Gauß' erste Herleitung der Normalverteilung

Edgar M. E. Wermuth

Fakultät für Angewandte Mathematik,
Physik und Allgemeinwissenschaften
Technische Hochschule Nürnberg

Nürnberg, Regiomontanusweg 1, 23. Nov. 2023

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Übersicht

Vorbemerkung

Gauß' erste Herleitung der Normalverteilung (1809) wird manchmal als zirkulär hingestellt.

Völlig zu unrecht. Diese bestechend einfache Überlegung *sollte* auch heute noch in den Lehrbüchern stehen.

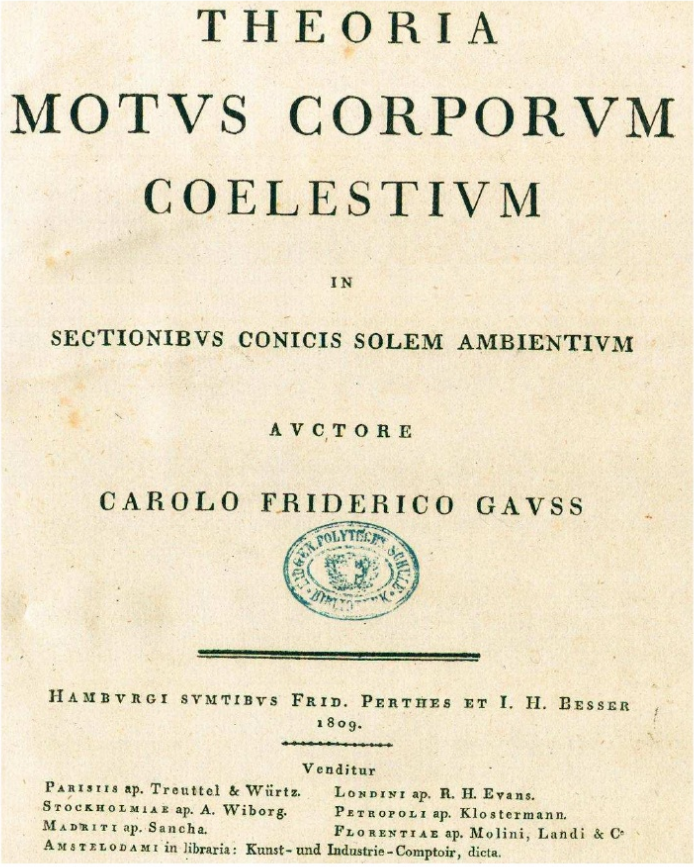
Gauß benutzte wesentlich schon die sogenannte Cauchysche Funktionalgleichung (1821).

Die Theoria Motus

Im Jahre 1807 wurde der 30-jährige Carl Friedrich Gauß (1777-1855) zum Direktor der Sternwarte Göttingen berufen, wohl auch aufgrund seines frühen Ruhms im Zusammenhang mit dem Planetoiden Ceres.

Zwei Jahre später legte er – nach den *Disquisitiones arithmeticae* (1801) – sein zweites großes Werk vor, die *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* („Theorie der Himmelskörper-Bewegungen um die Sonne längs Kegelschnitten").

Titelblatt der Theoria Motus I



THEORIA
MOTVS CORPORVM
COELESTIVM

IN

SECTIONIBVS CONICIS SOLEM AMBIENTIVM

AVCTORE

CAROLO FRIDERICO GAVSS

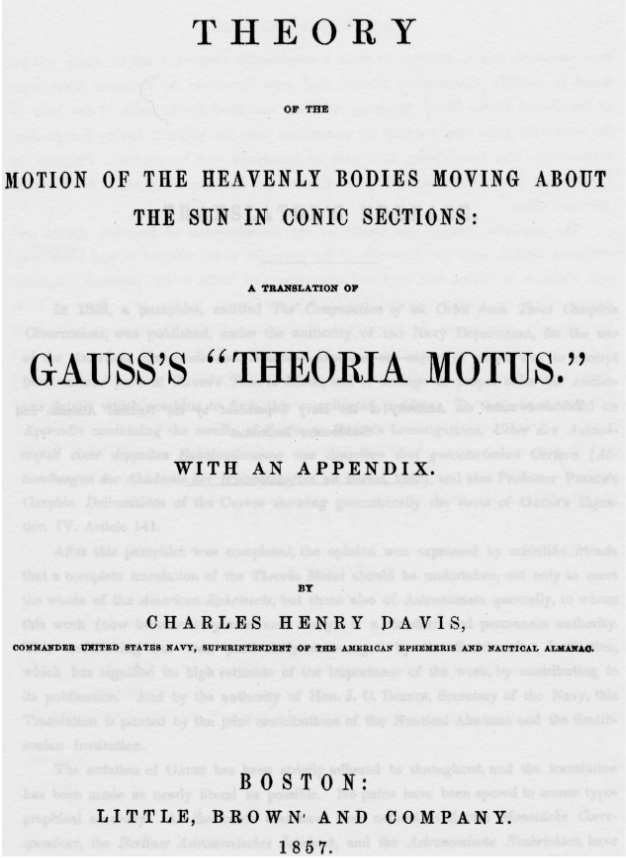
HAMBVRGI SVMTIBVS FRID. PERTHES ET I. H. BESSER
1809.

Venditur
PARISIIS ap. Treuttel & Würtz. LONDINI ap. R. H. Evans.
STOCKHOLMIAE ap. A. Wiborg. PETROPOLI ap. Klostermann.
MADRITI ap. Sancha. FLORENTIAE ap. Molini, Landi & C°
AMSTELODAMI in libraria: Kunst- und Industrie-Comptoir, dicta.

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Titelblatt der Theoria Motus II



THEORY

OF THE

MOTION OF THE HEAVENLY BODIES MOVING ABOUT
THE SUN IN CONIC SECTIONS:

A TRANSLATION OF

GAUSS'S "THEORIA MOTUS."

WITH AN APPENDIX.

BY

CHARLES HENRY DAVIS,
COMMANDER UNITED STATES NAVY, SUPERINTENDENT OF THE AMERICAN EPHEMERIS AND NAUTICAL ALMANAC.

BOSTON:
LITTLE, BROWN AND COMPANY.
1857.

Methode der kleinsten Quadrate I

Diese Methode, von Adrien-Marie Legendre (1752-1833) im Jahre 1805 („Méthode des moindres carrés“) erstmals publiziert, dient der Ausgleichung von Messfehlern, insbesondere bei astronomischen Beobachtungen.

Gauß benutzte die Methode schon *vor* Legendres Publikation bei seinen Rechnungen. In der *Theoria motus* gibt er eine erste *Begründung* für die Methode.

Legendre beschrieb die Methode zwar vorbildlich klar, gab aber *keine* Begründung.

Methode der kleinsten Quadrate II

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Worin besteht die Methode?

Definition (Methode der kleinsten Quadrate)

Gegeben gleich sorgfältige Messwerte $x_1, \ldots, x_n$ zur Bestimmung der Größen $\xi_i = f_i(p_1, \ldots, p_k)$ $(1 \leq i \leq n)$ mit den letztlich gesuchten *Parametern* (Einflussfaktoren) $p_1, \ldots, p_k$. Die p_j werden so gewählt, dass die

Fehlerquadratsumme $F(p_1, \ldots, p_k) = \sum_{i=1}^{n} \big(x_i - f_i(p_1, \ldots, p_k)\big)^2$

minimal ausfällt, also $\frac{\partial F}{\partial p_1} = \cdots = \frac{\partial F}{\partial p_k} = 0$ gilt.

Oft sind die f_i lineare Funktionen (zumindest näherungsweise). Dann ist es ein Problem der *multilinearen Regression*.

Der Maximum-Likelihood-Ansatz I

Was ist „Maximum Likelihood“?

Die beobachteten Werte $x_i = f_i(p_1, \ldots, p_k) + \delta_i \ (1 \leq i \leq n)$ zu den Funktionen f_i der gesuchten Parameter p_j enthalten zufällige *Fehler* δ_i. Aus der *Wahrscheinlichkeitsdichte* $\varphi(\delta)$ der Fehler ergibt sich als Maß für die Wahrscheinlichkeit des Auftretens der beobachteten x_i die „Likelihood-Funktion“

$$L(x_1, \ldots, x_n; p_1, \ldots, p_k) = \varphi(\delta_1) \cdots \varphi(\delta_n).$$

Dabei sei vorausgesetzt, dass $\varphi(-\delta) = \varphi(\delta)$.

Definition (Maximum-Likelihood-Prinzip)

Wähle die Parameter p_j so, dass L so groß wie möglich wird.

Der Maximum-Likelihood-Ansatz II

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Im 20. Jahrhundert erhielt die Maximum-Likelihood-Methode ihren *Namen* und wurde zu einem der bestbegründeten Hilfsmittel der Parameter-Schätzung bei Zufallsgrößen, nicht nur bei *Fehler*-Verteilungen. Bahnbrechend waren die Arbeiten von Ronald A. Fisher (1890-1962).

Zu Gauß' Zeiten galt dies noch nicht, weshalb er sorgfältige einfache Wahrscheinlichkeitsbetrachtungen zur Begründung und Erläuterung des Maximum-Likelihood-Prinzips anführt.

Er wendet es dann auf den einfachsten denkbaren Fall an, den der Bestimmung eines einzigen Parameters x durch eine Messreihe von Beobachtungswerten x_i $(1 \leq i \leq n)$.

Das arithmetische Mittel I

Gegeben eine Reihe $x_1, \ldots, x_n$ fehlerbehafteter, aber gleich sorgfältiger voneinander unabhängiger Messungen einer zu bestimmenden Größe x.

In aller Regel ist dann das arithmetische Mittel $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$ der beste Schätzwert für den *wahren Wert* der Größe x.

Davon gehe man so gut wie immer aus, konstatiert Gauß.

Kennzeichnend für das arithmetische Mittel:

$$(x_1 - \bar{x}) + \cdots + (x_n - \bar{x}) = 0$$

D.h.: Die Summe der Abweichungen vom Mittel ist 0.

Das arithmetische Mittel II

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Den Test der kleinsten Quadrate besteht das arithmetische Mittel mit Leichtigkeit:

$$\sum_{i=1}^{n}(x_i - c)^2 = \sum_{i=1}^{n}\Big((x_i - \bar{x}) + (\bar{x} - c)\Big)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - c)^2.$$

Wie sieht es aus mit der maximalen Likelihood?

Diese setzt ja eine *Fehlerdichte* voraus. Und dies nutzt Gauß für ein *inverses* Maximum-Likelihood-Argument:

Wie muss die Fehlerdichte aussehen, damit das arithmetische Mittel der Messwerte den Maximum-Likelihood-Test besteht, d.h. die „Likelihood-Funktion“ L maximiert?

Das arithmetische Mittel III

Voraussetzungen:
Die Fehlerdichte sei *symmetrisch* ($\varphi(-\delta) = \varphi(\delta)$) und *positiv* in der Umgebung von 0, ferner *stetig differenzierbar*.

Wenn dies auch praktisch nicht ganz exakt gelten muss, so Gauß, könne man doch *näherungsweise* davon ausgehen.

Die Maximumsstelle von $f(x) = \varphi(x_1 - x) \cdots \varphi(x_n - x)$ ist auch Maximumsstelle von $\ln f(x) = \ln \varphi(x_1 - x) + \cdots + \ln \varphi(x_n - x)$. Also verschwindet die Ableitung; d.h.

$$\frac{\varphi'(x_1 - x)}{\varphi(x_1 - x)} + \cdots + \frac{\varphi'(x_n - x)}{\varphi(x_n - x)} = 0.$$

Herleitung der Normalverteilung
Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Das arithmetische Mittel IV

Wenn die Likelihood stets durch das arithmetische Mittel der Werte x_i maximiert wird, *muss* also gelten:

$$\sum_{i=1}^{n}(x_i - x) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \frac{\varphi'(x_i - x)}{\varphi(x_i - x)} = 0.$$

Mit $\psi(u) := \varphi'(u)/\varphi(u)$ also

$$u_1 + \cdots + u_n = 0 \quad \Rightarrow \quad \psi(u_1) + \cdots + \psi(u_n) = 0.$$

Welche stetigen Funktionen ψ haben diese Eigenschaft?

Wählt man $u_2 = \cdots = u_n = u$ und $u_1 = (1 - n)u$, ergibt sich $\psi((1 - n)u) = (1 - n)\psi(u)$.

Hieraus ist ersichtlich, dass $\psi(u)/u$ konstant ist, sagt Gauß.

Das arithmetische Mittel V

In der Tat:

Mit $n = 2$ folgt $\psi(-u) = -\psi(u)$, also auch $\psi(0) = 0$ und damit allgemein

$$\psi(ku) = k\psi(u) \ (k \in \mathbb{Z}), \quad \psi(ru) = r\psi(u) \ (r \in \mathbb{Q}).$$

Mit $\psi(ru)/ru = \psi(u)/u$ ist aber $\psi(u)/u$ konstant. $\square$

Dies impliziert $\varphi'(x) = kx\varphi(x)$, zumindest für hinreichend kleine $|x|$. Da es keinen Sinn macht, dass die Fehlerdichte für positive x wächst, muss $k < 0$ gelten, also etwa $k = -2h^2$. Damit folgt sofort $\varphi(x) = c\,\mathrm{e}^{-h^2x^2}$. Mit $\int_{\mathbb{R}} \varphi = 1$ also

$$\boldsymbol{\varphi(x) = \frac{h}{\sqrt{\pi}}\mathrm{e}^{-h^2x^2}} \quad \text{wegen} \quad \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\mathrm{d}x = \sqrt{\pi}.$$

Das Laplacesche Integral*

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Das den Normierungsfaktor der Dichte ergebende uneigentliche Integral hat in dieser Form als erster wohl Pierre Simon de Laplace (1749-1827) angegeben, obwohl es 1781 in etwas anderer Gestalt schon bei Euler vorkommt:

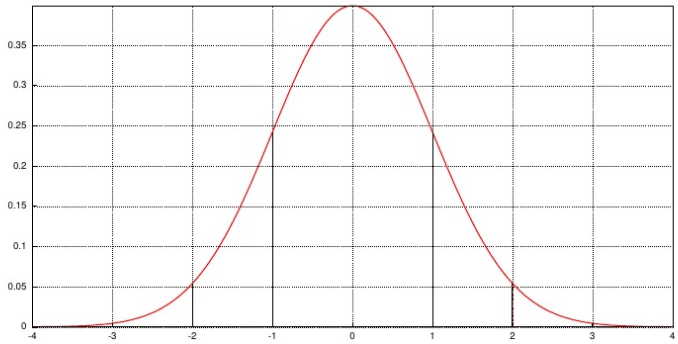
$$\int_0^{\infty} \mathrm{e}^{-ax}\,\frac{\cos bx}{\sqrt{x}}\mathrm{d}x = \sqrt{\frac{\pi(a+c)}{2c^2}}$$

mit $a \geq 0$, $c = \sqrt{a^2 + b^2} > 0$. Setzt man $a = 1, b = 0$ und substituiert $t = \sqrt{x}$, folgt das Laplacesche Integral.
Der eleganteste Beweis ist der *zweidimensionale* mittels Übergang zu Polarkoordinaten und stammt von Gauß:

$$\iint_{\mathbb{R}^2} \mathrm{e}^{-x^2-y^2}\mathrm{d}x\mathrm{d}y = \int_0^{2\pi}\int_0^{\infty} \mathrm{e}^{-r^2} r\mathrm{d}r\mathrm{d}\varphi\,.$$

Die Normalverteilungsdichte I

Gauß merkt an, dass bei der erhaltenen Dichtefunktion $\varphi(x) = \frac{h}{\sqrt{\pi}}\,\mathrm{e}^{-h^2x^2}$ die Konstante h ein Maß der *Messgenauigkeit* der Beobachtungen ist. Er betont außerdem, dass diese analytische Funktion nie ganz exakt richtig sein könne, da ja nicht beliebig große Fehler möglich seien; aber praktisch spiele das keine Rolle. In der Grafik ist $h = 1/\sqrt{2}$; $1/\sqrt{2\pi} \approx 0.399$. Heute heißt $\sigma^2 = \frac{1}{2h^2}$ *Varianz* der Dichte.



Die Normalverteilungsdichte II

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Satz (Arithmetisches Mittel als ML-Schätzwert)

Bei einer Reihe $x_1, \ldots, x_n$ unabhängiger gleich sorgfältiger Messungen zur Bestimmung einer Größe x ist genau dann stets $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ der Höchstwahrscheinlichkeits-Schätzwert für x, wenn die Fehlerdichte jeweils vom Typ $\varphi(x) = \frac{h}{\sqrt{\pi}}\,\mathrm{e}^{-h^2x^2}$ ist. Dabei ist h ein Maß für die Genauigkeit der Messungen.

Damit ist die Kleinste-Quadrate-Methode leicht zu begründen: Sind $x_1, \ldots, x_n$ fehlerbehaftete Messwerte zu den $f_i(p_1, \ldots, p_k)$, ist die Likelihood-Funktion (mit konstantem c)

$$L(x_1, \ldots, x_n; p_1, \ldots, p_k) = c\,\mathrm{e}^{-h^2 \sum_{i=1}^{n} (x_i - f_i(p_1, \ldots, p_k))^2}$$

maximal bei minimaler Fehlerquadratsumme. Bei bekannten unterschiedlichen x_i-Messgenauigkeiten ist eine *gewichtete* Fehlerquadratsumme zu betrachten, betont Gauß.

Nichtlineare Modelle

Gauß skizziert auch kurz eine geeignete Vorgehensweise im Falle *nichtlinearer* Einflussfunktionen $f_i(p_1, \ldots, p_k)$; dann ergeben ja die Gleichungen $\frac{\partial F}{\partial p_1} = \cdots = \frac{\partial F}{\partial p_k} = 0$ zur Minimierung der Fehlerquadratsumme ein nichtlineares System.

Ausgehend von einer Näherungslösung $(p_{1,0}, \ldots, p_{k,0})$ des Minimierungsproblems (z.B. beschafft durch Lösen des Systems $x_i = f_i(p_1, \ldots, p_k)$ $(1 \leq i \leq k)$), löst man ein lineares Problem mit den angenäherten Einflussfunktionen $f_i^*(p_1^*, \ldots, p_k^*) = f_i(p_{1,0}, \ldots, p_{k,0}) + \nabla f_i(p_{1,0}, \ldots, p_{k,0}) \cdot (p_1^*, \ldots, p_k^*)$ $(1 \leq i \leq n)$ (Taylorentwicklung erster Ordnung) und erhält verbesserte Werte $p_{j,1} = p_{j,0} + p_j^*$. Sind die Korrekturen noch zu groß, muss das Verfahren wiederholt werden.

Dies ist das *Gauß-Newton-Verfahren*.

Reaktion von Laplace I

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Pierre Simon de Laplace (1749-1827) trug jahrzehntelang Bahnbrechendes zur Wahrscheinlichkeitsrechnung bei. Insofern hatte sein Urteil großes Gewicht. Er kritisierte vor allem zweierlei:

1) Man müsse auch *andere* Fehler-Verteilungen als die Normalverteilung berücksichtigen.

2) Nicht größtmogliche Wahrscheinlichkeit sei das relevante Qualitätskriterium, sondern *im Durchschnitt möglichst kleine Abweichung* der Schätzung vom wahren Wert.

Außerdem zog er ein *anderes Fehlermaß* vor: Die Summe der *Absolutwerte* der Einzelabweichungen anstelle der Summe der *Quadrate*.

Reaktion von Laplace II

Laplace gab als Alternative zu Gauß eine „asymptotische“ Begründung der kleinsten Quadrate im linearen Fall.

In seiner *Théorie Analytique des Probabilités* (erste Auflage 1812) zeigte er: Bei *sehr großer* Beobachtungs-Zahl (mit beliebigen, aber unabhängigen gleichartigen Einzelfehlern) ist die Verteilung der Summen der absoluten wie der quadrierten Fehler annähernd *normal*, und kleinste Quadrate führen zum kleinsten mittleren absoluten Fehler.

Die Laplaceschen Resultate – nicht ganz rigoros bewiesen, aber dennoch analytisch beeindruckend – waren *ein wesentlicher Schritt in Richtung des allgemeinen zentralen Grenzwertsatzes*, über den „Satz von de Moivre/Laplace“ weit hinausgehend, den er auch kurz zuvor erstmals allgemein bewiesen hatte.

De Moivre / Laplace

Die Wahrscheinlichkeitsgewichte der Binomialverteilung streben mit wachsendem n gegen eine Normalverteilung.

Abraham de Moivre zeigte 1738 für $p = 1/2$ und Pierre Simon de Laplace 1810/12 für allgemeines $p \in (0, 1)$:

Satz

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \mathrm{e}^{-\frac{(np-k)^2}{2np(1-p)}} \quad (n \to \infty),$$

wenn beim Grenzübergang $|np - k| \leq c\sqrt{n}$ *eingehalten wird.*

$np = \mu, \quad np(1-p) = \sigma^2.$

Gauß ist Wegbereiter des Varianz-Begriffs

Gauß hielt die Laplaceschen Einwände in einigen Punkten für berechtigt und sah seine erste Kleinste-Quadrate-Begründung später als zwar relevant, aber insgesamt nicht ausreichend an. Immerhin wies Friedrich Wilhelm Bessel (1748-1846) empirisch nach: Astronomische Beobachtungsfehler sind zumeist *tatsächlich annähernd normalverteilt*.

Am *quadratischen* Fehlermaß statt des absoluten hielt Gauß fest und rechtfertigte dies unter anderem ganz anschaulich:

Zweimal einen einfachen Fehler lasse man sich lieber gefallen als einmal einen doppelten.

Vor allem bewiesen seine abschließenden Untersuchungen zur Fehlerrechnung, gut zehn Jahre später publiziert, die bessere Tauglichkeit *seines* Fehlermaßes. – Der Begriff *Varianz* ging aus Gauß' „mittlerem zu befürchtenden Fehler" hervor.

Lineare Modelle

Herleitung der Normalverteilung
Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Bei der *linearen Regression* geht es um „Fit-Funktionen" des Typs $\boldsymbol{y = c_0\, f_0(x) + \cdots + c_k\, f_k(x)}$, die an beobachtete Werte y_i an den Stellen x_i mit $1 \leq i \leq n$ anzupassen sind.

Bei der *multilinearen Regression* werden Funktionen $\boldsymbol{y = c_0 + c_1\, x_1 + \cdots + c_k\, x_k}$ an beobachtete Werte y_i an Stellen $(x_{1,i}, \ldots, x_{k,i})$ mit $1 \leq i \leq n$ angepasst.

Beide Fälle führen auf die Näherungslösung eines überbestimmten Gleichungssystems $A\mathbf{x} = \mathbf{b}$, $A \in \mathbb{R}^{n\times(k+1)}$, $\mathbf{b} \in \mathbb{R}^n$. Linear unabhängige Spalten von A vorausgesetzt (anderenfalls: Modell redundant), führt die Methode der kleinsten Quadrate zur Lösung $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$ („Gaußsche Normalgleichungen").

Theoria Combinationis (Gauß 1823)

Die *Theoria combinationis observationum erroribus minimis obnoxiae* vollendet Gauß' Fehlertheorie, schafft bis heute gültige Grundlagen der Fehlerrechnung. Zwei Haupt-Resultate:

Satz (Kleinste Quadrate sind optimal, 1821)

Der Kleinste-Quadrate-Schätzer der Parameter eines linearen Modells hat einen kleineren mittleren quadratischen Fehler als jeder andere lineare verzerrungsfreie Schätzer, unabhängig vom Verteilungsgesetz der Fehler und von der Anzahl der Daten.

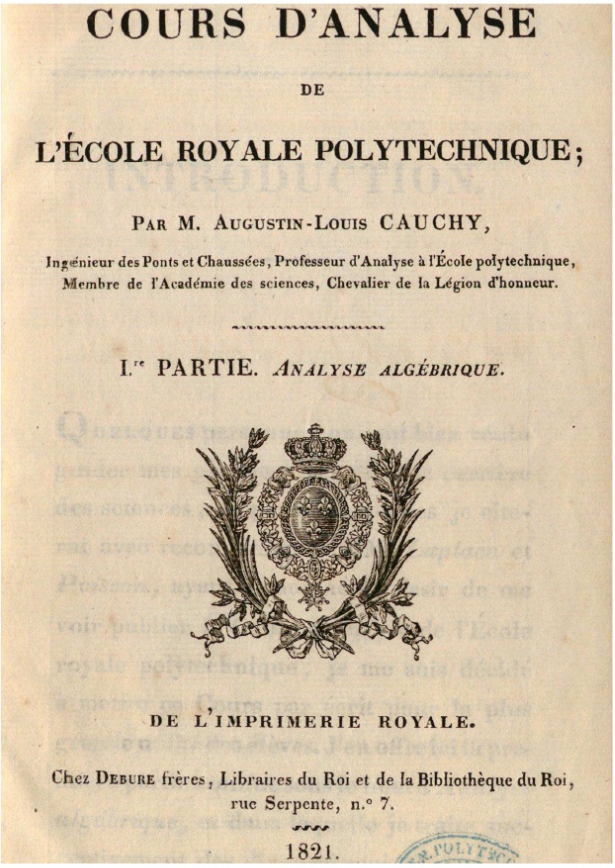
Satz (A-posteriori-Fehlerschätzung, 1823)

Verzerrungsfreier Schätzer des mittleren quadratischen Fehlers σ^2 der Beobachtungen $b_1, \ldots, b_n$ bei durch kleinste Quadrate bestimmten Einflussfaktoren $x_1, \ldots, x_k$ ist $\frac{1}{n-k}|\mathbf{b} - A\mathbf{x}|^2$.

Titel der *Analyse algébrique*

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy



COURS D'ANALYSE

DE

L'ÉCOLE ROYALE POLYTECHNIQUE;

PAR M. AUGUSTIN-LOUIS CAUCHY,

Ingénieur des Ponts et Chaussées, Professeur d'Analyse à l'École polytechnique, Membre de l'Académie des sciences, Chevalier de la Légion d'honneur.

I.re PARTIE. *ANALYSE ALGÉBRIQUE.*

DE L'IMPRIMERIE ROYALE.

Chez DEBURE frères, Libraires du Roi et de la Bibliothèque du Roi, rue Serpente, n.° 7.

1821.

Cauchys Funktionalgleichung I

Es ist leicht zu sehen, dass die Funktionalgleichung

$$f(x+y) = f(x) + f(y) \quad (x, y \in \mathbb{R}) \qquad (\star)$$

äquivalent ist zur von Gauß betrachteten Beziehung

$$x_1 + \cdots + x_n = 0 \;\Rightarrow\; f(x_1) + \cdots + f(x_n) = 0 \quad (n \in \mathbb{N},\ x_i \in \mathbb{R}).$$

Satz (Cauchy (1821))

Die einzigen stetigen Lösungen von $(\star)$ sind die Funktionen $f(x) = f(1)\,x$.

Es reicht, Stetigkeit *in einem einzigen Punkt* vorauszusetzen.

Cauchys Funktionalgleichung II

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Satz (Lösungen der Cauchyschen Funktionalgleichung)

a) Alle Lebesgue-messbaren Lösungen von $(\star)$ sind stetig.

b) Der Graph einer unstetigen Lösung von $(\star)$ liegt dicht im $\mathbb{R}^2$.

c) (A.M. Ostrowski, 1929) Sei f eine Lösung von $(\star)$. Gibt es eine Menge $M \subseteq \mathbb{R}$ positiven Lebesgue-Maßes, so daß $f(M) \cap (a, b) = \emptyset$ für ein Intervall (a, b), so ist f stetig.

Beweis zu b):
Zu einer unstetigen Lösung f von $(\star)$ gibt es Zahlen $a, b \in \mathbb{R} \setminus \{0\}$ mit $f(a)/a \neq f(b)/b$, also $b\,f(a) - a\,f(b) \neq 0$. Lineare Systeme $a\,x + b\,y = \alpha,\ f(a)\,x + f(b)\,y = \beta$ sind somit stets eindeutig lösbar, und zu $\varepsilon > 0$ gibt's *rationale* Zahlen ξ, η nahe x bzw. nahe y mit $|a\xi + b\eta - \alpha| < \varepsilon,\ |f(a)\xi + f(b)\eta - \beta| < \varepsilon$. Die Behauptung folgt nun wegen $f(a)\xi + f(b)\eta = f(a\xi) + f(b\eta) = f(a\xi + b\eta)$. $\square$

Cauchys Funktionalgleichung III

Gibt es überhaupt *unstetige* Lösungen von $(\star)$?

Diese Frage beantwortete erst 1905 – lange nach Gauß und Cauchy – Georg Hamel in seiner berühmten Arbeit *Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung:* $f(x+y) = f(x) + f(y)$.

$\mathbb{R}$ sei betrachtet als Vektorraum über $\mathbb{Q}$. Ausgehend von einer *Wohlordnungsrelation* $\prec$ auf $\mathbb{R}$, ist

$$\mathcal{B} := \{b \in \mathbb{R} \mid b \text{ ist nicht } \mathbb{Q}\text{-LK von } \prec\text{-Vorgängern von } b\}$$

algebraische Basis dieses Vektorraumes; dabei sei eine „$\mathbb{Q}$-LK“ eine endliche Linearkombination mit Koeffizienten aus $\mathbb{Q}$. Durch beliebige Wahl der Werte $f(b)$ mit $b \in \mathcal{B}$ und lineare Fortsetzung gemäß $f(x) = r_1\, f(b_1) + \cdots + r_n\, f(b_n)$ für $x = r_1\, b_1 + \cdots + r_n\, b_n$ erhält man *alle* Lösungen von $(\star)$.

Cauchys Funktionalgleichung IV

Herleitung der Normalverteilung

Vorbemerkung
Theoria Motus
Least Squares
Max.Likelihood
Arithm. Mittel
Normalvertlg
Laplace
Theoria Combi
Cauchy

Funktionalgleichung der wichtigsten mathematischen Funktion:

$$g(x+y) = g(x)\, g(y) \quad (x,\, y \in \mathbb{R}). \qquad (\star\star)$$

Gilt $g(x_0) = 0$, folgt $g(x) = g(x_0)\, g(x - x_0) = 0$ für alle x; ferner $g(x) = g(x/2)\, g(x/2) \geq 0$ für alle x. Also sind alle Lösungen außer der trivialen ($g \equiv 0$) durchweg positiv, und $f(x) := \ln g(x)$ ist eine Lösung der Cauchyschen Gleichung.

Da es total unstetige Lösungen von $(\star)$ gibt, charakterisiert $(\star\star)$ allein *nicht* die Exponentialfunktion.

Überraschend einfach kann man hingegen nachweisen:

$$g(x+y) \geq g(x)(1+y)\ (x, x+y \geq 0) \ \Leftrightarrow\ g(x) = c\,\mathrm{e}^x\ (x \geq 0),\ c \geq 0.$$

(Diese Funktionalungleichung drückt die *Konvexität* von g aus.)

Originalarbeiten

- Carl Friedrich Gauß, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, Hamburg 1809; auch: *Werke*, Bd. 7 („Theorie der Himmelskörper-Bewegungen um die Sonne längs Kegelschnitten“)
- –, *Theory of the Motion of Heavenly Bodies Moving About the Sun in Conic Sections*, Boston 1857 (Nachdruck: The Michigan Historical Reprint Series)
- –, *Theoria combinationis observationum erroribus minimis obnoxiae*, Göttingen 1823, 1828; auch: *Werke*, Bd. 4 („Theorie der am wenigsten fehlerbehafteten Kombination von Beobachtungen“)
- Augustin-Louis Cauchy, *Analyse algébrique*, Paris 1821