

Statistische Methoden
zur
Datenauswertung
im Fortgeschrittenen-Praktikum

PHYSIKALISCHES INSTITUT
RUPRECHT-KARLS-UNIVERSITÄT
HEIDELBERG

MARKUS KÖHLI
3/2014



Inhaltsverzeichnis

1	Einführung	3
1.1	Vorwort	3
1.2	Acknowledgements	3
1.3	Literatur	4
2	Grundlagen	5
2.1	Wahrscheinlichkeitsdichte	5
2.2	Erwartungswert	6
2.3	Zentrale Momente und Varianz	6
2.4	Funktionen von zwei Zufallsvariablen	8
2.5	Weitere Parameter: Median und Quantile	9
2.6	Beschreibung diskreter Daten	10
3	Verteilungsfunktionen	11
3.1	Binomialverteilung	11
3.2	Poisson-Verteilung	11
3.3	Gauß-Verteilung	12
3.3.1	Zentraler Grenzwertsatz	12
3.3.2	Gauß-Verteilung in zwei Parametern	13
3.4	Gleichverteilung	13
3.5	Breit-Wigner-Verteilung	14
3.6	Faltung von Verteilungen	15
4	Fehlerbehandlung	16
4.1	σ und Fehler des Mittelwertes	16
4.2	Fehlerfortpflanzung	17
4.3	Kovarianz und Korrelation	17
4.4	Kovarianzmatrix	18
4.5	Kovarianzmatrix systematischer Fehler	19
4.6	Schätzer	19
4.7	χ^2 -Verteilung	21
4.8	Methode der kleinsten Quadrate	23
4.9	Konfidenzintervalle	25
4.10	Maximum Likelihood Methode	27

1 Einführung

1.1 Vorwort

‘Zu jedem Messwert gehört ein Fehler’ - eine eherne Regel, welche spätestens nach dem ersten Semester verinnerlicht sein dürfte. So intuitiv wie auch logisch begründbar liegt das, dass ein Infragestellen bestenfalls metaphysisch erscheint.

In der Realität aber angekommen fällt dieses Prinzip oft zuerst den unterschiedlichsten Begründungen zum Opfer, welche nicht selten auf unvollständiger Kenntnis des Experimentes beruhen, meist jedoch mit dem qualitativen Erfolg gerechtfertigt werden.

Hubble¹ begründete die These über die Expansion des Universums auf der Beobachtung von umliegenden Galaxien im Nahbereich, welcher in astronomischen Maßstäben zu klein, beziehungsweise durch andere Relativbewegungen überlagert war. Die Behauptung stellte sich im Nachhinein zu seinen Gunsten als richtig heraus. Falsch dagegen waren zahllose Entdeckungen wie die des Polywassers², der Polymerstruktur des Wassers, durch Fedyaikin oder dass Krebs eine Infektionskrankheit sei³, wofür gar der Nobelpreis verliehen wurde.

Weit gefehlt, wer diese Beispiele einer eher qualitativen historischen Arbeitsweise zuschreibt, welche heute undenkbar erschiene. Die statistisch korrekte und nachvollziehbare Arbeitsweise fundiert heute wie damals das Gebäude der Wissenschaften. Da aber die Abkehr von diesem Grundgedanken immer wieder in verschiedener Façon in Erscheinung tritt, sollte ein dezenter Hinweis auf deren Wichtigkeit ab und an aufs Neue erfolgen⁴ - ganz im Sinne des Experimentators. Diesem Impetus will das Skript folgen.

1.2 Acknowledgements

Das vorliegende Skript basiert auf der Vorlesung ‘*Statistische Methoden im Fortgeschrittenen-Praktikum*’, gehalten von Volker BÜSCHER an der Albert-Ludwigs-Universität Freiburg.

¹ HUBBLE, E.P.: *A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae*, (1929) Proc. Natl. Acad. Sci. USA 15, S. 168-173

² FEDYAKIN, N.N.: *Change in the Structure of Water during Condensation in Capillaries*, (1962) Kolloid Zhurnal 24, S. 497

³ FIBIGER, J.: *Untersuchungen über eine Nematode (Spiroptera sp. n.) und deren Fähigkeit papillomatöse und carcinomatöse Geschwulstbildungen im Magen der Ratte hervorzurufen*, (1913) Zeitschrift für Krebsforschung, Band 13, Nummer 2, S. 217-280

⁴ VAUX, D.L.: *Research methods: Know when your numbers are significant*, (2012) Nature 492, S. 180-181

1.3 *Literatur*

Weiterführend sind folgende Werke empfohlen:

COWAN, G.: *Statistical Data Analysis*, Oxford Science Publications

BRANDT, S.: *Datenanalyse*, Spektrum Akademischer Verlag

BARLOW, R.J.: *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley-VCH

2 Grundlagen

2.1 Wahrscheinlichkeitsdichte

Die Wahrscheinlichkeit P einen Wert im Intervall $[x, x + dx]$ zu finden ist gegeben durch die **Wahrscheinlichkeitsdichtefunktion** $f(x)$ ('pdf' von engl. *probability density function*)

$$P(x' \in [x, x + dx]) = f(x) \cdot dx.$$

Die pdf folgt dabei der Normierungsbedingung

$$\int_{\Omega} f(x) dx = 1$$

auf dem Ereignisraum Ω

Die Wahrscheinlichkeit einen Wert x' unterhalb $b \in \Omega[-\infty, \infty]$ zu finden beträgt:

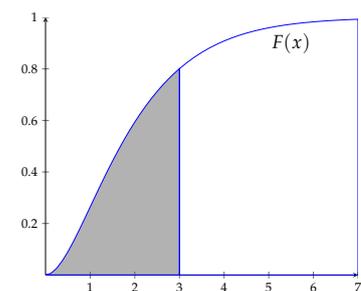
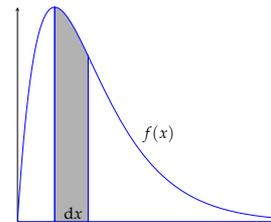
$$P(x' \leq b) = \int_{-\infty}^b f(x) dx = F(b),$$

respektive den Wert x' innerhalb eines Intervalls $[a, b]$ zu finden beträgt:

$$P(a \leq x' \leq b) = \int_a^b f(x) dx = F(b) - F(a).$$

Mit der so gebildeten Stammfunktion erhält man die **kumulative Verteilung** $F(x)$ von $f(x)$:

$$F(x) := \int_{-\infty}^x f(x') dx'.$$



$$\lim_{x \rightarrow \infty} F(x) = 1$$
$$\lim_{x \rightarrow -\infty} F(x) = 0$$

2.2 Erwartungswert

Der Erwartungswert $E(x)$ einer Zufallsvariable x mit Wahrscheinlichkeitsdichte $f(x)$ wird gebildet durch:

$$E(x) := \mu = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

Der Erwartungswert $E(h(x))$ einer beliebigen Funktion $h(x)$ wird gebildet durch:

$$E(h(x)) = \int_{-\infty}^{\infty} h(x) \cdot f(x) dx.$$

2.3 Zentrale Momente und Varianz

Erwartungswerte der Funktionen

$$h_l(x) = (x - c)^l$$

werden als **l-te Momente** der Variablen x und Punkt c bezeichnet.

Den **Spezialfall** stellen die l-ten Momente α_l um den Erwartungswert μ dar. Diese sind:

$$\alpha_0 = \int_{-\infty}^{\infty} (x - \mu)^0 f(x) dx = 1,$$

Normierung

$$\alpha_1 = \int_{-\infty}^{\infty} (x - \mu)^1 f(x) dx = \int_{-\infty}^{\infty} x f(x) dx - \mu = 0,$$

(Erwartungswert)

$$\alpha_2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Varianz

Sie enthält Information über die Streuung der Variablen x um den Mittelwert μ

Die Varianz ergibt sich nach

$$\alpha_2 = E \left((x - \mu)^2 \right) = \text{Var}(x) = \sigma^2(x).$$

Die Quadratwurzel aus der Varianz $\sqrt{\text{Var}(x)}$ entspricht der **Standardabweichung** $\sigma(x)$. Mittelwert und Standardabweichung sind allgemein die wichtigsten Größen zur statistischen Beschreibung einer Messreihe. Die Messunsicherheit in einem Experiment, der Messfehler, wird mit der Standardabweichung identifiziert.

Die Varianz kann folgendermaßen über die Erwartungswerte ausgedrückt werden:

$$\sigma^2(x) = E \left((x - \mu)^2 \right) = E(x^2) - (E(x))^2.$$

Höhere Momente

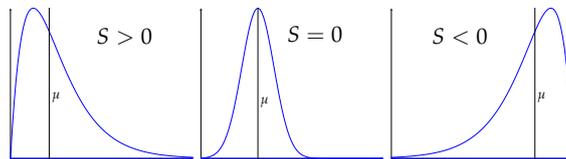
Die Momente α_l mit $l > 2$ werden als höhere Momente bezeichnet. Höhere Potenzen gewichten mehr und mehr die Randwerte der Verteilung.

$$\alpha_3 = E\left((x - \mu)^3\right).$$

Die Größe

$$S := \alpha_3 / \sigma^3$$

wird **Schiefe** genannt. Sie ist das erste von Null verschiedene ungerade zentrale Moment. Sie gewichtet Werte rechts und links des Erwartungswertes mit unterschiedlichem Vorzeichen. Ergibt S genau Null handelt es sich um eine symmetrische Verteilung. Für $S < 0$ bezeichnet man eine Verteilung als linksschief, das bedeutet sie fällt nach links langsamer ab als nach rechts. Bei $S > 0$ spricht man von rechtsschief.



$$\alpha_4 = E\left((x - \mu)^4\right)$$

Die **Kurtosis** K wird definiert über

$$K := \alpha_4 / \sigma^4 - 3.$$

Das vierte zentrale Moment einer Gauß-Verteilung beträgt genau 3. Durch die Subtraktion von 3 wird die Kurtosis darauf normiert, in wie fern eine Verteilung schmaler ($K < 0$), das heißt stärker zentriert um den Mittelwert liegt, oder breiter ($K > 0$) als eine Gauß-Verteilung ist.

Schiefe

Sie gibt ein Maß für die Asymmetrie der Verteilung von x an

Kurtosis

Durch die Gewichtung von 4 im Exponenten beschreibt sie die Verteilung von x an den Randwerten.

von gr. *κρῑτωσις* = Wölben

2.4 Funktionen von zwei Zufallsvariablen

Für eine zweidimensionale Wahrscheinlichkeitsdichtefunktion $f(x, y)$ der Zufallsvariablen x, y gilt:

$$P(x' \in [x, x + dx], y' \in [y, y + dy]) = f(x, y) \cdot dx dy$$

mit der Normierungsbedingung:

$$\int \int_{\Omega} f(x, y) dx dy = 1.$$

Die eindimensionalen Projektionen sind die **Randwertverteilungen**:

$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx,$$

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy,$$

mit den Erwartungswerten μ_x und μ_y .

Der Erwartungswert einer zweidimensionalen Funktion $h(x, y)$ wird analog zum eindimensionalen Fall gebildet:

$$E(h(x, y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dx dy.$$

Die Varianzen bezogen auf eine Variable ergeben sich nach:

$$\sigma^2(x) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x, y) dx dy,$$

$$\sigma^2(y) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_y)^2 f(x, y) dx dy.$$

Betrachtet man beide Variablen gleichzeitig, spricht man von der **Kovarianz**:

$$\left[\text{cov}(x, y) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x) (y - \mu_y) f(x, y) dx dy. \right]$$

Kovarianz

Die Kovarianz beschreibt damit die Verteilung nicht nur in Hinsicht auf ihre Breite, sondern auch auf die relative Lage der Ausdehnung in x, y -Richtung. Man spricht davon, in wie weit x und y korreliert sind und betrachtet den auf die relativen Breiten $\sigma_{x,y}$ normierten **Korrelationskoeffizienten** ρ :

Korrelation

$$\rho(x, y) := \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)}.$$

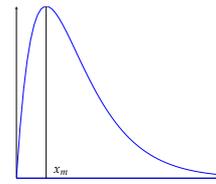
ρ nimmt Werte von -1 (negative Korrelation) bis 1 (positive Korrelation) an. In diesen Fällen führt eine Schwankung in x zu einer ebenso großen in y , für negative ρ entgegen gesetzt. Für $\rho = 0$ sind beide Wahrscheinlichkeitsdichten unkorreliert. Die Verteilung ist unabhängig in beiden Variablen.

2.5 Weitere Parameter: Median und Quantile

- Bei dem **wahrscheinlichsten Wert** x_m liegt das Maximum einer Verteilung.

Es gilt:

$$f(x_m) = \max_{x \in \Omega} |f(x)|.$$

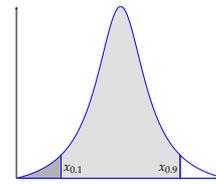


- Der **Median** gibt den Wert $x_{0.5}$ an, für welchen die kumulative Verteilungsfunktion $F(x)$ den Wert $1/2$ annimmt:

$$F(x_{0.5}) = \int_{-\infty}^{x_{0.5}} f(x) dx = 0.5.$$

- Verallgemeinert gibt das **Quantil** den Wert x_q für beliebige Bruchteile $q \leq 1$ an:

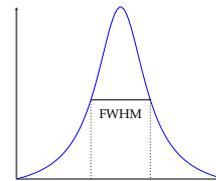
$$F(x_q) = \int_{-\infty}^{x_q} f(x) dx = q.$$



- Das **Full-Width-Half-Maximum (FWHM)** gibt die Breite der Verteilung auf halber Höhe des Maximums an. So werden Ausreißer im Randbereich ignoriert.

Für die Gauß-Verteilung gilt:

$$\text{FWHM} = 2.35 \sigma.$$



2.6 Beschreibung diskreter Daten

Die einem Experiment entstammenden Daten liegen in Form eines Datensatzes x_1, \dots, x_N vor. Die zugrundeliegende Wahrscheinlichkeitsdichte $f(x)$ ist im Allgemeinen nicht bekannt.

Verteilung und Parameter der Verteilung müssen aus den Messdaten bestimmt werden.

	Grundgesamtheit	Datensatz
Wahrscheinlichkeitsdichte	$f(x), F(x)$	$h(x)$ (Häufigkeitsvert.)
Erwartungswert	$E(x) = \mu$	Mittelwert \bar{x}
Varianz	$\sigma^2(x) = \text{Var}(x)$	Varianz $\sigma^2(x_1, \dots, x_N)$

Mittelwert:

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$$

Der Mittelwert stellt eine unverzerrte Schätzung des wahren Erwartungswertes μ dar.

$$\lim_{N \rightarrow \infty} \bar{x} = \mu$$

Varianz:

$$\text{Var}(x_1, \dots, x_N) = \sigma_N^2(x_i, \mu) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

im Allgemeinen ist der Erwartungswert μ a priori nicht bekannt, so dass μ mit Hilfe des Mittelwertes \bar{x} abgeschätzt werden muss. Die Größe $\sigma_N^2(x_i) = 1/N \sum_{i=1}^N (x_i - \bar{x})^2$ stellt jedoch keine unverzerrte Schätzung der zugrundeliegenden Grundgesamtheit dar.

Ein Freiheitsgrad der Datenmenge entfällt für die Berechnung des Mittelwertes, so dass der unverzerrte Schätzer gegeben ist durch:

$$\sigma_{N-1}^2(x_i) := \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2.$$

Kovarianz:

$$\text{cov}(x, y) := \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

3 Verteilungsfunktionen

3.1 Binomialverteilung

Beschreibung eines Experimentes mit zweiwertigem Ausgang A und \bar{A} :

$$P(A) = p,$$

$$P(\bar{A}) = 1 - p = q.$$

Bei der Durchführung von n unabhängigen Versuchen beträgt die Wahrscheinlichkeit dafür, dass Ereignis A insgesamt k -mal eintritt:

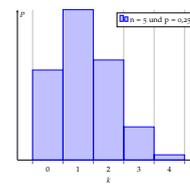
$$P(k, p, n) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Die Binomialverteilung hat folgende Eigenschaften:

$$E(k) = n \cdot p,$$

$$\sigma^2(k) = n \cdot p(1 - p),$$

$$\sigma(k) = \sqrt{n \cdot p(1 - p)}.$$



Binomialverteilung

Fehler der Binomialverteilung

3.2 Poisson-Verteilung

Die **Poisson-Verteilung**¹ beschreibt Prozesse, welche durch eine Binomialverteilung bei großem n und kleinem p beschrieben werden. Das Produkt $n \cdot p = \lambda$, welches für die mittlere Anzahl an Ereignissen im Beobachtungszeitraum steht, soll konstant bleiben.

$$\lim_{n \rightarrow \infty} P(k, p, n) = P(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Beispiel: radioaktiver Zerfall von 1 mg ^{238}U : ca. $2,5 \cdot 10^{18}$ Atome, $T_{1/2} = 4,47 \cdot 10^9$ a.

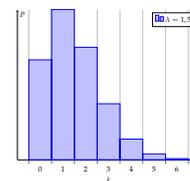
$P(k, \lambda)$ heißt Wahrscheinlichkeitsdichte der Poisson-Verteilung. Es gilt:

$$\sum_{k=0}^{\infty} P(k, \lambda) = 1,$$

$$E(k) = \lambda,$$

$$\sigma^2(k) = \lambda.$$

¹ nach Siméon Denis POISSON, Frankreich, Mathematiker und Physiker



Normierung

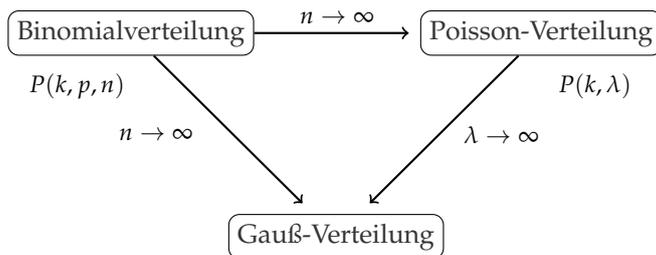
Erwartungswert

Varianz

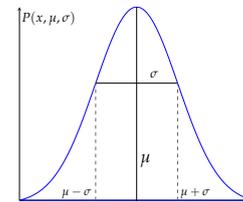
3.3 Gauß-Verteilung

Die **Gauß-Verteilung**² oder auch **Normalverteilung** ist in der Physik von zentraler Bedeutung. Durch sie lassen sich in guter Näherung Abweichungen von Messwerten ausdrücken. Daher fußt sowohl die Fehlerbeschreibung wie auch die Fehlerrechnung zu großen Teilen auf der Gauß-Verteilung.

Sie ergibt sich aus der Binomialverteilung für große Stichprobenmengen n und aus der Poisson-Verteilung für große Erwartungswerte λ .



² nach Johann Carl Friedrich GAUSS, Heiliges Römisches Reich Deutscher Nation, Mathematiker, Astronom, Geodät und Physiker



Im Gegensatz zu beiden genannten Funktionen besitzt die Gauß-Verteilung ein kontinuierliches Spektrum und ist symmetrisch. Mittelwert und Varianz werden als unabhängige Parameter behandelt. Ihre Wahrscheinlichkeitsdichte $P(x, \mu, \sigma)$ wird wie folgt beschrieben:

$$P(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Für die Wahrscheinlichkeitsdichte gilt:

$$\int_{-\infty}^{\infty} P(x, \mu, \sigma) = 1,$$

$$E(x) = \mu,$$

$$\text{Var}(x) = \sigma^2.$$

Gauß-Verteilung

Normierung

Erwartungswert

Varianz

3.3.1 Zentraler Grenzwertsatz

Sind x_i unabhängig verteilte Zufallsvariablen mit Mittelwert μ und Varianz σ^2 , so ist die Summe

$$X := \frac{1}{n} \sum_{i=1}^n x_i$$

im Grenzfalle $n \rightarrow \infty$ normalverteilt mit Mittelwert μ und Varianz σ^2/n

3.3.2 Gauß-Verteilung in zwei Parametern

Seien die **unabhängigen** Variablen x, y normalverteilt und sei o.B.d.A. $\mu_x = \mu_y = 0$, dann ist die Wahrscheinlichkeitsdichte

$$P(x, y) = P(x)P(y)$$

separabel in beiden Variablen und es gilt:

$$P(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)}$$

Analog zum eindimensionalen Fall ergibt sich die $1-\sigma$ -Variation für $P(x, y) = P(0, 0)e^{-1/2}$ zu:

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = 1.$$

Die Konturen konstanter Fehlerwahrscheinlichkeit sind Ellipsen.

In Vektordarstellung ausgedrückt wird die Ellipsengleichung für einen Vektor $\vec{x} = (x, y)^T$ zu:

$$(x, y) \begin{pmatrix} \frac{1}{\sigma_x^2} & 0 \\ 0 & \frac{1}{\sigma_y^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1,$$

das heißt

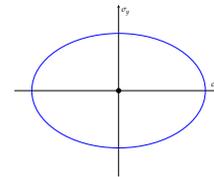
$$\vec{x}^T \mathbf{B} \vec{x} = 1.$$

Die Matrix \mathbf{B} ist die Inverse der Kovarianzmatrix \mathbf{C} ³

Die Wahrscheinlichkeitsdichte einer zweidimensionalen Gauß-Verteilung kann dann wie folgt ausgedrückt werden⁴:

$$P(x, y) = \frac{1}{2\sqrt{\det \mathbf{C}}} e^{-\frac{1}{2}\vec{x}^T \mathbf{C}^{-1} \vec{x}}.$$

Ellipsengleichung



³ für welche hier noch unkorrelierte x, y angenommen sind

⁴ das gilt auch für nicht verschwindende Kovarianzen, siehe Kapitel 4

3.4 Gleichverteilung

Durch die **Gleichverteilung** kann eine Vielzahl von Messungen beschrieben werden. So stellt sie den einfachsten Fall eines Detektors dar, welcher auf dem Ereignisraum in dem Intervall $[a, b]$ eine homogene Ansprechwahrscheinlichkeit besitzt.

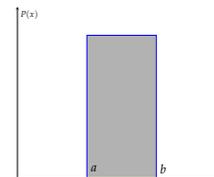
$$P(x) = \begin{cases} c = \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Sie besitzt die folgenden Eigenschaften:

$$E(x) = \frac{1}{2}(b - a),$$

$$\sigma^2(x) = \frac{(b - a)^2}{12}.$$

Die Standardabweichung einer Gleichverteilung insbesondere beträgt $\sigma = \frac{b-a}{\sqrt{12}}$ und nicht etwa $0.5 \cdot (b - a)$.



Erwartungswert

Varianz

3.5 Breit-Wigner-Verteilung

Durch eine **Breit⁵-Wigner⁶-Verteilung** oder auch **Lorentz-Kurve⁷** können Resonanzen beschrieben werden. Dies ist insbesondere dann relevant, wenn die natürliche Linienbreite aufgelöst werden kann, wie etwa bei Spektrallinien oder Energiespektren kurzlebiger Teilchen. Der Grenzfall für eine unverschobene Kurve $a = 0$ mit einer Halbwertsbreite von $\Gamma/2 = 1$ wird auch als Cauchy-Verteilung⁸ bezeichnet.

$$P(x, a, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(x - a)^2 + \left(\frac{\Gamma}{2}\right)^2}.$$

Sie besitzt die folgenden Eigenschaften:

$$E(x) = a,$$

$$\sigma^2(x) = \infty = \int_{-\infty}^{\infty} x^2 P(x) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x^2}{1 + x^2} dx.$$

Die Varianz und alle höheren Momente divergieren, da die Funktion nicht schnell genug abfällt. Daher wird für die Breit-Wigner-Verteilung die Breite durch das Full-Width-Half-Maximum (FWHM) angegeben:

$$\text{FWHM} := |x_2 - x_1| = \Gamma \quad f(x_1) = f(x_2) = \frac{1}{2}f(x)$$

⁵ nach Gregory BREIT, Ukraine, Physiker

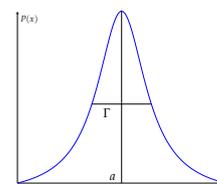
⁶ nach Eugene WIGNER, Königreich Ungarn, Physiker

⁷ Hendrik Antoon LORENTZ, Niederlande, Mathematiker und Physiker

⁸ nach Augustin-Louis CAUCHY, Frankreich, Mathematiker

Erwartungswert

Varianz



3.6 Faltung von Verteilungen

Eine **Faltung** oder auch **Konvolution**⁹ bewirkt, dass ein Funktionswert durch das durch eine zweite Funktion gewichtete Mittel der ihn umgebenden Werte ersetzt wird. Sie drückt damit die Auswirkung des Auflösungsvermögens der Apparatur auf eine Observable aus - wie die Unschärfe des Messfehlers auf die physikalische Wahrscheinlichkeitsdichte zu beschreiben ist.

⁹ von lat. *convolvere* = zusammenrollen

Seien Wahrscheinlichkeitsdichte der Observablen $f_a(x)$ und des Messfehlers $f_b(x)$ und ergebe sich der Messwert $u = a + b$, dann beschreibt das Faltungsintegral die Wahrscheinlichkeitsdichte $f_u(r)$

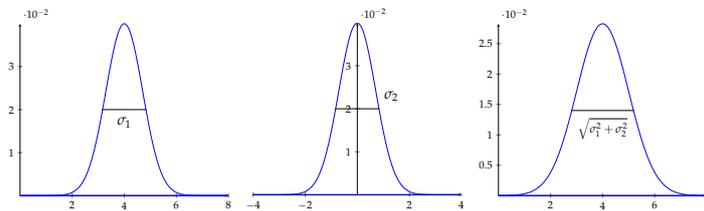
$$f_u(r) = \int_{-\infty}^{\infty} f_a(x)f_b(r-x) dx = \int_{-\infty}^{\infty} f_a(r-x)f_b(x) dx.$$

Ein wichtiger Fall ist die Faltung mit einer Gauß-Verteilung:

$$f_u(r) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} f(x)e^{-\frac{1}{2}\left(\frac{r-x}{\sigma}\right)^2} dx.$$

Wichtig: Die Faltung zweier Gauß-Funktionen ergibt ebenfalls eine Gauß-Funktion mit doppelter Varianz:

$$\sigma_{\text{neu}}^2 = \sigma_1^2 + \sigma_2^2.$$



Die Faltung einer Exponentialfunktion mit einer Gauß-Funktion ergibt eine Exponentialfunktion mit modifiziertem Exponenten.

4 Fehlerbehandlung

Vorbemerkung: (Mess-)Fehler, wie sie im physikalischen Sinne aufgefasst werden, unterliegen in ihrer Streuung um den wahren Wert einer Gauß-Verteilung. Daher wird der (Mess-)Fehler s mit der Standardabweichung σ einer Gauß-Verteilung identifiziert.

4.1 σ und Fehler des Mittelwertes

Für N wiederholte Messungen x_1, \dots, x_N mit Fehlern $\sigma_i = \sigma$ wird der **arithmetische Mittelwert** \bar{x} definiert als

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

arithmetisches Mittel

Der Fehler des arithmetischen Mittelwertes ergibt sich nach

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{N}}.$$

Sind die Genauigkeiten der Messpunkte individuell, existieren also Fehler σ_i , so wird das **gewichtete Mittel** \bar{x}_G verwendet:

$$\bar{x}_G = \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}.$$

gewichtetes Mittel

Der Fehler des gewichteten Mittels beträgt:

$$\sigma^2(\bar{x}_G) = \frac{1}{\sum \frac{1}{\sigma_i^2}}.$$

Diese Mittelwertbildung wird häufig zur Kombination von unterschiedlichen unabhängigen Messungen verwendet.

4.2 Fehlerfortpflanzung

Seien Observable x, y aufgrund ihrer statistischen Natur normalverteilt um die wahren Werte μ_x, μ_y mit Varianzen σ_x^2, σ_y^2 . Die Messwerte dieser Observablen seien \bar{x}, \bar{y} .

Um für eine allgemeine Funktion $f(x, y)$ den Fehler abzuschätzen, soll als Maß die Varianz von $f(\bar{x}, \bar{y}) - f(\mu_x, \mu_y)$ dienen. Mit der **Gauß'schen Fehlerfortpflanzung** schätzt man den Gesamtfehler durch eine Taylorentwicklung erster Ordnung ab:¹

$$f(\bar{x}, \bar{y}) = f(\mu_x, \mu_y) + \left. \frac{\delta f}{\delta x} \right|_{\mu_x, \mu_y} (\bar{x} - \mu_x) + \left. \frac{\delta f}{\delta y} \right|_{\mu_x, \mu_y} (\bar{y} - \mu_y) + (\dots).$$

Die Größe

$$f(\mu_x, \mu_y) - f(\bar{x}, \bar{y}) = \left. \frac{\delta f}{\delta x} \right|_{\bar{x}, \bar{y}} (\mu_x - \bar{x}) + \left. \frac{\delta f}{\delta y} \right|_{\bar{x}, \bar{y}} (\mu_y - \bar{y})$$

ist dann normalverteilt um Erwartungswert 0 und besitzt die Varianz:

$$\text{Var} [f(\mu_x, \mu_y) - f(\bar{x}, \bar{y})] = \left. \frac{\delta f}{\delta x} \right|_{\bar{x}, \bar{y}}^2 \sigma^2(x) + \left. \frac{\delta f}{\delta y} \right|_{\bar{x}, \bar{y}}^2 \sigma^2(y).$$

Der Fehler σ_f von $f(\bar{x}, \bar{y})$ ist gegeben durch

$$\sigma_f = \sqrt{\left(\left. \frac{\delta f}{\delta x} \right|_{\bar{x}, \bar{y}} \right)^2 \sigma^2(x) + \left(\left. \frac{\delta f}{\delta y} \right|_{\bar{x}, \bar{y}} \right)^2 \sigma^2(y)},$$

falls x und y unabhängig sind.

¹ Unter den Annahmen dass:

- höhere Terme vernachlässigbar
- partielle Ableitungen um μ_x, μ_y konsistent
- Ableitungen an \bar{x}, \bar{y} anstatt μ_x, μ_y

4.3 Kovarianz und Korrelation

Falls x und y **nicht unabhängig** voneinander sind, muss in der Fehlerfortpflanzung deren Korrelation ρ berücksichtigt werden. Sie gibt an, wie sich die Änderung des einen Parameters zum anderen verhält und kann ebenfalls über die multivariate Varianz berechnet werden. Es gilt für die sogenannte **Kovarianz**

$$\text{cov}(x, y) = \rho \sigma_x \sigma_y,$$

Kovarianz

sodass sich die Varianz von f berechnet zu

$$\sigma^2 (f(\bar{x}, \bar{y})) = \sigma^2(f(x, y)) = \left(\left. \frac{\delta f}{\delta x} \right|_{\bar{x}, \bar{y}} \right)^2 \sigma^2(x) + \left(\left. \frac{\delta f}{\delta y} \right|_{\bar{x}, \bar{y}} \right)^2 \sigma^2(y) + 2 \left(\left. \frac{\delta f}{\delta x} \right|_{\bar{x}, \bar{y}} \right) \left(\left. \frac{\delta f}{\delta y} \right|_{\bar{x}, \bar{y}} \right) \underbrace{\text{cov}(x, y)}_{\rho \sigma(x)\sigma(y)}.$$

Die Kovarianz ergibt sich dabei entweder

- aus dem experimentellen Aufbau. Eine Arrangement aus zwei Detektoren etwa, bei dem ein Ereignis entweder im einen oder zwingend im anderen Detektor erfasst wird, bedeutet eine Korrelation von $\rho = -1$ auf die Zählraten.

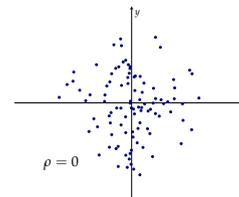
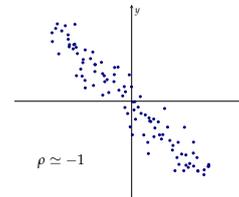
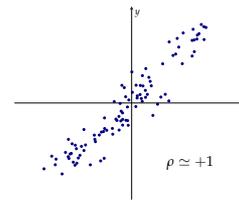
- aus einem bekannten funktionalen Zusammenhang. Dies ist insbesondere dann relevant wenn aus einer Kurvenanpassung mehrere Fitparameter verwendet werden, da diese in den seltensten Fällen unkorreliert sind.
- aus der Streuung der Daten. Die Korrelation wird dann wie in Kapitel 2.6 beschrieben berechnet.

Einige Beispiele für Korrelationen:

Es gilt:

$$-1 \leq \rho \leq +1.$$

Im Falle von $\rho > 0$ spricht man von positiver Korrelation, im Falle von $\rho < 0$ von negativer Korrelation. Ist $\rho = 0$ sind die Variablen unkorreliert.



4.4 Kovarianzmatrix

Für den Fall, bei welchem eine Funktion f auf einer Parametermenge x_1, \dots, x_n basiert, wird die bivariate Abhängigkeit der Parameter durch die Kovarianzmatrix C ausgedrückt. Diese enthält auf ihrer Diagonalen die Varianzen der Messgrößen x_i und die Kovarianzen $\text{cov}(x_i, x_j)$ auf der Nebendiagonalen.

$$C = \begin{pmatrix} \sigma_{x_1}^2 & & \text{cov}(x_i, x_j) \\ & \ddots & \\ \text{cov}(x_j, x_i) & & \sigma_{x_n}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 & & \rho \sigma_{x_i} \sigma_{x_j} \\ & \ddots & \\ \rho \sigma_{x_j} \sigma_{x_i} & & \sigma_{x_n}^2 \end{pmatrix}.$$

In linearer Näherung ist dann der Fehler auf f wie folgt zu berechnen

$$\sigma^2(f(x, y)) = \sum_{i,j=1}^n \left(\frac{\delta f}{\delta x_i} \frac{\delta f}{\delta x_j} C_{ij} \right)$$

oder in Vektorform:

$$\sigma^2(f(x, y)) = \nabla f^T \cdot C \cdot \nabla f.$$

Für den allgemeinen Fall, dass ein Satz von Funktionen f_1, \dots, f_m auf einer Parametermenge x_1, \dots, x_n basiert, ergibt sich für jedes Paar f_k, f_l eine (Ko-)Varianz der x_i, x_j , sodass der Fehler $\sigma^2(f(x, y))$ zur Fehlermatrix E_{kl} erweitert wird:

$$E_{kl} = \sum_{i,j=1}^n \left(\frac{\delta f_k}{\delta x_i} \frac{\delta f_l}{\delta x_j} \underbrace{\rho_{ij} \sigma_{x_i} \sigma_{x_j}}_{C_{ij}} \right).$$

Mit der Transformationsmatrix G

$$G_{ki} := \frac{\delta f_k}{\delta x_i}$$

erhält man in Vektorform:

$$E = G \cdot C \cdot G^T.$$

$$\begin{aligned} \dim C &= (n, n) \\ \dim G &= (m, n) \\ \dim E &= (m, m) \end{aligned}$$

4.5 Kovarianzmatrix systematischer Fehler

Unter der Annahme zwei Messgrößen x_1, x_2 besitzen einen statistischen Fehler $\sigma_{x_1}, \sigma_{x_2}$ sowie einen systematischen Fehler s :

$$\begin{aligned}x_1 &\pm \sigma_{x_1} \pm s, \\x_2 &\pm \sigma_{x_2} \pm s,\end{aligned}$$

dann sind beide Größen korreliert:

$$\text{cov}(x_1, x_2) = E(x_1, x_2) - E(x_1)E(x_2) = s^2,$$

woraus für die Kovarianzmatrix folgt

$$\mathbf{C} = \begin{pmatrix} \sigma_{x_1}^2 + s^2 & s^2 \\ s^2 & \sigma_{x_2}^2 + s^2 \end{pmatrix} = \begin{pmatrix} \sigma_{x_1}^2 + s^2 & \text{cov}(x_1, x_2) + s^2 \\ \text{cov}(x_2, x_1) + s^2 & \sigma_{x_2}^2 + s^2 \end{pmatrix}.$$

4.6 Schätzer

Die Fragestellung bei der Auswertung eines Experimentes besteht darin, einen allgemeinen und simplen Ansatz zur Beschreibung einer Gesetzmäßigkeit mit den gewonnenen Messungen zu vergleichen. Die Verteilungsfunktion, welche die Daten beschreibt, ist im allgemeinen nicht bekannt - wohl aber ihre Form, welche durch einen Satz an Parametern $\vec{\lambda} = (\lambda_1, \dots, \lambda_n)$ festgelegt ist. Ein einfaches Beispiel ist der radioaktive Zerfall - der komplette Datensatz ist lediglich von der Variable Lebensdauer abhängig, ein einzelner Datensatz hat keine spezifische Information.

$$N(t) = N_0 \cdot e^{-t/\tau}$$

Aus der Stichprobe sind der mit Fehlern behaftete Schätzwert $\vec{\lambda}$ für $\vec{\lambda}$ und dessen Varianz $\sigma(\vec{\lambda})$ zu bestimmen. Letzteres soll ein Maß für die Güte der Anpassung darstellen, ein sogenanntes Konfidenzniveau angeben..

Die Anforderungen an einen Schätzer $S(x_1, \dots, x_n)$ sind, dass er folgende Eigenschaften besitzt:

- **unverzerrt:** Der Erwartungswert für den Schätzer eines Parameters soll (unabhängig von der Anzahl der Messwerte) mit dem Parameter übereinstimmen:

$$E[S_\lambda(x_1, \dots, x_n)] = \lambda.$$

- **konsistent:** Die Varianz des Schätzers soll für große Stichproben gegen Null gehen

$$\lim_{n \rightarrow \infty} \sigma^2(S) = 0.$$

- **effizient:** Die Streuung im Parameterraum soll möglichst gering sein, um die Informationen möglichst effektiv zu nutzen. Das erfordert, dass die Varianz minimal sein muss:

$$E[(S - \lambda)^2] = \sigma^2(S) < \sigma^2(S_i),$$

wobei S_i ein beliebiger Schätzer ist.

Erwartungswert der Gauß-Verteilung

Für eine Gauß-Verteilung ist der Mittelwert $\bar{x} = 1/n \sum x_i$ der unverzerrteste Schätzer für den Erwartungswert. Es lässt sich zeigen, dass dessen Varianz σ^2 ebenfalls minimal ist und der Mittelwert somit die oben genannten Kriterien erfüllt.

Für andere Schätzer des Erwartungswertes lässt sich zeigen, dass diese nicht dem Kriterium der Effizienz genügen, so etwa die unten aufgeführten Beispiele des Medians und dem Mittel aus größtem und kleinstem Wert im Datensatz. Diese sind jedoch für andere Annahmen über die Verteilungsfunktion effizient - so schätzt der Median für eine zweiseitige Exponentialverteilung den Erwartungswert besonders gut ab.

Schätzer für μ	Varianz	effizient für
Mittelwert	σ^2/N	Gauß-Verteilung
Median	$\sigma^2/N \cdot \pi/2$	zweiseitige Exponentialverteilung
$(x_{\min} + x_{\max})/2$	$\sigma^2/N \cdot N\pi^2/(12 \ln N)$	Gleichverteilung

Varianz der Gauß-Verteilung

Der Schätzer für die Varianz

$$\sigma^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$$

wäre mit dem Faktor $1/(n)$ keine unverzerrte Schätzung. Anschaulich wurde bereits ein Freiheitsgrad des Datensatzes für die Bestimmung des Mittelwertes \bar{x} verwendet, welcher für die Bestimmung weiterer Parameter nicht mehr zur Verfügung steht.

Korrelation

Für den empirischen Korrelationskoeffizienten

$$\hat{\rho} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

sind die besten Schätzer für den Fehler

$$\sigma(\hat{\rho}) = \frac{1 - \hat{\rho}^2}{\sqrt{N-1}} \quad \text{für } N > 50,$$

$$\sigma(\hat{\rho}) = \frac{1}{\sqrt{N-3}} \quad \text{für } N < 50.$$

Beispiele für Stichprobenfunktionen

4.7 χ^2 -Verteilung

Beim Vergleich von Signifikanzen verschiedener Modellierungen wird die Passgenauigkeit an die Daten durch die Abweichung der Messungen x_i von der Vorhersage μ , bezogen auf ihren jeweiligen Fehler σ , bestimmt. Man ist an der Stichprobenfunktion χ^2 interessiert:

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}.$$

wobei der Parameter μ in der Regel noch nicht bekannt ist und mit dem Mittelwert \bar{x} abgeschätzt wird. Man geht ferner davon aus, dass es sich bei x um eine normalverteilte Zufallsvariable handelt.

In anderen Worten: Die χ^2 -Verteilung besteht aus einer Summe $\chi_1^2 + \dots + \chi_n^2$ von normalverteilt streuenden Messwerten, welche auf ihren jeweiligen Mittelwert und Varianz normiert werden.

Die χ^2 -Verteilung selbst ist für $2\lambda = n$ Messungen verteilt nach

$$f(\chi^2) = \frac{1}{2^\lambda \Gamma(\lambda)} (\chi^2)^{\lambda-1} e^{-\frac{\chi^2}{2}}$$

mit der Eulerschen² Gammafunktion Γ

$$\Gamma(x + 1) = \int_0^\infty t^{x-1} e^{-t} dt$$

mit $\Gamma(x + 1) = x\Gamma(x),$
 $\Gamma(1) = 1,$
 $\Gamma(x + 1) = x!$

² nach Leonhard EULER, Schweiz, Mathematiker

Mit der Definition der empirischen Varianz s^2 folgt auch, dass

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma^2} = (n - 1) \frac{s^2}{\sigma^2}.$$

Der Erwartungswert der χ^2 -Verteilung entspricht somit der Anzahl der Freiheitsgrade n_F

$$E(\chi^2) = n - 1 = n_F.$$

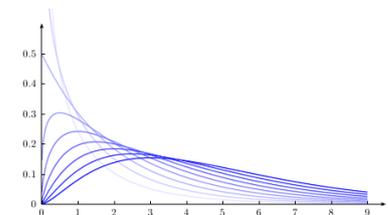
Die Varianz der χ^2 -Verteilung beträgt

$$\text{Var}(\chi^2) = 2n_F.$$

Ihre Verteilungsfunktion lautet:

$$F(\chi^2) = \int_0^{\chi^2} f(u) du.$$

Anzahl der Freiheitsgrade n_F :
 Anzahl der Messwerte n minus Anzahl der zu bestimmenden Parameter n_P
 $n_F = n - n_P$



χ^2 für $n = 1 \dots 5$ in Schritten von 0,5

Anwendung: Hypothesentest

Unter korrekter Annahme des Modells und rein statistischen Fehlern sollten Daten x_1, \dots, x_n χ^2 -verteilt sein. Die Wahrscheinlichkeit P , für eine gegebene Anzahl an n Freiheitsgraden einen schlechteren Wert als χ^2 zu erhalten ergibt sich aus der Verteilungsfunktion:

$$P(\chi_{n_f}^2) = 1 - F(\chi_{n_f}^2).$$

Man erwartet insbesondere, dass jeder Freiheitsgrad im Mittel die Varianz σ^2 besitzt und somit χ^2 um 1 erhöht:

$$\frac{\chi^2}{n} \approx 1.$$

Der Wert χ^2/doF ³ soll damit nahe bei 1 liegen und wird oft dazu verwendet, die Qualität eines Modells abzuschätzen. Falls sich ein deutlich von 1 verschiedener Wert ergibt, kann dies in mehrere Ursachen begründet liegen:

³ *dof = degrees of freedom*

- die Modellierung ist falsch oder unzureichend,
- die Gauß-Statistik für die Verteilung der Fehler ist falsch,
- es gibt unberücksichtigte systematische Fehler,
- die angenommene Standardabweichung σ ist zu groß (χ^2 wird zu klein) oder zu klein (χ^2 wird zu groß).

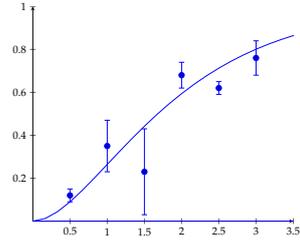
4.8 Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate soll einen unverzerrten und konsistenten Schätzer liefern. Unter aus einer Messung gegebenen Wertepaaren $x_i, y_i \pm \sigma_i$ wird die Funktion $f(x, \lambda)$ bestimmt, indem χ^2 **minimiert** wird. Dies liefert den besten Schätzwert für λ

$$\min(\chi^2) = \min \left(\sum_{i=1}^n \frac{(x_i - f(x_i, \lambda))^2}{\sigma_i^2} \right).$$

Für Parameter $\lambda_1, \dots, \lambda_n$ ergeben sich n Gleichungen für n Unbekannte

$$\frac{d\chi^2}{d\lambda_i} = 0.$$



Beispiel: **Ausgleichsgerade**

$$f(x, \lambda) = f(x, m, c) = mx + c$$

$$\chi^2 = \sum_{i=1}^n \frac{(x_i - mx_i - c)^2}{\sigma_i^2}.$$

Unter Annahme, dass $\sigma_i = \sigma$ für $i = 1 \dots n$ erhält man nach

$$\frac{d\chi^2}{dc} = \frac{1}{\sigma^2} (-2) \sum (y_i - mx_i - c) = 0,$$

$$\frac{d\chi^2}{dm} = \frac{x_i}{\sigma^2} (-2) \sum (y_i - mx_i - c) = 0,$$

als Schätzer

$$\hat{m} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{cov}(x, y)}{\sigma^2(x)}$$

$$\hat{c} = \bar{y} - \hat{m}\bar{x}$$

mit den Fehlern

$$\sigma^2(\hat{m}) = \sum \left(\frac{\delta \hat{m}}{\delta x_i} \right)^2 \sigma_i^2 \rightarrow \sigma^2(\hat{m}) = \frac{\sigma^2}{n(\overline{x^2} - \bar{x}^2)},$$

$$\sigma^2(\hat{c}) = \frac{\sigma^2 \bar{x}^2}{n(\overline{x^2} - \bar{x}^2)},$$

$$\text{cov}(\hat{m}, \hat{c}) = -\frac{\sigma^2 \bar{x}}{n(\overline{x^2} - \bar{x}^2)}.$$

Wichtig: Die Kovarianz von m und c hängt vom Erwartungswert \bar{x} ab. Durch geeignete Parametrisierung ergibt dieser Null und Steigung sowie Achsenabschnitt sind nicht mehr miteinander korreliert. Dies geschieht durch die Wahl der Geradendarstellung von

$$f(x, m, c) = m(x - \bar{x}) + c$$

und es ergeben sich die unkorrelierten Lösungen

$$\hat{m} = \frac{\overline{xy}}{x^2}, \quad \sigma^2(\hat{m}) = \frac{\sigma^2}{nx^2},$$

$$\hat{c} = \bar{y}, \quad \sigma^2(\hat{c}) = \frac{\sigma^2}{n}.$$

**Verallgemeinerung:
beliebige Funktionen und beliebige Fehlermatrizen**

Sei eine Messung bestehend aus Paaren

$$(x_i, y_i) = (\vec{x}, \vec{y}) \quad \text{mit } i = 1 \dots n,$$

welche nicht unabhängig sind, also die Kovarianzmatrix \mathbf{C} nicht verschwindende Nebendiagonalelemente $\text{cov}(y_i, y_j)$ besitzt.

Diese Messwerte sollen an ihren Stellen x_i beschrieben werden durch eine Funktion $\vec{f}(\vec{x}, \vec{\lambda})$, welche einen Satz an m Parametern $\vec{\lambda}$ besitzt.

Die χ^2 -Funktion lautet dann:

$$\chi^2 = (\vec{y} - \vec{f}(\vec{x}, \vec{\lambda}))^T \mathbf{C}^{-1} (\vec{y} - \vec{f}(\vec{x}, \vec{\lambda})).$$

Das χ^2 -Minimum wird bestimmt durch $\frac{d\chi^2}{d\lambda_i} = 0$, das heißt der Gradient von χ^2 verschwindet bezüglich der Parameter λ .

Es wird zusätzlich angenommen, dass \vec{f} linear in in den Parametern λ_i sei, gemäß

$$\vec{f}(\vec{x}, \vec{\lambda}) = \sum_{k=1}^m a_k(\vec{x}) \lambda_k,$$

dann lässt sich abkürzend schreiben

$$\vec{f}(\vec{x}, \vec{\lambda}) = \mathbf{A} \vec{\lambda} \quad \text{mit } A_{ij} = a_j(x_i),$$

sodass die χ^2 -Funktion geschrieben werden kann als

$$\chi^2 = (\vec{y}^T - \vec{\lambda}^T \mathbf{A}^T) \mathbf{C}^{-1} (\mathbf{A} \vec{\lambda}).$$

Die Lösung zur Bestimmung der Schätzer der Parameter aus der Minimierung von χ^2 lautet:

$$\vec{\lambda} = \underbrace{(\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A})^{-1}}_{\mathbf{G}} \mathbf{A}^T \mathbf{C}^{-1} \vec{y}$$

Die Kovarianzmatrix der Parameter \mathbf{C}_λ ergibt sich aus der Transformation der Kovarianzmatrix der Messerwerte:

$$\mathbf{C}_\lambda = \mathbf{G} \mathbf{C} \mathbf{G}^T,$$

womit man erhält:

$$\mathbf{C}_\lambda = (\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A})^{-1}.$$

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Kovarianzmatrix \mathbf{C} :

$$\mathbf{C} = \begin{pmatrix} \sigma_{y_1}^2 & & & \text{cov}(y_i, y_j) \\ & \ddots & & \\ \text{cov}(y_j, y_i) & & & \sigma_{y_n}^2 \end{pmatrix}$$

Parameter $\vec{\lambda} = (\lambda_1, \dots, \lambda_m)$

etwa:

$$f(\vec{x}, \vec{\lambda}) = \lambda_0 + \lambda_1 x + \lambda_2 x^2$$

nicht $f(x, \lambda) = e^{-\lambda t}$

$$\nabla_\lambda \chi^2 = -2\mathbf{A}^T \mathbf{C}^{-1} (\vec{y} - \mathbf{A} \vec{\lambda}) = 0$$

$$\mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} \vec{\lambda} = \mathbf{A}^T \mathbf{C}^{-1} \vec{y}$$

4.9 Konfidenzintervalle

Das Resultat einer Messung beträgt

$$\hat{\lambda} \pm \sigma(\hat{\lambda})$$

Was bedeutet das? Wie lautet die Interpretation des Ergebnisses?
Wo liegt der wahre Wert?

Der zu dem Messergebnis $\hat{\lambda}$ gehörende Fehler $\sigma(\hat{\lambda})$ soll eine Aussage über das Vertrauensintervall geben, in welchem das Resultat mit einer gewissen Wahrscheinlichkeit liegt. Da in der Regel die Wahrscheinlichkeitsdichten für λ nicht bekannt sind gibt es zwei verschiedene Herangehensweisen an die obigen Fragen.

Ein einfaches Beispiel stellt der radioaktive Zerfall dar. Der Fehler der Anzahl der gemessenen Zerfälle wird mit der Wurzel aus deren Mittelwert abgeschätzt. Damit hängt der Vertrauensbereich direkt von der (zufällig) bestimmten Anzahl an Ereignissen ab. Es könnte vernünftiger erscheinen anzunehmen es gäbe einen wahren Wert für das Ergebnis mitsamt der Wahrscheinlichkeitsdichte, gegeben durch die Wurzel dessen. Die Einschätzung der Belastbarkeit des Ergebnisses sollte dann in Hinblick auf diese vorab festgesetzte Wahrscheinlichkeit getroffen werden und niemals durch sich selbst?

Der Frequentist-Ansatz⁴:

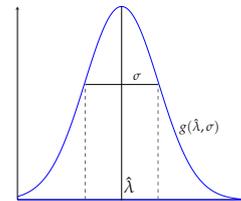
Es wird vorab der Ereignisraum Ω festgelegt, aber keine Aussagen über Wahrscheinlichkeiten getroffen. Die Wahrscheinlichkeit eines Ereignisses E wird nach der Messung bestimmt durch dessen relative Häufigkeit n in Bezug zur Gesamtheit G festgelegt:

$$P(E) \approx \frac{n_E}{n_G}$$

Es wird angenommen, dass sich im Limes für beliebig große Stichproben aus diesem Zufallsexperiment die so bestimmte *posterior*-Wahrscheinlichkeit dem wahren Wert annähert.

Der Bayes-Ansatz⁵:

Um ein Ergebnis interpretieren zu können muss dieses immer unter seiner vorab feststehenden Häufigkeit, der *A-priori*-Wahrscheinlichkeitsdichte, betrachtet werden. Diese wird durch den Experimentator festgesetzt und unterliegt damit seiner subjektiven Einschätzung. Die Validität des Bayesschen Wahrscheinlichkeitsstatistik hängt damit davon ab, wie vernünftig die Vorabannahmen getroffen werden können.

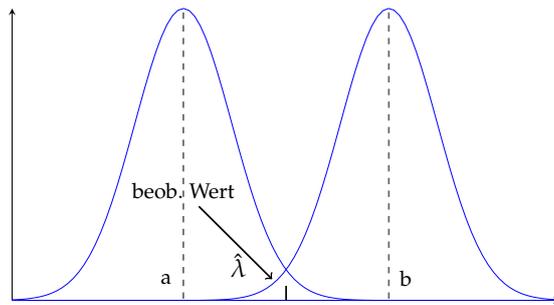


⁴ von engl. 'Häufigkeit'

⁵ nach Thomas BAYES, England, Mathematiker und Pfarrer

Konfidenzintervalle:

Ausgehend von einem Frequentist-Ansatz verwendet man um Vertrauensintervalle zu bestimmen nicht die Wahrscheinlichkeitsdichte der Parameter λ , sondern die Wahrscheinlichkeitsdichte der Messwerte bei festem Parameter. Es werden bei dieser Methode die Grenzen bestimmt, in welchen der Messwert mit der gesuchten Wahrscheinlichkeit liegt.



Das Konfidenzintervall $[a, b]$ wird so festgelegt, dass die Werte a, b gesucht werden, für welche nach

$$\int_{\hat{\lambda}_{\text{beob}}}^{\infty} g(\hat{\lambda}, a) d\hat{\lambda} = \alpha \quad \int_{-\infty}^{\hat{\lambda}_{\text{beob}}} g(\hat{\lambda}, b) d\hat{\lambda} = \beta$$

a stellt eine untere Grenze für den Wert λ dar
 b stellt eine obere Grenze für den Wert λ dar

die Ausschlusswahrscheinlichkeiten α, β existieren. Der wahre Wert liegt dann mit einer Wahrscheinlichkeit von $P = 1 - \alpha - \beta$ im Intervall $[a, b]$.

Das Konfidenzniveau zu P setzt sich so aus den einseitigen Konfidenzintervallen zusammen.

$\hat{\lambda}$ normalverteilt:

Unter der Annahme, dass $\hat{\lambda}$ normalverteilt ist, es sich dabei also um einen Parameter handelt der eine Observable mit Messfehler beschreibt, sind die Wahrscheinlichkeitsdichten symmetrisch und es gilt:

$$\alpha = \beta = \frac{\epsilon}{2} \Rightarrow 1 - \epsilon = \int_{\lambda - \delta}^{\lambda + \delta} e^{-\left(\frac{\hat{\lambda} - \lambda_0}{\sigma_\lambda}\right)^2} d\lambda_0 \frac{1}{\sqrt{2\pi} \sigma_\lambda}$$

Hieraus ergeben sich folgende Konfidenzintervalle in welchen der wahre Wert liegt:

δ	1σ	2σ	3σ
$1 - \epsilon$	68,3 %	95,4 %	99,7 %

4.10 Maximum Likelihood Methode

Während die Methode der χ^2 -Minimierung auf der Annahme normalverteilter Messwerte beruht, gilt die *Maximum Likelihood Methode* für allgemeine Wahrscheinlichkeitsdichten. Sie findet ihre Anwendung immer dann, wenn weniger die Betrachtung einzelner Messungen als vielmehr deren Gesamtheit an ein Modell angepasst werden soll. So etwa wenn ein Modell weniger Parameter - wahlweise als analytische Funktion oder aus Monte-Carlo-Simulationen - einem gesamten Histogramm einbeschrieben werden soll.

Seien Messwerte x_1, \dots, x_n gemäß einer Wahrscheinlichkeitsdichte $f(x, \vec{\lambda})$ verteilt, ist die Likelihood-Funktion L das Produkt der Auftrittswahrscheinlichkeiten eines jeden x_i :

$$L(x, \vec{\lambda}) = \prod_{i=1}^n f(x_i, \vec{\lambda}).$$

Likelihood-Funktion

Die beste Anpassung wird für den Parameter erreicht, für den das Produkt der Einzelwahrscheinlichkeiten am größten ist. Dieser findet sich, indem das Maximum der Likelihood-Funktion $\left| L(x, \vec{\lambda}) \right|_{\text{sup}}$ bestimmt wird.

Man bildet zusätzlich (aufgrund des Produktes vieler kleiner Zahlwerte) den Logarithmus der Likelihood-Funktion, die **Log-Likelihood-Funktion**

$$\mathcal{L}(x, \vec{\lambda}) = \ln L(x, \vec{\lambda}) = \sum_{i=1}^n \ln f(x_i, \vec{\lambda}).$$

Log-Likelihood-Funktion

Der beste Schätzer $\vec{\lambda}$ wird, wie oben ausgeführt, durch die Maximierungsbedingung an \mathcal{L} bestimmt:

$$\frac{\delta \mathcal{L}}{\delta \lambda_j} = \frac{\delta}{\delta \lambda_j} \ln L(x_i, \lambda_j) = 0,$$

$$\frac{\delta^2 \mathcal{L}}{\delta \lambda_j \delta \lambda_k} \Big|_{\lambda=\hat{\lambda}} = \text{negativ definit.}$$

Beispiel:

der radioaktive Zerfall: $f(t, \tau) = 1/\tau \cdot e^{-t/\tau}$ beinhaltet in seinem Modell als einzigen Parameter die Lebensdauer τ . Es werden n Zerfälle mit den Zeiten t_i gemessen. Die Log-Likelihood-Funktion lautet dann:

$$\mathcal{L}(t_1, \dots, t_n, \tau) = \sum_{i=1}^n \left(-\ln \tau - \frac{t_i}{\tau} \right).$$

Über die Maximierungsbedingung erhält man den Schätzer der Lebensdauer $\hat{\tau}$:

$$\frac{\delta \mathcal{L}}{\delta \tau} = \sum_{i=1}^n \left(-\frac{1}{\tau} + \frac{t_i}{\tau^2} \right) = 0 \implies \hat{\tau} = \frac{1}{n} \sum_{i=1}^n t_i = \bar{t}.$$

Der Schätzer für die Lebensdauer nach der Maximum Likelihood Methode ist gerade das arithmetische Mittel der gemessenen Zeiten.

Maximum Likelihood für Histogramme

Für geringe Stichprobengrößen, wie sie etwa in Histogrammen mit **wenigen Einträgen pro Bin** vorkommen, kann man nicht von einer Normalverteilung ausgehen, um einen Schätzwert für die wahre Anzahl an Einträgen abzugeben (da für eine Gauß-Funktion nicht zuletzt die untere Konfidenzgrenze leicht unter Null fallen kann). Es muss entweder eine Binomialverteilung angenommen oder eine Poisson-Statistik vorausgesetzt werden, siehe Kapitel 3.

Da eine χ^2 -Minimierung dies nicht leistet, kann die Maximum Likelihood Methode verwendet werden. Man nimmt an, dass die Stichprobenmenge N_i in jedem Bin i Poisson-verteilt ist, wobei der Mittelwert der Poisson-Verteilung durch die Fitfunktion $f(x, \vec{\lambda})$ ⁶ abgeschätzt wird.

Das heißt für jedes Bin gilt eine Wahrscheinlichkeit von:

$$P_f(N_i) = \frac{e^{-f(x_i, \vec{\lambda})} f(x_i, \vec{\lambda})^{N_i}}{N_i!}$$

und die (Log-)Likelihood-Funktion setzt sich aus den Poisson-Wahrscheinlichkeiten der einzelnen Bins zusammen

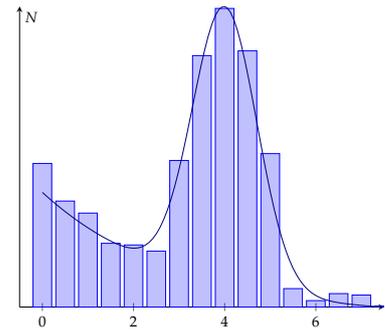
$$L(x_1, \dots, x_n, \vec{\lambda}) = \prod_{i=1}^n \frac{e^{-f(x_i, \vec{\lambda})} f(x_i, \vec{\lambda})^{N_i}}{N_i!},$$

$$\mathcal{L}(x_1, \dots, x_n, \vec{\lambda}) = \sum_{i=1}^n \left(-f(x_i, \vec{\lambda}) + N_i \ln(f(x_i, \vec{\lambda})) - \ln(N_i!) \right).$$

Der letzte Summand hängt nicht von der Messung ab, die Log-Likelihood-Funktion für Histogramme lautet dann:

$$\mathcal{L}(x_1, \dots, x_n, \vec{\lambda}) = \sum_{i=1}^n \left(-f(x_i, \vec{\lambda}) + N_i \ln(f(x_i, \vec{\lambda})) \right).$$

Um die Schätzer für die Parameter $\vec{\lambda}$ zu finden, wird \mathcal{L} ebenso maximiert.



⁶ genauer:

Der Mittelwert der Fitfunktion über der Bin-Breite $\langle f(x, \vec{\lambda}) \rangle_{x_i \pm \Delta x_i / 2}$

Fehlerabschätzung bei Maximum Likelihood

Für eine Fehlerabschätzung lässt sich bei gefundenem Schätzer $\hat{\lambda}$ die Log-Likelihood-Funktion um ihr Maximum entwickeln:

$$\mathcal{L}(x_1, \dots, x_n, \vec{\lambda}) \approx \mathcal{L}(x_1, \dots, x_n, \hat{\lambda}) + \underbrace{0}_{\text{1. Ord.}} + \frac{1}{2} (\lambda_j - \hat{\lambda}_j) (\lambda_k - \hat{\lambda}_k) \underbrace{\frac{\delta^2 \mathcal{L}}{\delta \lambda_j \delta \lambda_k} \Big|_{\lambda=\hat{\lambda}}}_{-\mathbf{B}_{jk}} + \dots$$

Es ergibt sich um das Maximum der Log-Likelihood-Funktion

$$\mathcal{L}(x_1, \dots, x_n, \vec{\lambda}) \approx \mathcal{L}_{\max} - \frac{1}{2} (\vec{\lambda} - \hat{\lambda})^\top \mathbf{B} (\vec{\lambda} - \hat{\lambda})$$

und daher für die Likelihood-Funktion:

$$L(x_1, \dots, x_n, \vec{\lambda}) \approx L_{\max} e^{\frac{1}{2} (\vec{\lambda} - \hat{\lambda})^\top \mathbf{C}^{-1} (\vec{\lambda} - \hat{\lambda})}.$$

Auch hier kann die Kovarianzmatrix \mathbf{C} wie in Kapitel 3.3 mit dem Inversen der Matrix \mathbf{B} identifiziert werden. Für unkorrelierte Parameter ergibt sich die Standardabweichung aus den Diagonalelementen

$$\sigma_{\lambda_j}^2 = C_{jj} = B_{jj}^{-1} = \left(-\frac{\delta^2 \mathcal{L}}{\delta \lambda_j \delta \lambda_j} \Big|_{\lambda=\hat{\lambda}} \right)^{-1}.$$