geschätzt, so ergibt sich hieraus eine Schätzung für $x_{1/2}$. In diesem Fall ist Φ die Linkfunktion.

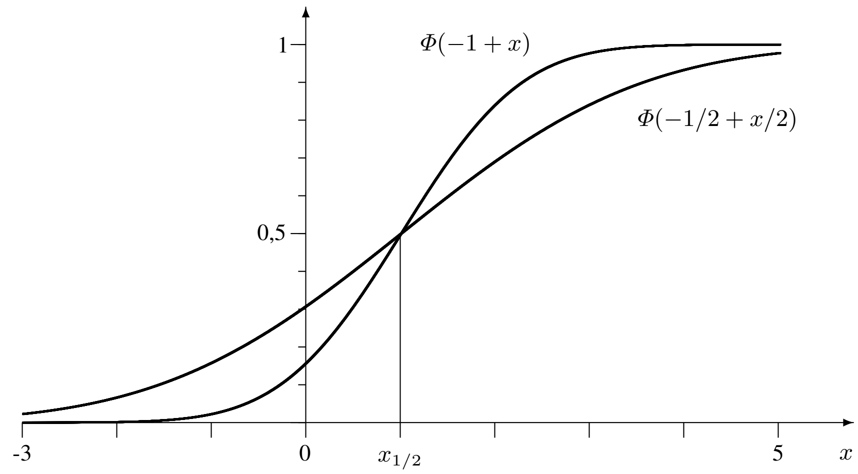

Abb. 10 Zwei Wirkungskurven mit gleicher mittlerer tödlicher Dosis $x_{1/2}$ und verschiedenem β

Auch bei verallgemeinerten linearen Modellen arbeitet man mit Maximum-Liklihood-Methoden. Während es im Satz von Gauß auf die Varianz ankommt, ist hier der Begriff „Quasi-Likelihood" zentral, und es ergeben sich Analogien. Eine Darstellung der Theorie findet man im Buch von McCullagh und Nelder.

Auch andere Ideen von Gauß wirken fort. Zum Beispiel ist die Beweismethode, die Gauß bei der Herleitung des Fehlerfortpflanzungsgesetzes verwendet hat, zur so genannten δ-Methode verallgemeinert worden. Dies wird z.B. im Buch von van der Vaart beschrieben.

Wir sehen: Gauß ist durch seine Arbeiten noch sehr lebendig.

Danksagung Ich danke den Kollegen, die mich mit nützlichen Hinweisen und Literaturangaben unterstützt haben. Hervorheben möchte ich Bradley Efron, Richard Gill, Axel Munk, Michael Odenkirchen, Pete Stewart und Stephen Stigler.

Besonders habe ich von den Büchern von Anders Hald und Stephen Stigler zur Geschichte der Statistik profitiert. Dort finden sich auch viele Literaturangaben, die hier weggelassen sind. Ich danke auch Erich Berger für viele Anregungen und Hilfen.

Literatur

1. A. Celmiņš: The method of Gauss in 1799. Stat. Sci. 13, 123–135 (1998)
2. P. Brosche und M. Odenkirchen: C. F. Gauß und die Einführung der Methode der kleinsten Quadrate. Mitt. Gauß-Ges. Gött. 33, 11–20 (1996)
3. J. Dutka: On Gauss' Priority in the Discovery of the Method of Least Squares. Arch. Hist. Exact Sci. 49, 355–370 (1996)
4. C. F. Gauß: Theory of Combination of Observations Least Subject to Errors. Part One, Part Two, Supplement. Translated by G. W. Stewart, with translators introduction and afterword. Classics in Applied Mathematics, SIAM, Philadelphia (1995) [Translation of Gauss' work of 1821, 1823 and 1828]
5. P. McCullagh and J. A. Nelder: Generalized Linear Models. Chapman and Hall, London-New York (1983)

6. K. Reich: Im Umfeld der Theoria Motus: Gauß' Briefwechsel mit Perthes, Laplace, Delambre und Legendre. Abhandl. Akad. Wissensch. Göttingen, Vandenkoek und Ruprecht, 3. Folge Nr. 18 (2001)
7. S. M. Stigler: Gauss and the invention of least squares. Ann. Statist. 9, 465–474 (1981)
8. S. M. Stigler: The History of Statistics — The Measurement of Uncertainty before 1900. The Belknap Press of Harvard Univ. Press, Cambridge MA (1981)
9. A. W. van der Vaart: Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Statistics (1998)