Ein Gaußsches inverses Maximum-Likelihood-Argument

©Edgar M. E. Wermuth, Fakultät AMP, TH Nürnberg

Begleit-Skriptum zum Vortrag an der Regiomontanus-Sternwarte, 23. Nov. 2023

Zusammenfassung

Gauß' erste Herleitung der Normalverteilung (1809) wird manchmal als zirkulär hingestellt,[1] meines Erachtens völlig zu unrecht. Im Gegenteil sollte diese bestechend einfache Überlegung auch heute noch in den Lehrbüchern stehen. Gauß benutzte wesentlich schon die sogenannte Cauchysche Funktionalgleichung (1821). Dies scheint nicht allgemein bekannt zu sein.[2]
Keywords: Normal distribution, least squares, maximum likelihood, Cauchy's functional equation.

1 Vorbemerkung

Im dritten Abschnitt des zweiten Buchs seiner *Theoria motus*[3] gibt Carl Friederich Gauß eine erste Begründung der Methode der Kleinsten Quadrate, eines wesentlichen Hilfsmittels zur Minimierung des Einflusses von Messfehlern bei astronomischen Beobachtungen. Er erkannte dabei als erster die allgemeine Rolle der sogenannten Normalverteilung, deren Dichte daher auch Gaußsche Glockenkurve genannt wird. Seine raffiniert einfache und doch bahnbrechende Herleitung der Normalverteilung verdiente es, allgemeiner bekannt zu sein.

2 Gauß' Herleitung der Normalverteilung

Durch n unabhängige gleich genaue Messungen zur Bestimmung einer Größe x habe man die Messwerte $x_1, \ldots, x_n$ erhalten.

Ist φ die Wahrscheinlichkeitsdichte der Messfehler, so ist zur Bestimmung des Maximum-Likelihood-Schätzers von x die Funktion

$$L(x) = \varphi(x_1 - x)\varphi(x_2 - x)\cdots\varphi(x_n - x)$$

zu maximieren. Mit $\tilde{\varphi} := \varphi'/\varphi$ ergibt sich nach Logarithmieren durch Ableiten nach x die Bestimmungsgleichung für den ML-Schätzer:

$$\tilde{\varphi}(x_1 - x) + \tilde{\varphi}(x_2 - x) + \cdots + \tilde{\varphi}(x_n - x) = 0\,.$$

Wenn das arithmetische Mittel $\bar{x}$ der beste Schätzer bei gleich genauen Einzelmessungen ist (wovon, betonte Gauß, man allgemein stets ausgehe), sollte es mit dem ML-Schätzer übereinstimmen, also die gerade aufgestellte Gleichung für $x = \bar{x}$ erfüllt sein.

Das arithmetische Mittel $\bar{x}$ ist charakterisiert durch

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) = 0\,.$$

[1] So z.B. bei Stigler [15], p. 141. Hald [9], p. 57ff., gibt eine angemessenere Darstellung.
[2] Nachträgliche Recherche ergab, dass es bei Aczél [1] erwähnt wird (p. 24 und p. 44), wenn auch etwas zu unpräzise.
[3] Siehe [4], Artikel 175–179; siehe auch [5]. Deutsche Übersetzung: [14], p. 237–243.

Gauß bemerkte nun:[4]

Gilt $\sum\limits_{k=1}^{n} u_k = 0 \;\Rightarrow\; \sum\limits_{k=1}^{n} f(u_k) = 0$ für eine stetige Funktion f, ist $\dfrac{f(u)}{u}$ konstant.

Beweis:
Mit untereinander gleichen Werten $u_2 = u_3 = \cdots = u_n = u$, so dass also $u_1 = (1-n)u$, folgt $f((1-n)u) = -(n-1)f(u)$, insbesondere (Fall $n = 2$) $f(-u) = -f(u)$ und damit $f(ku) = kf(u)$ $(k \in \mathbb{Z})$, also $f(ru) = rf(u)$ $(r \in \mathbb{Q})$, und $\frac{f(ru)}{ru} = \frac{f(u)}{u}$ ist konstant. ■

Dies aber bedeutet $\tilde{\varphi}(x) = ax$ (a eine Konstante), also $\varphi'(x) = ax\varphi(x)$ und $a < 0$, d.h. $a = -2h^2$ und damit $\varphi(x) = C\mathrm{e}^{-h^2x^2}$. Da das Gesamtintegral über φ den Wert 1 ergeben muss (Fehler-Wahrscheinlichkeitsdichte!), folgt mit dem Laplaceschen Integral[5] $\int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\mathrm{d}x = \sqrt{\pi}$, dass $C = h/\sqrt{\pi}$.

Fazit: Die Fehlerdichte ist $\boldsymbol{\varphi(x) = \frac{h}{\sqrt{\pi}}\,\mathrm{e}^{-h^2x^2}}$, die *Normalverteilung* mit Mittelwert $\mu = 0$.

Gauß merkte an, h entspreche der *Messgenauigkeit.* Er betonte, diese Fehlerdichte könne natürlich nicht *ganz exakt* richtig sein, da beliebig große Fehler unmöglich sind; aber praktisch ausreichend genau entspreche dem ihr starkes Abfallen.

Mit der Annahme annähernd normalverteilter Messfehler konnte nun Gauß die Kleinste-Quadrate-Methode durch ein Maximum-Likelihood-Argument ganz einfach begründen:
Ausgehend von Messwerten $x_1, \ldots, x_n$ gleicher Genauigkeit zur Bestimmung von Größen ξ_i, die irgendwelche Funktionen $\xi_i = f_i(p_1, \ldots, p_k)$ $(1 \le i \le n)$ der letztlich zu bestimmenden Parameter p_j $(1 \le j \le k)$ sind, ist nach dem Maximum-Likelihood-Prinzip die Funktion

$$L(p_1, \ldots, p_k) = \varphi\big(x_1 - f_1(p_1, \ldots, p_n)\big)\cdots\varphi\big(x_n - f_n(p_1, \ldots, p_k)\big) = c\mathrm{e}^{-h^2\sum_{i=1}^{n}(x_i - f_i(p_1,\ldots,p_n))^2}$$

(c ein konstanter Faktor) zu maximieren und somit $\big(x_1 - f_1(p_1, \ldots, p_n)\big)^2 + \cdots + \big(x_n - f_n(p_1, \ldots, p_n)\big)^2$, die Fehlerquadratsumme, zu *minimieren.*

Gauß hob noch hervor, dass man im Falle bekannter *ungleicher* Messgenauigkeiten durch Skalierung gleiche Messgenauigkeit erzielen könne und dann eine dementsprechend *gewichtete* Fehlerquadratsumme minimieren müsse.

3 Ergänzende Bemerkungen

Die erste Darstellung der Methode der Kleinsten Quadrate gab 1805 der bedeutende französische Mathematiker Adrien-Marie Legendre; siehe [13]. Er prägte auch schon den Begriff „Kleinste Quadrate" (méthode des moindres carrés).[6]

Berühmt-berüchtigt ist sein Prioritätsstreit mit Gauß. Dieser erwähnte in seiner *Theoria motus* Legendre nur ganz am Rande und betonte, die Methode schon seit ca. 1795 sozusagen privat in seinen Rechnungen verwendet zu haben („principium nostrum", siehe Artikel 186). Legendres Werk machte die Methode erstmals publik, stellte sie sehr klar dar, lieferte aber, anders als Gauß, keine theoretische Rechtfertigung.

Gauß' Überlegungen fanden insbesondere Beachtung bei Pierre Simon Laplace, der wenig später den von Abraham de Moivre ca. 80 Jahre früher im Spezialfall $p = 1/2$ bewiesenen Grenzwertsatz auf alle $p \in (0, 1)$ ausdehnte: die Normalverteilung ergibt sich als Grenzfall $n \to \infty$ bei diskreten Binomial-

[4] Dies ist offenbar äquivalent mit „Cauchys Funktionalgleichung" $f(x+y) = f(x) + f(y)$; siehe [3], p. 104ff. Die Existenz *unstetiger* Lösungen dieser Gleichung wurde erst viel später mittels des Auswahlaxioms durch Georg Hamel bewiesen [10]; siehe auch Hille [11], p. 418ff.
[5] Das in etwas anderer Gestalt aber schon 1781 von Euler bestimmt wurde: $\int_0^\infty \exp(-ax)\,\cos(bx)\,\mathrm{d}x/\sqrt{x} = \sqrt{\pi(a+c)/2c^2}$ mit $a \ge 0,\ c = \sqrt{a^2+b^2} > 0$, da $\int_0^\infty x^{b-1}\exp(-ax)\,\mathrm{d}x = \Gamma(b)/a^b$ für $a > 0$, $b > 0$, also auch (Holomorphie!) für $\Re a > 0$, $\Re b > 0$; man substitutiere $t = \sqrt{x}$ bei passenden a, b. Der heute übliche elegante und kurze Beweis mittels Transformation auf Polarkoordinaten wird manchmal Riemann zugeschrieben, scheint aber wohl schon auf Gauß zurückzugehen.
[6] Deutsche Übersetzung des Anhangs über die Kleinste-Quadrate-Methode: [14], p. 234–236.

wahrscheinlichkeiten.[7] Laplace kritisierte die Gaußsche Begründung der Kleinste-Quadrate-Methode in zweierlei Hinsicht: Erstens müsse man auch *andere* Fehlerverteilungen berücksichtigen, und zweitens komme es weniger auf Maximal-Wahrscheinlichkeit, sondern vor allem auf durchschnittlich möglichst geringe Fehler der Parameter-Schätzung an.[8]

Gauß hielt die Laplacesche Kritik für nicht ganz unberechtigt. Aber erst mehr als zehn Jahre später fand er Gelegenheit, seine Theorie abzurunden und die bis heute gültigen Grundlagen der Fehlerrechnung zu schaffen, und zwar in seiner *Theoria combinationis* [6]; siehe auch [7].[9]

Friedrich Wilhelm Bessel wies nach, dass astronomische Beobachtungsfehler tatsächlich meist normalverteilt sind; siehe [9], p. 58, und [14], p. 280–292.

In der *Theoria combinationis* zeigte Gauß (Art. 19–21), dass unter allen unverzerrten linearen Parameter-Schätzern der Kleinste-Quadrate-Schätzer den kleinsten durchschnittlichen Fehler aufweist. Historisch *falsch* wird dieser Gaußsche Kleinste-Quadrate-Satz manchmal als „Gauß-Markoff-Theorem" bezeichnet.

Als Fehlermaß benutzte Gauß die Fehlerquadratsumme, die er ausdrücklich der von Laplace präferierten Summe der Absolutfehler vorzog und auch mit Hinweis auf die erzielten Resultate rechtfertigte. Die *Theoria combinationis* etablierte also auch den wahrscheinlichkeitstheoretischen Grundbegriff der *Varianz*.

Neben dem bekannten Gaußschen Fehlerfortpflanzugsgesetz[10] (Art. 18) behandelte die *Theoria combinationis* auch die Frage der A-Posteriori-Schätzung mittlerer Fehler (Art. 37f.): Gleich genaue Beobachtungen b_i der Größen $f_i(p_1, \ldots, p_k)$ $(1 \le i \le n)$ vorausgesetzt, setzt man die Kleinste-Quadrate-Schätzwerte p_j^* der Parameter p_j in die Funktionen ein und erhält dann mit $\frac{1}{n-k}\sum_{i=1}^n \big(b_i - f_i(p_1^*, \ldots, p_k^*)\big)^2$ den erwartungstreuen Schätzer des mittleren quadratischen Fehlers. Die empirische Varianz $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ ist ein einfacher Spezialfall.

Im folgenden *Anhang* sind die zwei erwähnten wichtigen Aussagen der *Theoria combinationis* in moderner Form bewiesen; mit Begriffen der Matrizenrechnung ist dies in sehr kompakter Form möglich. Außerdem sind anhand zweier Beispiele die Least-Squares- und Maximum-Likelihood-Methode einander gegenübergestellt und wird ein sehr kurzer Beweis des Satzes von de Moivre und Laplace formuliert.

A Beweis des Kleinste-Quadrate-Satzes

Die Beobachtungsgrößen b_i $(1 \le i \le n)$ seien *lineare* Funktionen der Parameter x_j $(1 \le j \le k)$, also $\mathbf{b} = A\mathbf{x}$ mit $\mathbf{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times k}$. Die Spalten von A seien linear unabhängig. Die b_i enthalten *voneinander unabhängige* Beobachtungsfehler ε_i mit Mittelwert 0 und *demselben* mittleren Fehler σ. Der Vektor $\mathbf{x}_0$ der wahren Werte der Parameter erfüllt also die Gleichung $\mathbf{b} = A\mathbf{x}_0 + \boldsymbol{\varepsilon}$.

Die Kleinste-Quadrate-Lösung der Aufgabe $|\mathbf{b} - A\mathbf{x}| = \min$ ist gegeben durch $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} =: M\mathbf{b}$. Wegen $MA = E$ also $\mathbf{x} - \mathbf{x}_0 = M\mathbf{b} - MA\mathbf{x}_0 = M\boldsymbol{\varepsilon}$.

Der Kleinste-Quadrate-Schätzer ist somit linear und verzerrungsfrei, und sein mittlerer quadratischer Fehler ist $\mathrm{E}(|\mathbf{x} - \mathbf{x}_0|^2) = \mathrm{E}(|\boldsymbol{\varepsilon}^T M^T M\boldsymbol{\varepsilon}|)$. Mit $M = (m_{ji})_{\substack{1\le j\le k\\ 1\le i\le n}}$ folgt also $\mathrm{E}(|\mathbf{x} - \mathbf{x}_0|^2) = \mathrm{E}\big(\sum_j(\sum_i m_{ji}\varepsilon_i)^2\big) = \sigma^2\sum_{j,i} m_{ji}^2$, da ja $\mathrm{E}(\varepsilon_i\varepsilon_j) = 0$ für $i \ne j$. Man beachte: $\sum_{j,i} m_{ji}^2 = \mathrm{Spur}\, M^TM = \mathrm{Spur}\, MM^T$.

[7] Der Laplacesche Beweis findet sich auch in seinem erstmals 1812 erschienenen Buch [12], einem der einflussreichsten Werke in der Geschichte der Wahrscheinlichkeitsrechnung.

[8] Hinsichtlich des Prioritätsstreites ist bemerkenswert, dass Laplace im zitierten Werk ausdrücklich feststellte, man müsse Gauß zubilligen, unabhängig von Legendre, dem Erst-Publizierenden, ebenfalls die Methode der Kleinsten Quadrate entdeckt zu haben (Livre II, Ende Chap. IV; [1]1812, p. 347; Oeuvres, tome 7, p. 353). Es wäre aber wohl diplomatisch vorteilhafter gewesen, wenn Gauß sich etwas anerkennender öffentlich zu Legendres Erstpublikation geäußert hätte.

[9] Dass es so lange dauerte, liegt an den vielen anderen Arbeiten, in die Gauß involviert war. Seine anstrengenden Landvermessungsaktivitäten zeitigten zudem noch andere bedeutende mathematische Früchte: 1827 publizierte er seine *Disquisitiones generales circa superficies curvas*, Grund-Werk und Ausgangspunkt der modernen Differenzialgeometrie.

[10] Sind $x_1, \ldots, x_n$ unabhängige Messgrößen, $\sigma_1^2, \ldots, \sigma_n^2$ ihre mittleren quadratischen Fehler, so ist, *kleine* Fehler vorausgesetzt, der mittlere quadratische Fehler σ^2 einer Funktion $F(x_1, \ldots, x_n)$ gegeben durch $\sigma^2 = \big(\frac{\partial F}{\partial x_1}\big)^2\sigma_1^2 + \cdots + \big(\frac{\partial F}{\partial x_n}\big)^2\sigma_n^2$.

Sei nun $M'\mathbf{b}$ ein *anderer* linearer verzerrungsfreier Schätzer zum Modell $A\mathbf{x}$.

Es gilt $M'\mathbf{b} = (M' - M)A\mathbf{x}_0 + MA\mathbf{x}_0 + M'\boldsymbol{\varepsilon}$, also $M'\mathbf{b} - \mathbf{x}_0 = M'\boldsymbol{\varepsilon} + (M' - M)A\mathbf{x}_0$. Verzerrungsfreiheit bedeutet somit $(M' - M)A\mathbf{x}_0 = \mathbf{0}$ für beliebige $\mathbf{x}_0$.

Damit folgt auch $\mathrm{E}(|\mathbf{x}' - \mathbf{x}_0|^2) = \mathrm{E}(|M'\mathbf{b} - \mathbf{x}_0|^2) = \mathrm{E}(|M'\boldsymbol{\varepsilon}|^2) = \sigma^2\mathrm{Spur}(M'M'^T)$, und wegen $M'A = MA$, also $M'M^T = MM^T = MM'^T$, folgt $\mathrm{Spur}\, M'M'^T = \mathrm{Spur}\,(M' - M)(M' - M)^T + \mathrm{Spur}\, MM^T > \mathrm{Spur}\, MM^T$ für $M' \ne M$. ■

B Beweis der A-Posteriori-Fehlerschätzformel

Seien b_i die fehlerbehafteten Beobachtungsgrößen und $\mathbf{x}$ die Kleinste-Quadrate-Schätzung der Parameter.

Mit den Bezeichnungen aus Anhang A gilt $\mathbf{b} - A\mathbf{x}_0 = \boldsymbol{\varepsilon} \approx \mathbf{b} - A\mathbf{x} = (\mathbf{b} - A\mathbf{x}_0) + A(\mathbf{x}_0 - \mathbf{x}) = \boldsymbol{\varepsilon} - AM\boldsymbol{\varepsilon} =: B\boldsymbol{\varepsilon}$. Der erste Ausdruck ergibt die *wahren* (aber unbekannten), der zweite die nachträglich *geschätzten* Fehler der Beobachtungsgrößen. Der mittlere quadratische Einzelfehler $\mathrm{E}(\varepsilon_i^2) = \sigma^2$ soll nun anhand dessen *verzerrungsfrei* geschätzt werden. Zu dem Zweck wird $\mathrm{E}(|B\boldsymbol{\varepsilon}|^2) = \sigma^2\mathrm{Spur}\, B^TB$ (siehe Anhang A) ermittelt.

Sei $V = \langle\!\langle \mathbf{v}_1, \ldots, \mathbf{v}_k \rangle\!\rangle$ der Spaltenraum von A, aufgespannt durch eine Orthonormalbasis (ONB), analog $V^\perp = \langle\!\langle \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n \rangle\!\rangle$ sein orthogonales Komplement im $\mathbb{R}^n$.

Es gilt $B\mathbf{v} = \mathbf{0}$ $(\mathbf{v} \in V)$, da $MA = E$, also $BA = (E - AM)A = A - A$; ferner $B\mathbf{v} = \mathbf{v}$ $(\mathbf{v} \in V^\perp)$, da $A^T\mathbf{v} = \mathbf{0}$, also auch $M\mathbf{v} = \mathbf{0}$ für $\mathbf{v} \in V^\perp$.

Da stets $\mathrm{Spur}\, M_1M_2 = \mathrm{Spur}\, M_2M_1$, ergibt sich mit $Q := (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ und $Q' := (\mathbf{0}, \ldots, \mathbf{0}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$, dass $\mathrm{Spur}\, B^TB = \mathrm{Spur}\, Q^TB^TBQ = \mathrm{Spur}\, Q'^TQ' = n - k$ wegen $Q^T = Q^{-1}$.

Somit schließlich $\mathrm{E}(|B\boldsymbol{\varepsilon}|^2) = (n-k)\sigma^2$ und $\sigma^2 \approx \frac{1}{n-k}|\mathbf{b} - A\mathbf{x}|^2$. ■

C Kleinste Quadrate und Höchstwahrscheinlichkeit im Vergleich

a) Die Größen y_k seien exponentialverteilt mit Parameter $(bx_k)^{-1}$ $(1 \le k \le n)$; die x_k stehen für gewisse „Messpunkte". Dabei $x_1, \ldots, x_n, b > 0$, und es gilt $\mathrm{E}(y_k) = bx_k$ sowie $\mathrm{Var}(y_k) = (bx_k)^2$.

Maximum-Likelihood-Schätzung für b:

$L(y_1, \ldots, y_n; b) = \prod_{k=1}^n \frac{1}{bx_k}\mathrm{e}^{-y_k/(bx_k)}$ und damit $\frac{\partial \ln L}{\partial b} = -\sum_{k=1}^n \Big(\frac{1}{b} - \frac{y_k}{b^2x_k}\Big)$. Als ML-Schätzer folgt $b_{\mathrm{ML}} = \dfrac{1}{n}\displaystyle\sum_{k=1}^n \frac{y_k}{x_k}$. Der Schätzer ist linear und verzerrungsfrei, und $\mathrm{Var}(b_{\mathrm{ML}}) = \dfrac{b^2}{n}$.

Nun der Kleinste-Quadrate-Ansatz:

Da wir bekannte *unterschiedliche* mittlere Fehler bx_k haben, ist eine *gewichtete* Fehlerquadratsumme anzusetzen, $F(b) = \sum_{k=1}^n \frac{1}{x_k^2}(y_k - bx_k)^2$. $F'(b) = 0$ ergibt den Least-Squares-Schätzer $b_{\mathrm{LS}} = \dfrac{1}{n}\displaystyle\sum_{k=1}^n \frac{y_k}{x_k}$.

Dasselbe Ergebnis wie vorher, obwohl außer der Messgenauigkeit nichts an Wissen über die Fehlerverteilung benutzt wurde.

Benutzte man stattdessen die *ungewichtete* Fehlerquadratsumme $\tilde{F}(b) = \sum_{k=1}^n (y_k - bx_k)^2$, erhielte man den Schätzer $\tilde{b} = \sum_k y_kx_k \,/\, \sum_k x_k^2$, zwar auch linear und verzerrungsfrei, aber *schlechter* als der ML-Schätzer: $\mathrm{Var}(\tilde{b}) = b^2\sum x_k^4 \,/\, \big(\sum x_k^2\big)^2 \ge b^2/n$ (Cauchy/Schwarz-Ungleichung). Siehe auch [2], p. 352f.

b) Ein Würfel wird n-mal jeweils so lange geworfen, bis eine Sechs erscheint; die jeweiligen Wurfanzahlen k_i *vor* dem Auftreten der Sechs sind die Beobachtungsgrößen. Zu bestimmender Parameter: $p = \mathrm{P}(\text{Werfen einer Sechs})$.

Mit $\mathrm{P}(k_i = k) = (1-p)^k p$ gilt $\mathrm{E}(k_i) = \frac{1-p}{p}$, $\mathrm{Var}(k_i) = \frac{1-p}{p^2}$. Aus $L(k_1, \ldots, k_n; p) = p^n (1-p)^{k_1+\cdots+k_n}$ folgt der ML-Schätzer $p_{\mathrm{ML}} = \frac{n}{n+k_1+\cdots+k_n}$.

Wenn auch $\mathrm{E}(1/p_{\mathrm{ML}}) = 1/p$ (und sich für den Parameter $q := 1/\mathrm{P}(\text{Werfen einer Sechs})$ ein linearer verzerrungsfreier Schätzer ergäbe), ist dieser *nichtlineare* Schätzer *nicht* verzerrungsfrei. Mit einigem Umformungsaufwand kann man zeigen:

$$\mathrm{E}(p_{\mathrm{ML}}) = \sum_{k_1,\ldots,k_n \geq 0} \frac{n}{n+k_1+\cdots+k_n} p^n (1-p)^{k_1+\cdots+k_n} = \cdots = \int_0^1 \frac{nt^{n-1}\,\mathrm{d}t}{1+(1/p-1)t} \in \left(p + \frac{p(1-p)}{n+1},\, p + \frac{1-p}{n+1}\right).$$

Die Kleinste-Quadrate-Schätzung:

Modell $k = 1/p - 1$, nichtlinear. Mit der ungewichteten Fehlerquadratsumme $F(p) = \sum (k_i - 1/p + 1)^2$ (alle Beobachtungsgrößen haben dieselbe mittlere quadratische Abweichung) folgt wiederum $p_{\mathrm{LS}} = p_{\mathrm{ML}}$.

D Der Satz von de Moivre und Laplace

Gauß machte zum ersten Mal den Gedanken plausibel, dass die e^{-x^2}-Wahrscheinlichkeitsdichte sozusagen ubiquitär ist, was ihr ja dann den Namen „Normal"-Verteilung einbrachte. Aber schon 1738 zeigte Abraham de Moivre für $p = 1/2$, dass die Gewichte der Binomialverteilung sich im wesentlichen mit wachsendem n immer mehr an diese Wahrscheinlichkeitsdichte annähern. Im schwierigeren unsymmetrischen Allgemeinfall $p \in (0,1)$ – dessen Gültigkeit wohl auch schon de Moivre vermutete – gelang der Nachweis erst Pierre Simon de Laplace, und zwar kurz nach Gauß' Publikation der *Theoria motus*. Hier ein stromlinienförmiger Kurzbeweis, der sich nur auf die Asymptotik konzentriert, ohne alle Fehlerabschätzungen: Wir benutzen **(1)** $n! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$, **(2)** $\left(1+\frac{a}{n}\right)^n = \mathrm{e}^{n\ln(1+\frac{a}{n})} \sim \mathrm{e}^{a-\frac{a^2}{2n}}$ $(|a| \leq c\sqrt{n})$ sowie **(3)** $\frac{n}{k(n-k)} = \frac{1}{np(1-p)}\left(1+\frac{np-k}{k}\right)\left(1-\frac{np-k}{n-k}\right) \sim \frac{1}{np(1-p)}$ $(|np-k| \leq c\sqrt{n})$. Damit

$$\boldsymbol{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \underset{(1)}{\sim} \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{n}{k(n-k)}} \cdot \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k}$$
$$\underset{(3)}{\sim} \frac{1}{\sqrt{2\pi np(1-p)}} \left(1+\frac{np-k}{k}\right)^k \left(1-\frac{np-k}{n-k}\right)^{n-k}$$
$$\underset{(2)}{\sim} \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \mathrm{e}^{np-k-\frac{(np-k)^2}{2k}} \cdot \mathrm{e}^{-(np-k)-\frac{(np-k)^2}{2(n-k)}} = \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \mathrm{e}^{-\frac{(np-k)^2}{2}\left(\frac{1}{k}+\frac{1}{n-k}\right)}$$
$$\underset{(3)}{\sim} \boldsymbol{\frac{1}{\sqrt{2\pi np(1-p)}} \cdot \mathrm{e}^{-\frac{(np-k)^2}{2np(1-p)}}} \qquad (n \to \infty, \quad |np-k| \leq c\sqrt{n})$$

Die entscheidende asymptotische Beziehung, auf der diese Herleitung fußt, ist (1), die *Stirlingsche Formel*. Es war aber de Moivre selbst, der diese Formel hergeleitet hat. Bis auf den konstanten Faktor $\sqrt{2\pi}$; den hat, auf der Basis des Wallis-Produktes, Stirling beigesteuert.
Auch für die Stirling-Formel folgt hier ein möglichst kurzer Beweis:
$n! = \frac{(n+1)^n}{\prod_{k=1}^n \left(1+\frac{1}{k}\right)^k} = \frac{(n+1)^{n+1}}{\prod_{k=1}^n \left(1+\frac{1}{k}\right)^{k+1}}$, also $\boldsymbol{(n!)^2 = \frac{(n+1)^{2n+1}}{\prod_{k=1}^n \left(1+\frac{1}{k}\right)^{2k+1}}}$. Ferner
$(2k+1)\ln(1+\frac{1}{k}) = (2k+1)\sum_{i=1}^\infty (-1)^{i+1} \frac{1}{i \cdot k^i} = \cdots = 2 + \frac{1}{k^2} \underbrace{\sum_{i=0}^\infty \frac{(-1)^i}{i+4+2/(i+1)} \cdot \frac{1}{k^i}}_{\in (0,\frac{1}{6})}$, also

$n! = \frac{n^{n+1/2} \cdot \left(1+\frac{1}{n}\right)^{n+1/2}}{\prod_{k=1}^n \left(1+\frac{1}{k}\right)^{k+1/2}} = \sqrt{n} \cdot \left(\frac{n}{\mathrm{e}}\right)^n \cdot \frac{\left(1+\frac{1}{n}\right)^{n+1/2}}{\mathrm{e}^{\sum_{k=1}^n \theta_k \cdot \frac{1}{k^2}}}$ mit $\theta_k \in (0, \frac{1}{12})$. Daher $\frac{n!}{\sqrt{n} \cdot \left(\frac{n}{\mathrm{e}}\right)^n} \searrow \mathrm{e}^{1-c}$ wegen $\left(1+\frac{1}{n}\right)^{n+\frac{1}{2}} \searrow \mathrm{e}$. Und $\prod_{k=1}^\infty \left(1-\frac{1}{4k^2}\right) = \frac{2}{\pi} = \lim_{n\to\infty} \frac{((2n)!)^2}{4^{2n} \cdot (n!)^4} \cdot (2n+1)$ (Wallis) ergibt $\mathrm{e}^{1-c} = \sqrt{2\pi}$.

Literatur

[1] János Aczél, *Vorlesungen über Funktionalgleichungen und ihre Anwendungen*, Basel 1961

[2] Leo Breiman, *Statistics – With a View Toward Applications*, Boston 1973

[3] Augustin Louis Cauchy, *Analyse algébrique*, Paris 1821

[4] 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")

[5] –, *Theory of the Motion of Heavenly Bodies Moving About the Sun in Conic Sections*, Boston 1857 (Nachdruck: The Michigan Historical Reprint Series)

[6] –, *Theoria combinationis observationum erroribus minimis obnoxiae*, Göttingen 1821, 1823, 1828; auch: *Werke*, Bd. 4 („Theorie der am wenigsten fehlerbehafteten Kombination von Beobachtungen")

[7] –, *Theory of the Combination of Observations Least Subject to Errors*, zweisprachig (Latein/Englisch), Übersetzung und Kommentar: G. W. Stewart, Philadelphia 1995

[8] –, *Werke*, Bände 1–12 (10, 11 in je zwei Teilbänden), 1863–1929 (im Internet frei verfügbar)

[9] Anders Hald, *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713–1935*, New York 2010

[10] Georg Hamel, *Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung:* $f(x) + f(y) = f(x+y)$, Mathematische Annalen 60 (1905), 459-462

[11] Einar Hille, *Methods in Classical and Functional Analysis*, Reading (Mass.) 1972

[12] Pierre Simon Laplace, *Théorie analytique des probabilités*, Paris [1]1812, [3]1820; auch *Oeuvres complètes, tome 7*, 1886 (Nachdruck: Éditions Jacques Gabay)

[13] Adrien-Marie Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris 1805

[14] Ivo Schneider (Hrsg.), *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*, Darmstadt 1988

[15] Stephen M. Stigler, *The History of Statistics*, Cambridge (Mass.) 1986

Nachbemerkung:

Man stellt fest, dass viele Autoren, so etwa Stigler und Hald in den genannten Werken, aber z.B. auch Ulrich Krengel in einem Vortrag an der Universität Göttingen über die Beiträge von Gauß (*Von der Bestimmung von Planetenbahnen zur modernen Statistik: Carl Friedrich Gauß – Werk und Wirkung*, Math. Semesterberichte 53 (2006), p. 1–16), nur summarisch auf die Gaußsche Herleitung der Normalverteilung eingehen, so als hätten sie den Eindruck, die Argumentation sei zwar elementar, aber im einzelnen recht tüftelig, eben „typisch Gauß". Selbst Aczél zitiert in seinem Standardwerk Gauß nur ungenau und schreibt, Gauß habe 1809 vor Cauchy „in einer weniger exakten Form denselben Gedankengang verwendet" ([1], p. 44).

Ich halte diese Anmerkung Aczéls für *irreführend*. Meines Erachtens gibt es an der Exaktheit der Formulierung bei Gauß *nichts* auszusetzen. Im Prinzip ist es ein *ganz einfaches* Argument, das Gauß typischerweise in seiner Monographie für Experten lakonisch knapp als „Zweizeiler" formulierte, anders als Cauchy in seinem Lehrbuch für Studenten.

Wie auch immer – die Gaußsche Herleitung verdient es, expliziter angegeben zu werden und auch heute noch *genau* allgemein bekannt zu sein. Und der Bezug zur Cauchyschen Funktionalgleichung ist interessant und bemerkenswert: Es ist eine substanzielle und wichtige Anwendung der Gleichung ein dutzend Jahre *vor* Cauchy!

Die Cauchysche Funktionalgleichung

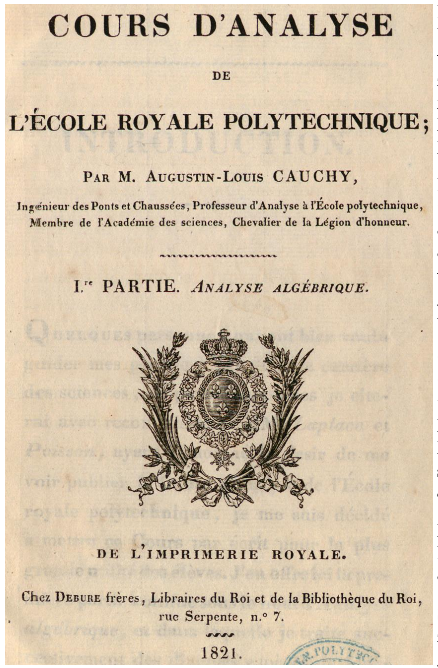
©Edgar M. E. Wermuth, Fakultät AMP, TH Nürnberg

Ergänzungs-Skriptum zum Vortrag an der Regiomontanus-Sternwarte, 23. Nov. 2023

Zusammenfassung

Es werden einige Grundtatsachen dargestellt über die Cauchysche Funktionalgleichung, die erstmals in Cauchys *Analyse algébrique* (1821) ausführlich diskutiert wurde.
Keywords: Algebraical analysis, Cauchy's functional equation.

1 Vorbemerkung


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.

Der 32-jährige Augustin Louis Cauchy (1789–1857) publizierte 1821 als Professor an der wenige Jahrzehnte zuvor von Gaspard Monge gegründeten und von Napoleon zur Militärakademie erklärten École Polytechnique in Paris, zu deren frühen Studenten (ab 1805) er selbst gehört hatte, das als bahnbrechendes Werk berühmt gewordene Lehrbuch *Analyse algébrique*.

Dieses Werk formuliert erstmals nahezu modern den Stetigkeitsbegriff und eine systematische Konvergenztheorie der unendlichen Reihen, es werden mustergültig die Gleichungen dritten und vierten Grades sowie der Fundamentalsatz der Algebra abgehandelt, und auch Cauchys eleganter Beweis der Ungleichung zwischen arithmetischem und geometrischem Mittel ist darin dargestellt.

Cauchy wurde berühmt durch wichtige Beiträge zu vielen Gebieten – er war einer der produktivsten Mathematiker überhaupt –, vor allem aber durch seine bahnbrechende Entwicklung der komplexen Analysis auf Basis komplexer Kurvenintegrale. (Später wurde entdeckt, dass Gauß wesentliche dieser Konzepte schon vorher entwickelt hatte, ohne sie zu publizieren.)

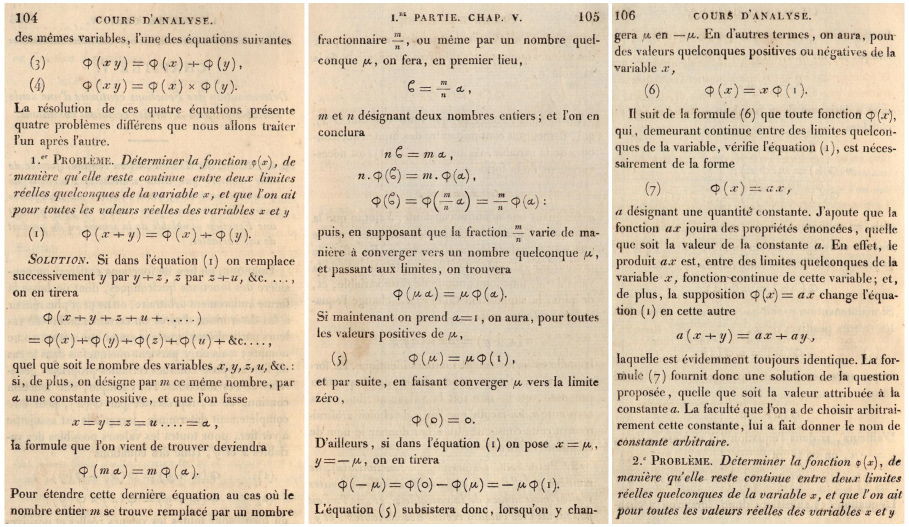
Es gibt nur wenige Mathematiker, deren Name mit so vielen Konzepten, Formeln und Resultaten verknüpft ist wie der Name Cauchy.

Schon als Jugendlicher wurde Cauchy, dessen Familie zeitweise in Armut lebte, von Joseph Louis Lagrange gefördert.

In der *Analyse algébrique* behandelte Cauchy auch erstmals ausführlich die nach ihm benannte grundlegende Funktionalgleichung sowie eine Reihe analoger Gleichungen.

2 Die Funktionalgleichung

Zunächst sei Cauchys Original reproduziert.


104 COURS D'ANALYSE.

des mêmes variables, l'une des équations suivantes

(3) $\varphi(xy) = \varphi(x) + \varphi(y)$,

(4) $\varphi(xy) = \varphi(x) \times \varphi(y)$.

La résolution de ces quatre équations présente quatre problèmes différens que nous allons traiter l'un après l'autre.

1.er PROBLÈME. *Déterminer la fonction $\varphi(x)$, de manière qu'elle reste continue entre deux limites réelles quelconques de la variable x, et que l'on ait pour toutes les valeurs réelles des variables x et y*

(1) $\varphi(x+y) = \varphi(x) + \varphi(y)$.

SOLUTION. Si dans l'équation (1) on remplace successivement y par $y+z$, z par $z+u$, &c. ..., on en tirera

$$\varphi(x+y+z+u+\ldots) = \varphi(x)+\varphi(y)+\varphi(z)+\varphi(u)+\text{\&c.}\ldots,$$

quel que soit le nombre des variables x, y, z, u, &c.: si, de plus, on désigne par m ce même nombre, par α une constante positive, et que l'on fasse

$$x = y = z = u \ldots = \alpha,$$

la formule que l'on vient de trouver deviendra

$$\varphi(m\alpha) = m\varphi(\alpha).$$

Pour étendre cette dernière équation au cas où le nombre entier m se trouve remplacé par un nombre

I.re PARTIE. CHAP. V. 105

fractionnaire $\frac{m}{n}$, ou même par un nombre quelconque μ, on fera, en premier lieu,

$$\beta = \frac{m}{n}\alpha,$$

m et n désignant deux nombres entiers; et l'on en conclura

$$n\beta = m\alpha,$$
$$n.\varphi(\beta) = m.\varphi(\alpha),$$
$$\varphi(\beta) = \varphi\left(\frac{m}{n}\alpha\right) = \frac{m}{n}\varphi(\alpha):$$

puis, en supposant que la fraction $\frac{m}{n}$ varie de manière à converger vers un nombre quelconque μ, et passant aux limites, on trouvera

$$\varphi(\mu\alpha) = \mu\varphi(\alpha).$$

Si maintenant on prend $\alpha = 1$, on aura, pour toutes les valeurs positives de μ,

(5) $\varphi(\mu) = \mu\varphi(1)$,

et par suite, en faisant converger μ vers la limite zéro,

$$\varphi(0) = 0.$$

D'ailleurs, si dans l'équation (1) on pose $x = \mu$, $y = -\mu$, on en tirera

$$\varphi(-\mu) = \varphi(0) - \varphi(\mu) = -\mu\varphi(1).$$

L'équation (5) subsistera donc, lorsqu'on y chan-

106 COURS D'ANALYSE.

gera μ en $-\mu$. En d'autres termes, on aura, pour des valeurs quelconques positives ou négatives de la variable x,

(6) $\varphi(x) = x\varphi(1)$.

Il suit de la formule (6) que toute fonction $\varphi(x)$, qui, demeurant continue entre des limites quelconques de la variable, vérifie l'équation (1), est nécessairement de la forme

(7) $\varphi(x) = ax$,

a désignant une quantité constante. J'ajoute que la fonction ax jouira des propriétés énoncées, quelle que soit la valeur de la constante a. En effet, le produit ax est, entre des limites quelconques de la variable x, fonction continue de cette variable; et, de plus, la supposition $\varphi(x) = ax$ change l'équation (1) en cette autre

$$a(x+y) = ax + ay,$$

laquelle est évidemment toujours identique. La formule (7) fournit donc une solution de la question proposée, quelle que soit la valeur attribuée à la constante a. La faculté que l'on a de choisir arbitrairement cette constante, lui a fait donner le nom de *constante arbitraire*.

2.e PROBLÈME. *Déterminer la fonction $\varphi(x)$, de manière qu'elle reste continue entre deux limites réelles quelconques de la variable x, et que l'on ait pour toutes les valeurs réelles des variables x et y*

Auf etwa zweieinhalb Seiten seines Lehrbuchs erörtert Cauchy in aller Ausführlickeit die Gesamtheit der *stetigen* Lösungen der Funktionalgleichung

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

und kommt zu dem Ergebnis, dass dies genau die Funktionen

$$f(x) = cx \quad (x \in \mathbb{R}) \tag{1}$$

mit einem beliebigen $c \in \mathbb{R}$ sind.

Eine etwas knappere Formulierung der *Herleitung*:
Mit $y = 0$ folgt aus ($\star$), dass $f(0) = 0$, und mit $y = -x$ ergibt sich $f(-x) = -f(x)$ $(x \in \mathbb{R})$. Per Induktion folgt $f(nx) = nf(x)$, $f(x/n) = f(x)/n$ für $n \in \mathbb{N}$ und damit insgesamt

$$f(rx) = rf(x) \quad (x \in \mathbb{R},\ r \in \mathbb{Q}). \tag{2}$$

Insbesondere also $f(r) = rf(1)$ $(r \in \mathbb{Q})$; im Falle der *Stetigkeit*$^{*)}$ also, da die rationalen Zahlen dicht liegen, $f(x) = cx$ $(x \in \mathbb{R})$ mit $c = f(1)$. Dass umgekehrt *jede* Funktion vom Typ (1) die Gleichung ($\star$) erfüllt, ist offensichtlich. ■ $^{*)}$ Stetigkeit in *einem* Punkt reicht: $f(x) - f(x_1) = f(x - x_1 + x_0) - f(x_0)$.

Es ist plausibel, die Funktionen vom Typ (1) als *triviale* Lösungen der Funktionalgleichung ($\star$) zu bezeichnen und zu fragen: Gibt es noch *andere* Funktionen, welche die Gleichung erfüllen?

Da die stetigen Lösungen auch durch

$$\frac{f(x)}{x} = c \quad (x \in \mathbb{R} \setminus \{0\}) \tag{3}$$

charakterisiert werden können, gibt es also im Falle einer *nicht*trivialen Lösung f von ($\star$) zwei Zahlen $a, b \in \mathbb{R} \setminus \{0\}$ mit $f(a)/a \neq f(b)/b$ oder $bf(a) - af(b) \neq 0$.

Jedes lineare Gleichungssystem

$$\begin{aligned} a\,x + b\,y &= \alpha \\ f(a)\,x + f(b)\,y &= \beta \end{aligned}$$

ist somit eindeutig lösbar. Zu gegebenem $\varepsilon > 0$ gibt es nun stets *rationale* Zahlen ξ und η nahe bei x bzw. nahe bei y mit $|a\xi + b\eta - \alpha| < \varepsilon,\ |f(a)\xi + f(b)\eta - \beta| < \varepsilon$.

Wegen $f(a)\xi + f(b)\eta = f(a\xi) + f(b\eta) = f(a\xi + b\eta)$ ist folglich gezeigt:

$$\text{Der Graph jeder unstetigen Lösung } f \text{ von } (\star) \text{ liegt } \textit{dicht} \text{ im } \mathbb{R}^2. \tag{4}$$

Unstetige Lösungen von $(\star)$, sollte es sie geben, sind also total unstetig.

Allerdings sind Funktionen mit im $\mathbb{R}^2$ dicht liegenden Graphen nicht ganz so exotisch, wie es auf den ersten Blick scheint: Man kann solche Funktionen leicht explizit konstruieren, indem man von einer Abzählung aller rationalen Zahlen und damit auch aller rationalen Zahlenpaare ausgeht. Indem man immer größere Teile des $\mathbb{R}^2$ durch immer feinere Quadratnetze überdeckt und in jedem Teilquadrat einen Punkt mit rationalen Koordinaten auswählt und dabei keine x-Koordinate sich wiederholen lässt, kann man eine dichte Funktion definieren, die im Lebesgueschen Sinne messbar und äquivalent zur Nullfunktion ist.

Es gilt auch:

$$\text{Eine (im Lebesgue-Sinne) } \textit{messbare} \text{ Lösung von } (\star) \text{ ist stetig.} \tag{5}$$

Beweis:
Mit $\mu(M)$ sei das Lebesgue-Maß einer Teilmenge von $\mathbb{R}$ bezeichnet, ferner gelte (mit *messbarem* f)

$$M_{k,0} := \{x \in (0,1) \mid k \leq f(x) < k+1\} \quad (k \in \mathbb{Z})\,.$$

Dann folgt $(0,1) = \bigcup_{k\in\mathbb{Z}} M_{k,0}$, also auch $\sum_{k\in\mathbb{Z}} \mu(M_{k,0}) = 1$.

Wäre nun f eine *unstetige* Lösung von $(\star)$, gäbe es nach (4) zu jedem $j \in \mathbb{N}$ ein $x_j \in (0, 2^{-j})$ mit $2j \leq f(x_j) < 2j+1$. Mit $M_{k,j} := (0,1) \cap (x_j + M_{k,0})$ folgte $k + 2j \leq f(x) < k + 2j + 2$ für $x \in M_{k,j}$. Die $M_{k,j}$ wären also *disjunkte* messbare Teilmengen von $(0,1)$, und $\mu(M_{k,j}) \geq \mu(M_{k,0}) - 2^{-j}$ ergäbe

$$1 \geq \sum_{j=1}^{n} \mu(M_{k,j}) > n\,\mu(M_{k,0}) - 1$$

für beliebige $n \in \mathbb{N}$. Daher $\mu(M_{k,0}) = 0$, im Widerspruch zu $\sum_{k\in\mathbb{Z}} \mu(M_{k,0}) = 1$. ■

(Kürzeres Argument bei *Integrierbarkeit*: $(\star) \Rightarrow \int_x^{x+1} f(t)\mathrm{d}t = f(x) + \int_0^1 f(t)\mathrm{d}t$; also ist f stetig.)

Von A. M. Ostrowski stammt folgende schärfere Aussage, deren Beweis nicht mehr ganz so einfach ist:[1]

$$\text{Gilt } (\star) \text{ und gibt es } a<b \text{ und } M\subseteq\mathbb{R}, \text{ so daß } \mu(M)>0 \text{ und } (a,b)\cap f(M)=\emptyset, \text{ so ist } f \text{ stetig.} \tag{6}$$

Noch immer nicht klar: *Gibt* es überhaupt unstetige Lösungen?

Diese Frage hat Georg Hamel 1905 *positiv* beantwortet, indem er den ein Jahr zuvor von Ernst Zermelo aus dem *Auswahlaxiom*[2] der Mengenlehre geschlussfolgerten *Wohlordnungssatz*[3] auf diese Frage anwandte. Auf der Basis der anderen Mengenlehre-Axiome sind Auswahlaxiom und Wohlordnungssatz logisch gleichwertig; aber das Auswahlaxiom folgt unmittelbar aus dem Wohlordnungssatz, während der umgekehrte Schluss um einiges schwieriger ist.

[1] Alexander Ostrowski, *Über die Funktionalgleichung der Exponentialfunktion und verwandte Funktionalgleichungen*, Jahresber. Deut. Math.-Ver. 38 (1929), S. 54-62; insbesondere Abschnitt IV, S. 58.

[2] Ist M eine nichtleere Menge, so gibt es eine Auswahl-Funktion, die jeder nichtleeren Teilmenge von M eines ihrer Elemente zuordnet.

[3] Jede Menge M lässt sich wohlordnen. D.h.: Es gibt eine Relation $\prec\ \subseteq M \times M$ (wobei $(x,y) \in\ \prec\ \ :\Leftrightarrow\ x \prec y$) mit den beiden Eigenschaften: a) $x \prec y \Rightarrow y \not\prec x\ (x,y \in M)$; b) $\emptyset \neq N \subseteq M \ \Rightarrow\ \exists x \in N$ mit $x \prec y$ für alle $y \in N \setminus \{x\}$. (Eine solche Relation $\prec$ heißt *Wohlordnung*).

Wir folgern nun die Existenz unstetiger Lösungen von $(\star)$ aus dem Wohlordnungssatz.

Sei $\prec$ eine Wohlordnungsrelation auf $\mathbb{R}$. Für jedes $x \in \mathbb{R}$ bezeichne $A(x) := \{y \in \mathbb{R} \mid y \prec x\}$ das zu x gehörige $\prec$-*Anfangsstück* von $\mathbb{R}$. Eine *algebraische Basis* des Vektorraumes $\mathbb{R}$ über dem Skalarenkörper $\mathbb{Q}$ (!) ist gegeben durch ($\mathrm{LK}_{\mathbb{Q}}$:= *endliche* algebraische Linearkombination *mit rationalen Koeffizienten*)

$$B := \{x \in \mathbb{R} \setminus \{0\} \mid x \text{ nicht } \mathrm{LK}_{\mathbb{Q}} \text{ von Elementen aus } A(x)\}. \tag{7}$$

Beweis: Sind $x_1, x_2, \ldots, x_n$ verschiedene Elemente aus B und dabei o.B.d.A. $x_1 \prec x_2 \prec \cdots \prec x_n$, also $x_k \in A(x_{k+1})$ $(1 \leq k < n)$, so folgt aus $r_1x_1 + \cdots + r_nx_n = 0$ mit $r_k \in \mathbb{Q}$ $(1 \leq k \leq n)$, dass $r_n = 0$, also auch $r_{n-1} = \cdots = r_1 = 0$; d.h.: $x_1, \ldots, x_n$ sind linear unabhängig. Außerdem ist jedes $x \in \mathbb{R}$ als $\mathrm{LK}_{\mathbb{Q}}$ von B-Elementen darstellbar. Denn wäre x die $\prec$-früheste *nicht* darstellbare reelle Zahl, also ganz $A(x)$ darstellbar und damit x nach Definition auch durch $A(x)$ *nicht* darstellbar, folgte $x \in B$ nach (7); ein Widerspruch. ■

Analog folgt die Existenz algebraischer Basen in beliebigen Vektorräumen, weshalb man heute – nach Georg Hamel – von *Hamel-Basen* spricht.[4]

Die Basis B ist unendlich, sogar *überabzählbar unendlich*, da dies für $\mathbb{R}$ gilt und die Menge aller endlichen rationalen Linearkombinationen von abzählbar vielen Elementen abzählbar ist.

Legt man nun Werte $f(x)$ $(x \in B)$ *irgendwie* fest, ergibt lineare Fortsetzung bzgl. des Koeffizientenkörpers $\mathbb{Q}$ eine Lösung von $(\star)$, und so werden *alle möglichen Lösungen* von $(\star)$ erfasst.

Allerdings ist diese „Konstruktion“ aller Lösungen der Funktionalgleichung nur *scheinbar* konkret. Denn die zugrundeliegende Wohlordnung der Menge $\mathbb{R}$ lässt sich nicht explizit angeben.[5] Das Auswahlaxiom ist ein reines Existenz-Prinzip. Deshalb waren Argumentationen, die sich auf dieses Axiom stützten, auch bei einigen der bedeutendsten Mathematiker wie Henri Poincaré, Henri Lebesgue, Hermann Weyl oder Luitzen E. J. Brouwer sehr umstritten. Man sprach längere Zeit von einer „Grundlagenkrise“ der Mathematik, zumindest von einem Grundlagen*streit*. Der Streit verblasste zwar mit der Zeit, aber ohne dass die Probleme wirklich endgültig überzeugend entschieden wurden.[6]

Untersuchungen zu den Grundlagen der Mengenlehre haben später zu der Erkenntnis geführt, dass man *ohne* das Auswahlaxiom die Existenz unstetiger Lösungen von $(\star)$ *nicht* nachweisen kann.[7]

Gilt

$$f(cx) = cf(x)\ \ (c \in \mathbb{R}) \text{ für ein } x \in \mathbb{R} \setminus \{0\}, \tag{8}$$

folgt, indem man c/x anstelle von c einsetzt, dass $f(c) = c\,f(x)/x$ $(c \in \mathbb{R})$. Im Falle einer unstetigen Lösung von $(\star)$ gibt es also zu *jedem* $x \neq 0$ ein $c \in \mathbb{R}$ mit $f(cx) \neq cf(x)$; dabei handelt es sich natürlich immer um ein *irrationales* c. Gilt hingegen

$$f(cx) = cf(x)\ \ (x \in \mathbb{R}) \text{ für ein } c \in \mathbb{R} \setminus \{0\}, \tag{9}$$

folgt daraus *nichts*. Denn bei beliebigen Lösungen von $(\star)$ gilt es für alle *rationalen* c, bei stetigen Lösungen für *alle* c, und auch bei unstetigen Lösungen kann es für gewisse *irrationale* c gelten. Um letzteres einzusehen, sei $\mathbb{R}$ als Vektorraum über dem Körper $\mathbb{A}$ der *reellen algebraischen Zahlen* (= der reellen

[4] Der Titel von Hamels Arbeit lautet dementsprechend: *Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung:* $f(x+y) = f(x) + f(y)$, Math. Annalen 60 (1905), S. 459–462. In dieser Arbeit wird auch erstmals (4) bewiesen.

[5] Zwar kann man ganz konkret unendliche linear unabhängige Teilmengen von $\mathbb{R}$ als Vektorraum über $\mathbb{Q}$ angeben, nämlich z.B. $\{\ln p \mid p \text{ Primzahl}\}$; aber so erhält man keine Basis.

[6] Diskussionsbeiträge bedeutender Forscher zur Grundlagen-Problematik sind zusammengestellt in dem Werk: Paul Benacerraf, Hilary Putnam, *Philosophy of Mathematics – Selected Readings*, Cambridge [2]1983. Viele weitere Bücher befassen sich mit mathematischen Grundlagenfragen. Nur zwei herausragende seien genannt: Hermann Weyl, *Philosophie der Mathematik und Naturwissenschaft*, München 1966; Philip J. Davis, Reuben Hersh, *Erfahrung Mathematik*, Basel 1985.

[7] Robert M. Solovay, *A model of set theory in which every set of reals is Lebesgue measurable*, Ann. of Math. 92 (1970), pp. 97–152. In Solovays Modell gilt nicht das allgemeine Auswahlaxiom, aber das *Prinzip der rekursiven Auswahl.* Wesentliche, nach einem Ergebnis von S. Shelah (1984) unverzichtbare Voraussetzung von Solovays Modell: die Existenz *unerreichbarer Kardinalzahlen.* Zum Prinzip der rekursiven Auswahl: Thomas J. Jech, *The Axiom of Choice*, Amsterdam 1973, Nachdruck New York 2008; zu unerreichbaren Kardinalzahlen: Azriel Levy, *Basic Set Theory*, Berlin 1979.

Wurzeln von Polynomgleichungen $p(x) = 0$ mit *ganzzahligen* Koeffizienten) betrachtet;[8] wir definieren analog zu (7) eine Hamel-Basis B', durch welche jedes $x \in \mathbb{R}$ als $\mathrm{LK}_{\mathbb{A}}$ eindeutig darstellbar ist. Auch diese Hamel-Basis ist überabzählbar, da $\mathbb{A}$ bekanntlich abzählbar ist, wie schon Cantor zeigte. Werden nun Werte $f(x)$ $(x \in B')$ irgendwie beliebig festgelegt und dann f bzgl. Linearkombinationen mit $\mathbb{A}$-Koeffizienten linear fortgesetzt, hat man eine Lösung von $(\star)$, für die $f(ax) = af(x)$ $(x \in \mathbb{R})$ für alle $a \in \mathbb{A}$, also nicht nur für beliebige rationale a, sondern zusätzlich für $a = \sqrt{2}, a = \sqrt{3}, a = \sqrt{2 + \sqrt{3}}$, usw.

Es gibt auch unstetige Lösungen von $(\star)$, für die $f(\pi x) = \pi f(x)$ $(x \in \mathbb{R})$ gilt; dazu geht man aus vom Skalarenkörper $\mathbb{Q}[\pi]$, einem transzendenten Erweiterungskörper von $\mathbb{Q}$, ebenfalls noch abzählbar unendlich; die f-Definition analog zum Fall zuvor ist dann klar.

Durch Verfeinerung der zuletzt benutzten Überlegungen haben Daróczy (1961), Losonczi (1964), Kemperman (1969) gezeigt, wie man Lösungen von $(\star)$ konstruiert, die

$$f(\tau_1 x) = \tau_2 x \quad (x \in \mathbb{R}) \tag{10}$$

für zwei *verschiedene* transzendente Zahlen τ_1, τ_2 erfüllen oder sogar für mehrere solche Zahlenpaare gleichzeitig.[9]

3 Formulierungsvarianten der Funktionalgleichung

Man kann $(\star)$ folgendermaßen modifizieren:

$$f(x + y) = f(x) + f(y) \quad \big(x, y, x + y \in (0, a)\big) \tag{11}$$

mit einem $a > 0$. Wie bisher folgt unmittelbar $f(rx) = rf(x)$ für $r \in \mathbb{Q}$ und $x, rx \in (0, a)$. Die einzige Möglichkeit, f auf $(-a, a)$ fortzusetzen, lautet $f(0) := 0,\ f(-x) := -f(x)\ (0 < x < a)$.

Es ist klar, dass dann die Funktionalgleichung für $x, y, x + y \in (-a, a)$ gilt: Etwa für $0 < x \le y < a$ folgt $f(x) + f(-y) = f(x) - f(y) = f(x) - \big(f(y - x) + f(x)\big) = -f(y - x) = f(x - y)$.

Mittels der notwendigen Eigenschaft $f(nx) = nf(x)$ ist dann f eindeutig auf ganz $\mathbb{R}$ fortsetzbar, da mit $x_1, x_2 \in (-a, a)$ und $x = n_1 x_1 = n_2 x_2$ ja $x_2 = \frac{n_1}{n_2} x_1$ und damit $f(x_2) = \frac{n_1}{n_2} f(x_1)$ folgt. Kurz:

$$\text{Jede Lösung von (11) ist Restriktion genau einer Lösung von } (\star). \tag{12}$$

Analog reicht es auch, von

$$f(x + y) = f(x) + f(y) \quad (x, y > a) \tag{13}$$

mit einem $a > 0$ auszugehen. Man kann dann f auf das Intervall $(a/2, \infty)$ fortsetzen durch die (natürlich zwangsläufige) Festlegung

$$f(x) := f(2x)/2 \quad (a/2 < x \le a).$$

Dann folgt für $x, y > a/2$: $f(x) + f(y) = (f(2x) + f(2y))/2 = f(2(x + y))/2 = 2f(x + y)/2 = f(x + y)$. Durch eine Folge solcher Schritte dehnt man f auf das Intervall $(0, \infty)$ aus und damit letztlich auf ganz $\mathbb{R}$, im Einklang mit der Funktionalgleichung.

Geht man zu einem *anderen* Definitionsintervall über, gilt dies nicht mehr ohne weiteres.

Gelte etwa $f(x + y) = f(x) + f(y)$ $\big(x, y, x + y \in (a, b)\big)$. Nur Argumente $> 2a$ sind als *Summe*, nur Argumente $< b - a$ als *Differenz* zweier Argumente aus (a, b) darstellbar. Was bedeutet: Werte $f(x)$ mit $b - a \le x \le 2a$ sind durch die Bedingung $f(x+y) = f(x)+f(y)$ $\big(x, y, x+y \in (a, b)\big)$ nicht gebunden. Also ist f im Falle $a < b \le 3a$ nicht unbedingt zu einer Funktion gemäß (13) fortsetzbar; denn die f-Werte sind völlig willkürlich, falls $b \le 2a$, und für $2a < b \le 3a$ wäre bei Fortsetzbarkeit der Wert von $f(b - a)$ durch $f(b - a) = rf(x)$ mit $r \in \mathbb{Q}$ und $b - a = rx,\ x \in (a, b - a)$ festgelegt.

[8] Man kann elementar linear-algebraisch zeigen, dass Summe und Produkt zweier algebraischer Zahlen wieder algebraisch sind, also die algebraischen Zahlen einen *Körper* bilden. Die *reellen* algebraischen Zahlen sind demnach ein Unterkörper.

[9] Einar Hille, *Methods in Classical and Functional Analysis*, Reading (Mass.) 1972; siehe Chap. 14.2: *Cauchy's Equation and Generalizations*.

H. W. Pexider diskutierte 1903 die Funktionalgleichung

$$f(x + y) = g(x) + h(y) \quad (x, y, x + y \in \mathbb{R}). \tag{14}$$

Diese ist – entgegen dem ersten Anschein – nur unwesentlich allgemeiner als $(\star)$. Denn es folgt sofort $f(0) = g(0) + h(0)$, $f(x) = g(x) + h(0)$ und $f(y) = g(0) + h(y)$, also auch $f(x + y) = f(x) + f(y) - f(0)$. Damit erfüllt $\tilde{f}(x) := f(x) - f(0)$ die Gleichung $\tilde{f}(x + y) = \tilde{f}(x) + \tilde{f}(y)$.
Fazit: f, g und h unterscheiden sich jeweils um eine *Konstante* von einundderselben Lösung $\tilde{f}$ von $(\star)$, wobei $f(0) = g(0) + h(0)$ gilt. Klar ist, dass umgekehrt *jedes* Tripel $f = \tilde{f} + a + b,\ g = \tilde{f} + a,\ h = \tilde{f} + b$ mit Konstanten $a, b \in \mathbb{R}$ die Gleichung (14) erfüllt.

Nun die Gaußsche Variante der Funktionalgleichung, die implizit formuliert ist in Artikel 177 der *Theoria motus* (1809):[10]

$$x_1 + x_2 + \dots + x_n = 0 \ \Rightarrow\ f(x_1) + f(x_2) + \dots + f(x_n) = 0 \quad \big(n \in \mathbb{N},\ x_1, x_2, \dots, x_n \in (-a, a)\big) \tag{15}$$

mit einem $a > 0$. Mit $n = 1, x_1 = 0$ folgt $f(0) = 0$, mit $n = 2$ ergibt sich $f(-x) = -f(x)$ und damit auch $f(x) + f(y) - f(x + y) = 0$ $(x, y, x + y \in (-a, a))$, also (11), auf das Intervall $(-a, a)$ ausgedehnt. Es gibt also zu jeder Lösung von (15) eine eindeutig bestimmte Fortsetzung zu einer Lösung von $(\star)$. Umgekehrt folgt aus $(\star)$ offensichtlich per Induktion (15) für jedes $a > 0$.

Bei Gauß steht, bedingt durch den Zusammenhang, in dem die Überlegung angestellt wird, φ' anstelle von f und $M - p, M' - p, M'' - p, \dots$ anstelle von $x_1, x_2, x_3, \dots$ (Siehe die ersten sechs Zeilen von S. 260 der deutschen Übersetzung der *Theoria motus*, wiedergegeben am Ende dieses Artikels.)

Er wählt speziell $x_2 = x_3 = \dots = x_n = -x_1/(n - 1) =: x$, folgert $(n - 1)f(x) = -f((1 - n)x)$ und stellt lakonisch fest, daraus ergebe sich leicht die Konstanz von $f(x)/x$. (Er setzt ja – wie Cauchy – Stetigkeit von f voraus.) Die *Theoria motus* richtet sich, anders als Cauchys Lehrbuch für junge Studierende, an Forscher und formuliert deshalb solche einfachen mathematischen Zusammenhänge erheblich knapper. Meines Erachtens ist ohne jede Einschränkung festzustellen:[11]

$$\text{Schon Gauß hat 1809 die stetigen Lösungen der Cauchyschen Funktionalgleichung bestimmt.} \tag{16}$$

4 Anwendungen und verwandte Funktionalgleichungen

Additive Funktionen, also Lösungen von $(\star)$, spielen naturgemäß eine große Rolle in der Mathematik. Die wichtigste Funktion der Mathematik überhaupt genügt der Funktionalgleichung

$$g(x + y) = g(x)g(y) \quad (x, y \in \mathbb{R}). \tag{17}$$

Triviale Lösung: $g(x) \equiv 0$. Für eine *nicht*triviale Lösung folgt $g(x) \ne 0$ $(x \in \mathbb{R})$ und damit $(x = y)$ $g(x) > 0$ $(x \in \mathbb{R})$. Also erfüllt $f := \ln g$ die Gleichung $(\star)$. Mithin ist exp nicht allein durch (17) charakterisiert; Hinzunahme von $g(x) \ge 1 + x$ in einer Umgebung von $x = 0$ legt die Lösung fest. Auch

$$g(x + y) \ge g(x)(1 + y) \quad (x, x + y \ge 0) \tag{18}$$

[10] *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, Hamburg 1809. Deutsche Übersetzung von Carl Haase: *Theorie der Bewegung der Himmelskörper, welche in Kegelschnitten die Sonne umlaufen*, Gotha 1877. Englische Ausgabe: *Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections*, Boston 1857. Übers.: Charles Henry Davis; eine lesenswerte Kurz-Biografie von Davis findet man bei David L. Roberts, *The Republic of Numbers*, Baltimore 2019. Die Charakterisierung der stetigen Lösungen von (15) ist ein einfacher, aber wesentlicher Überlegungsschritt bei der ersten Gaußschen Herleitung der *Normalverteilung*.

[11] Dies scheint nicht allgemein bekannt zu sein. In seinem bekannten Lehrbuch über Funktionalgleichungen aus dem Jahre 1961 erwähnt J. Aczel zwar Gauß in einer Fußnote (S. 44), merkt aber nur knapp an, die Funktionalgleichung komme bei Gauß „in einer weniger exakten Form“ vor. Das ist meines Erachtens so ungerechtfertigt wie irreführend.

allein charakterisiert die Exponentialfunktion, da $g(x) \geq g(0)(1+x/n)^n, g(0) \geq g(x)(1-x/n)^n$ folgt und damit: (18) $\Leftrightarrow$ $g(x) = c\mathrm{e}^x$ $(x \geq 0)$ mit einem $c \geq 0$.
(Wegen $g(0)\mathrm{e}^{x+y} \geq g(0)\mathrm{e}^x(1+y)$ und $\mathrm{e}^y > 1+y$ $(y \neq 0)$ muss $g(0) \geq 0$ gelten.)

Eine weitere verwandte Funktionalgleichung (Beispiel: *Betrags*-Multiplikativität) ist

$$g(xy) = g(x)g(y) \quad (x, y \in \mathbb{R}). \tag{19}$$

Für Lösungen $\not\equiv 0$ von (19) gilt $g(x) \neq 0$ $(x \neq 0)$, also $g(x) > 0$ $(x > 0)$ wegen $x = \sqrt{x}\sqrt{x}$. Somit erfüllt $f(x) := \ln g(\ln x)$ $(x > 1)$ die Gleichung (13), und $g := \exp \circ f \circ \exp$ ist Lösung von (19) für $x > 0$. Wegen $g(-x) = g(x)g(-1)$ ist durch $g(-x) = g(x)$ $(x > 0)$ oder durch $g(-x) = -g(x)$ $(x > 0)$ die Lösung auf negative Werte auszudehnen. Wegen $g(0) = g(0)g(x)$ $(x \in \mathbb{R})$ folgt schließlich $g(0) = 0$ im Falle $g \not\equiv 1$.

In der Geometrie spielen die Lösungen von $(\star)$ insbesondere bei der Einführung von Vektoren in der Ebene eine Rolle; siehe etwa G. Hessenberg, J. Diller, *Grundlagen der Geometrie*, de Gruyter, Berlin 1967, Seite 54ff. Beim Nachweis, dass $\lambda\mathbf{v} + \lambda\mathbf{w} = \lambda(\mathbf{v} + \mathbf{w})$ für Skalare λ und ebene Vektoren $\mathbf{v}$ und $\mathbf{w}$ gilt, wird benutzt, dass die einzigen monotonen Lösungen von $(\star)$ die trivialen sind.

Auch in der Stochastik tritt (1) auf. Siehe z.B. Sheldon M. Ross, *Applied Probability Models with Optimization Applications*, Dover, New York 1992 (urspr. Holden-Day, San Francisco 1970), Seite 9 und Seite 26f. (Anwendung auf stationäre Punktprozesse).

W. B. Jurkat zeigte:[12]

$$F(x+y) = F(x) + F(y) \quad ((x, y) \in \mathbb{R}^2 \setminus \mathcal{A}) \tag{20}$$

mit einer Lebesgue-*Nullmenge* $\mathcal{A} \subset \mathbb{R}^2$ gilt genau dann, wenn es eine fast überall mit $F(x)$ übereinstimmende Funktion $f : \mathbb{R} \to \mathbb{R}$ gibt, die $(\star)$ erfüllt; f ist eindeutig bestimmt.

[12] Wolfgang B. Jurkat, *On Cauchy's Functional Equation*, Proc. Amer. Math. Soc. 16 (1965), 683-686.

Eine Seite aus der deutschen Übersetzung (Gotha 1877, Übersetzer Carl Haase) der *Theoria motus corporum coelestium in sectionibus conicis solem ambientium* (Gauß 1809, siehe Fußnote weiter oben):

so wird allgemein sein müssen $\varphi'(M-p) + \varphi'(M'-p) + \varphi'(M''-p) +$ etc. $= 0$, wenn für p der Werth $\frac{1}{\mu}(M + M' + M'' +$ etc.$)$ substituirt wird, welches positive Ganzes nun auch durch μ ausdrückt sein mag. Setzt man daher voraus $M' = M'' =$ etc. $= M - \mu N$, so wird allgemein, d. h. für jeden ganzen positiven Werth für μ, sein $\varphi'(\mu - 1)N = (1-\mu)\varphi'(-N)$, woraus man leicht sieht, dass allgemein $\frac{\varphi'\Delta}{\Delta}$ eine constante Grösse sein müsse, welche ich mit k bezeichnen will. Hieraus wird $\log \varphi\Delta = \frac{1}{2}k\Delta\Delta +$ Const., oder wenn man die Basis der hyperbolischen Logarithmen mit e bezeichnet und die Constante $= \log \varkappa$ setzt,

$$\varphi\Delta = \varkappa e^{\frac{1}{2}k\Delta\Delta}.$$

Ferner sieht man leicht ein, dass k nothwendig negativ sein müsse, damit Ω in der That ein Grösstes werden könne, weshalb wir setzen $\frac{1}{2}k = -hh$; und da vermittelst des eleganten, zuerst von Laplace*) gefundenen Theorems das Integral $\int e^{-hh\Delta\Delta}\mathrm{d}\Delta$, von $\Delta = -\infty$ bis zu $\Delta = +\infty$, wird $= \frac{\sqrt{\pi}}{h}$ (wobei π den halben Kreisumfang für den Radius $= 1$ bezeichnet), so wird unsere Function werden:

$$\varphi\Delta = \frac{h}{\sqrt{\pi}} e^{-hh\Delta\Delta}.$$

178.

Die so eben ermittelte Function kann zwar nicht in aller Strenge die Wahrscheinlichkeiten der Fehler ausdrücken; denn da die möglichen Fehler
(213) stets in gewisse Grenzen eingezwängt sind, so müsste die Wahrscheinlichkeit grösserer Fehler immer $= 0$ herauskommen, während unsere Formel stets einen begrenzten Werth darstellt. Dennoch aber ist dieser Mangel, an welchem jede analytische Function ihrer Natur nach laboriren muss, für jeden praktischen

*) In v. Zach „Monatliche Correspondenz" Band 21, S. 280 äussert Gauss: „Dass Euler schon das Theorem gefunden hat, woraus der schöne, von mir Laplace beigelegte Lehrsatz sehr leicht abgeleitet werden kann, fiel mir selbst schon früher ein, als aber die Stelle S. 212 schon abgedruckt war; ich wollte es aber nicht unter die Errata setzen, weil Laplace wenigstens das obige Theorem doch erst in der dort gebrauchten Form aufgestellt hat." *Anmerkung des Uebersetzers.*