

Gauss'sche Fehlerrechnung

T. Ihn

24. Oktober 2016

Inhaltsverzeichnis

1	Modell und Likelihood	1
2	Alle Standardabweichungen σ_i sind bekannt, bzw. die Kovarianzmatrix der Daten ist bekannt: Minimieren der χ^2-Funktion.	6
2.1	Allgemeine Behandlung	6
2.1.1	Priorverteilung der Parameter θ	6
2.1.2	Anwendung des Bayesschen Theorems: von der Priorverteilung zur Posteriorverteilung.	7
2.1.3	Posteriorverteilung für die Parameter.	8
2.2	Spezialfall 1: Konstante Fitfunktion und unkorrelierte Fehler	9
2.3	Spezialfall 2: lineare Fitfunktion und unkorrelierte Fehler	11
2.4	Genauigkeit der Schätzwerte	13
3	Unkorrelierte Fehler, bei denen die Standardabweichungen σ_i nicht bekannt, aber alle gleich σ sind: Methode der kleinsten Fehlerquadrate (non-linear least squares)	20
3.1	Allgemeine Behandlung	20
3.1.1	Priorverteilung der Parameter θ und σ	20
3.1.2	Posteriorverteilung für die Parameter	21
3.2	Spezialfall 1: konstante Fitfunktion	23
3.3	Spezialfall 2: lineare Fitfunktion	24
3.4	Genauigkeit der Schätzwerte	24
3.4.1	Genauigkeit aus der Posteriorverteilung der Parameter	24
3.4.2	Marginalisieren des Parameters σ	30
3.5	Spezialfall 2: lineare Fitfunktion	40

1 Modell und Likelihood

Es liegt uns aus einer Messung eine Serie von N Wertepaaren (x_i, y_i) vor.

Erste Annahme: additives Fehlermodell. Wir nehmen an, die Messgrösse y hänge mit der Stellgrösse x über

$$y_i = f(x_i; \theta) + \epsilon_i$$

zusammen, wobei θ für die M Parameter des Modells steht. Man nennt y die *abhängige Variable*, x heisst *unabhängige Variable* des Modells. Die Funktion f beschreibt die systematische Abhängigkeit der wahren Werte der abhängigen Variablen y von der unabhängigen Variablen x . Die Werte ϵ_i beschreiben statistische Abweichungen der gemessenen abhängigen Variablen von den wahren Werten. Die unabhängige Variable wird als exakt bekannt angenommen.

Bemerkung 1. *Die Annahme, dass die Werte x_i der unabhängigen Variablen exakt bekannt sind ist praktisch nicht erfüllbar und stellt somit eine vereinfachende Abstraktion dar. Praktisch muss man die statistischen Abweichungen δx_i der x_i von den wahren Werten mit dem Bereich Δx vergleichen, über den die x_i gemessen wurden. Ebenso muss man die statistischen Abweichungen ϵ_i der y_i von den wahren Werten mit dem Bereich Δy vergleichen, in dem alle gemessenen y -Werte liegen. Dann sollte $\delta x_i / \Delta x \ll \epsilon_i / \Delta y$ gelten, damit die obige Annahme näherungsweise erfüllt ist.*

Beispiel 1. *Die Funktion $f(x; \theta)$ kann im einfachsten Fall eine Konstante sein, z.B.*

$$y_i = \mu + \epsilon_i.$$

In diesem Fall ist die Funktion unabhängig von x . Das Modell besitzt einen Parameter μ . Dieser Fall tritt auf, wenn man ein und dieselbe Grösse wiederholt unter denselben Bedingungen misst.

Beispiel 2. *Die Funktion $f(x; \theta)$ beschreibt eine Gerade, d.h.*

$$y_i = \underbrace{ax_i + b}_{f(x_i; a, b)} + \epsilon_i.$$

Hier besitzt das Modell die beiden Parameter a und b .

Beispiel 3. *Die Funktion $f(x; \theta)$ beschreibt eine exponentiell abfallende Abhängigkeit gemäss*

$$y_i = \underbrace{Ae^{-\lambda x_i}}_{f(x_i; A, \lambda)} + \epsilon_i.$$

Dieses Modell besitzt die Parameter A und λ , wobei $A > 0$ und $\lambda > 0$.

Zweite Annahme: normalverteilte Abweichungen von der theoretischen Funktion. Die ϵ_i sind normalverteilt gemäss

$$\text{pdf}(\epsilon_i|\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \frac{\epsilon_i^2}{\sigma_i^2}\right).$$

Die Grösse σ_i nennt man die *Standardabweichung* (eng. ‘standard deviation’) des i -ten Messwerts. Die Normalverteilung heisst auch Gauss-Verteilung oder Gausssche Normalverteilung. Daher kommt die Bezeichnung *Gaussische Fehlerrechnung*.

Bemerkung 2. Die Schreibweise $\text{pdf}()$ bezeichnet eine *Wahrscheinlichkeitsdichtefunktion* (*probability density function*). Die Schreibweise $\text{pdf}(\cdot|\cdot)$ kennzeichnet eine *bedingte Wahrscheinlichkeit*, wobei die Bedingung rechts vom senkrechten Strich ‘|’ zu finden ist. Wir lesen $\text{pdf}(\epsilon_i|\sigma_i)$ als ‘die Wahrscheinlichkeit, dass die statistische Abweichung des Messpunktes vom wahren Wert ϵ_i ist, unter der Bedingung dass die Genauigkeit des Messgeräts σ_i exakt bekannt ist.’

Bemerkung 3. Die *Wahrscheinlichkeitsdichte* $\text{pdf}(\epsilon_i|\sigma_i)$ hat *vorhersagenden* (*prediktiven*) Charakter. Sie sagt die *Wahrscheinlichkeit* für in der Zukunft zu messende Daten (bzw. deren Fehler) voraus, unter der Annahme dass die Parameter σ_i der Messapparatur bekannt sind.

Bemerkung 4. Die Parameter σ_i beschreiben die *Genauigkeit* (engl. *precision*) des i -ten Messwerts und werden häufig in graphischen Darstellungen als Fehlerbalken eingezeichnet. Ein Fehlerbalken erstreckt sich in der Regel von $y_i - \sigma_i$ bis $y_i + \sigma_i$ (wenn nichts anderes angegeben ist).

Bemerkung 5. Die Werte σ_i sind bei unmittelbar gemessenen Grössen in der Regel *zeitunabhängige Eigenschaften* des Messverfahrens. Messen wir z.B. eine Länge mit einem Lineal, dann ist die Messgenauigkeit $\sigma = 1/6$ mm (wenn das Lineal Millimeterstriche hat). Messen wir einen elektrischen Strom mit einem Amperemeter, dessen Genauigkeit $10 \mu A$ ist, dann bestimmt dies den Wert von σ_i .

Bemerkung 6. Die Normalverteilung enthält für jeden Messwert i einen neuen Parameter σ_i . Haben wir aus der Funktion $f(x; \theta)$ bereits $M \geq 1$ Parameter, so kommen nun N weitere Parameter hinzu, so dass unser Modell gesamthaft $N + M$ Parameter enthält. Da wir nicht hoffen können aus N Messpunkten (x_i, y_i) $N + M$ Parameter zu bestimmen, ist das System *unterbestimmt*, wenn alle diese Parameter unbekannt sind.

Bemerkung 7. Sind die Werte σ_i hingegen bekannt (d.h. wir kennen die Fehlerbalken jedes Wertes y_i), so ist das System *überbestimmt*, sobald $N > M$ ist. Das ist einer der typischen Fälle der Gaussischen Fehlerrechnung. Wir merken uns daher, dass wir immer mehr Messpunkte brauchen, als unser Modell unbekannt Parameter enthält.

Bemerkung 8. Häufig sind die $\sigma_i \equiv \sigma$ unabhängig vom Messpunkt i . Z.B. ist die Ablesegenauigkeit eines Lineals oder eines anderen analogen Messgeräts nicht vom Messwert abhängig. Dann ist die Zahl der Parameter des Modells $M + 1$. Haben wir $N > M + 1$ Messpunkte, so liegt ein überbestimmtes System vor. Falls σ nicht bekannt ist, kann es in diesem Fall aus der Messung bestimmt werden, wie jeder andere Parameter auch.

Der Fall statistisch unabhängiger Fehler ϵ_i . Bei statistischer Unabhängigkeit der ϵ_i haben wir

$$\text{pdf}(\{\epsilon_i\}|\{\sigma_i\}) = \frac{1}{(2\pi)^{N/2} \prod_{i=1}^N \sigma_i} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma_i^2}\right),$$

wobei N die Zahl der Datenpunkte ist. Somit wird die Stichprobenverteilung

$$\boxed{\text{pdf}(\{y_i\}|\{x_i\}, \theta, \{\sigma_i\}) = \frac{1}{(2\pi)^{N/2} \prod_{i=1}^N \sigma_i} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{(y_i - f(x_i; \theta))^2}{\sigma_i^2}\right)}, \quad (1)$$

Bemerkung 9. Die Schreibweise $\{\epsilon_i\}$ ist eine Abkürzung für $\{\epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_N\}$. Die Wahrscheinlichkeit $\text{pdf}(\{\epsilon_i\}|\{\sigma_i\})$ ist die Verbundwahrscheinlichkeit (joint probability density) für das Auftreten der zu messenden ϵ_i -Werte.

Bemerkung 10. Um von der Wahrscheinlichkeitsverteilung der einzelnen ϵ_i zu dieser Verbundwahrscheinlichkeit zu kommen, haben wir die Produktregel der Wahrscheinlichkeitsrechnung

$$\text{prob}(A, B|C) = \text{prob}(A|C)\text{prob}(B|A, C)$$

verwendet. Statistische Unabhängigkeit von A und B bedeutet $\text{prob}(B|A, C) = \text{prob}(B|C)$, so dass

$$\text{prob}(A, B|C) = \text{prob}(A|C)\text{prob}(B|C).$$

Bemerkung 11. Hier stellt sich die wichtige Frage, wie man praktisch feststellen kann, ob die ϵ_i statistisch unabhängig sind. Aufeinanderfolgende Wertepaare (x_i, y_i) werden in einem gewissen zeitlichen Abstand Δt gemessen. Diesen zeitlichen Abstand muss man mit der ‘Antwortzeit’ (engl. response time) τ des Messsystems vergleichen. Das ist die Zeit, die das Messsystem braucht, bis sich der Messwert y nach einer Änderung der Grösse x auf den neuen Messwert eingependelt hat. Ist $\Delta t \gg \tau$, kann man in der Regel von statistisch unabhängigen Messfehlern ausgehen.

Bemerkung 12. Bei einigen Messgeräten (z.B. bei digitalen Multimetern, auch bei Lock-in Verstärkern) kann man eine Zeitkonstante einstellen, die in der Regel bestimmt, wie lange das Messgerät mittelt, bevor es einen Wert ausgibt. Diese Zeitkonstante entspricht in etwa der Antwortzeit τ .

Bemerkung 13. Bei Multimetern der Firma HP wird die Zeitkonstante nicht in Sekunden, sondern in NPLC oder PLC angegeben. Das steht für ‘number of power line cycles’ und ist in der Schweiz mit der Netzfrequenz 50 Hz (Periode 20 ms) das entsprechende Vielfache dieser Schwingungsperiode. NPLC 10 bedeutet demnach eine Mittelung über $10 \times 20 \text{ ms} = 200 \text{ ms}$.

Bemerkung 14. Alle Messgeräte haben eine gewisse Bandbreite Δf , die in Hertz (Hz) angegeben wird. Die Zeitkonstante τ ist in etwa durch $1/2\pi\Delta f$ gegeben. (Umgekehrt haben Messgeräte mit einer Antwortzeit τ eine Bandbreite von etwa $\Delta f = 1/2\pi\tau$.)

Bemerkung 15. Es gibt graphische und auch quantitative Verfahren eine statistische Abhängigkeit der ϵ_i zu prüfen, bzw. quantitativ zu bestimmen. Diese Diskussion verschieben wir jedoch auf einen späteren Zeitpunkt. An dieser Stelle kennen wir die ϵ_i ja noch nicht, sondern beschreiben nur ihre statistischen Eigenschaften.

Statistisch abhängige Fehler ϵ_i . In diesem Fall ist

$$\text{pdf}(\{\epsilon_i\}|W) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(W)}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \epsilon_i W_{ij} \epsilon_j\right),$$

wobei N die Zahl der Datenpunkte ist und W die $N \times N$ Kovarianzmatrix der Fehler. Somit wird die Stichprobenverteilung

$$\text{pdf}(\{y_i\}|\{x_i\}, \theta, W) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(W)}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (y_i - f(x_i; \theta)) W_{ij} (y_j - f(x_j; \theta))\right).$$

Der Fall statistisch unabhängiger Fehler ergibt sich daraus als Spezialfall mit $W_{ij} = \delta_{ij}/\sigma_i^2$.

Bemerkung 16. Dieser Fall kommt in den Praktika des Physikstudiums eher selten vor, kann aber im Labor durchaus eine wichtige Rolle spielen.