

Statistische Methoden in der Experimentalphysik

PHYSIKALISCHES INSTITUT

RUPRECHT-KARLS-UNIVERSITÄT
HEIDELBERG

MARKUS KÖHLI, KLAUS REYGERS
VERSION 1.03
22. MAI 2023



Inhaltsverzeichnis

1	Einführung	5
1.1	Vorwort	5
1.2	Danksagung	5
1.3	Literatur	6
2	Grundlagen	7
2.1	Wahrscheinlichkeitsdichte	7
2.2	Erwartungswert	8
2.3	Zentrale Momente und Varianz	8
2.4	Funktionen von zwei Zufallsvariablen	10
2.5	Weitere Parameter: Median und Quantile	11
2.6	Beschreibung diskreter Daten	12
3	Verteilungsfunktionen	13
3.1	Binomialverteilung	13
3.2	Poisson-Verteilung	13
3.3	Gauß-Verteilung	14
3.3.1	Zentraler Grenzwertsatz	14
3.3.2	Gauß-Verteilung in zwei Parametern	15
3.4	Gleichverteilung	15
3.5	Breit-Wigner-Verteilung	16
3.6	Faltung von Verteilungen	17
4	Fehlerbehandlung	19
4.1	Unsicherheit des Mittelwertes	19
4.2	Fehlerfortpflanzung	20
4.3	Kovarianz und Korrelation	20
4.4	Kovarianzmatrix	21
4.5	Kovarianzmatrix und systematische Unsicherheiten	22
4.6	Schätzer	22
4.7	χ^2 -Verteilung	24
4.8	Methode der kleinsten Quadrate	26
4.9	Konfidenzintervalle	28
4.10	Maximum Likelihood Methode	29
4.11	Bayessche Parameterschätzung	32
5	Durchführung	35

1 Einführung

1.1 Vorwort

‘Zu jedem Messwert gehört eine Messunsicherheit’ - eine elementare Regel, die spätestens nach dem ersten Semester verinnerlicht sein dürfte. Diese Prämisse erscheint intuitiv und fundiert.

In der Praxis aber fällt dieses Prinzip oft zuerst den unterschiedlichsten Begründungen zum Opfer, welche nicht selten auf unvollständiger Kenntnis des Experimentes beruhen, meist jedoch mit einem 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 z.B. die des Polywassers², der Polymerstruktur des Wassers. Es wäre verfehlt, historische Negativbeispiele einer eher qualitativen Arbeitsweise zuzuschreiben, 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 erfolgen³ - ganz im Sinne des Experimentators.

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

1.2 Danksagung

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

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:

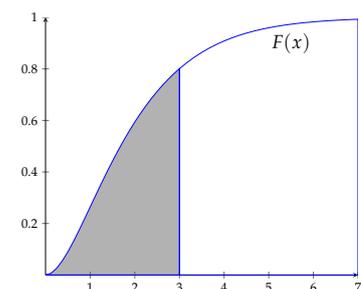
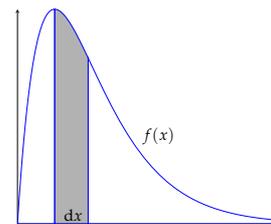
$$P(x' \leq b) = \int_{-\infty}^b f(x) dx = F(b),$$

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** $F(x)$ der Wahrscheinlichkeitsdichtefunktion $f(x)$:

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



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

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 $\sqrt{\text{Var}(x)}$ wird als **Standardabweichung** $\sigma(x)$ bezeichnet. Mittelwert und Standardabweichung sind allgemein die wichtigsten Größen zur statistischen Beschreibung einer Messreihe. Die Messunsicherheit in einem Experiment, der Messfehler, wird meist 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

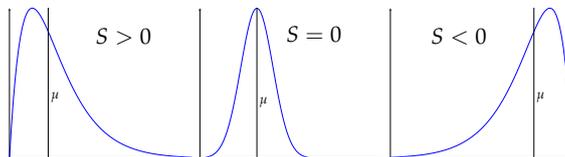
Die Momente α_l mit $l > 2$ werden als höhere Momente bezeichnet. Höhere Potenzen gewichten mehr und mehr die Randwerte der Verteilung.

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

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, inwiefern eine Verteilung schmäler ($K < 0$), 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**:

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

Kovarianz

Die Kovarianz ist ein Maß für die gemeinsame Variabilität von zwei Zufallsvariablen. Wenn die größeren Werte der einen Variablen hauptsächlich mit den größeren Werten der anderen Variablen korrespondieren und das Gleiche für die kleineren Werte gilt (d. h. die Variablen zeigen tendenziell ein ähnliches Verhalten), ist die Kovarianz positiv. Im umgekehrten Fall, wenn größere Werte der einen Variablen hauptsächlich kleineren Werten der anderen entsprechen, ist die Kovarianz negativ. Die Stärke der Korrelation wird mit dem dimensionslosen **Korrelationskoeffizienten** ρ quantifiziert, der auf

Korrelationskoeffizienten

die relativen Breiten $\sigma_{x,y}$ normiert ist:

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

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 $\rho = -1$ entgegengesetzt.

Die Variablen x und y werden als **unabhängig** voneinander bezeichnet, wenn sich die gemeinsame Verteilung $f(x, y)$ schreiben lässt als

$$f(x, y) = f_x(x) \cdot f_y(y).$$

Der Korrelationskoeffizient für unabhängige Variablen ist $\rho = 0$, d.h. unabhängige Variablen sind unkorreliert. Der Umkehrschluss gilt jedoch nicht, d.h., ein Korrelationskoeffizient $\rho = 0$ bedeutet nicht, dass x und y unabhängig sind.

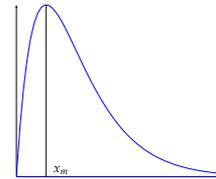
unabhängige Variablen

Unabhängige Variablen sind unkorreliert. Der Umkehrschluss gilt nicht.

2.5 Weitere Parameter: Median und Quantile

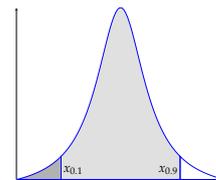
- Bei dem **wahrscheinlichsten Wert** x_m , dem **Modalwert** oder **Modus**, liegt das Maximum einer Verteilung. Es gilt:

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



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

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

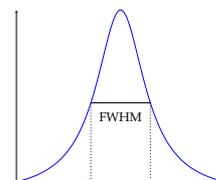


- Verallgemeinert gibt das **Quantil** den Wert x_q für beliebige Bruchteile $q \leq 1$ an:

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

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

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



2.6 Beschreibung diskreter Daten

Die einem Experiment (Stichprobe) entnommenen Daten liegen in Form eines Datensatzes x_1, \dots, x_N vor. Die zugrundeliegende Wahrscheinlichkeitsdichte $f(x)$ ist nicht immer 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 (=erwartungstreue) 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 erwartungstreue Schätzfunktion der Varianz. Es kann gezeigt werden, dass ein erwartungstreuer 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:

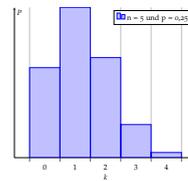
$$P(k, p, n) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

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

Varianz und Standardabweichung der Binomialverteilung

3.2 Poisson-Verteilung

Im Grenzfall einer unendlichen Zahl n von Experimenten und verschwindender Wahrscheinlichkeit p geht für den Fall, dass das Produkt $n \cdot p = \lambda$ endlich bleibt, die Binomialverteilung in die **Poisson-Verteilung**¹ über:

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

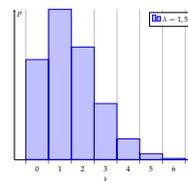
Ein klassisches Anwendungsbeispiel für die Poisson-Verteilung ist die Anzahl radioaktiver Zerfälle in einem gegebenen Zeitraum. Für die Poisson-Verteilung $P(k, \lambda)$ gilt: $P(k, \lambda)$ heißt Wahrscheinlichkeitsdichte der Poisson-Verteilung. Es gilt:

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

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

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

¹ nach Siméon Denis POISSON, Frankreich, Mathematiker und Physiker



Normierung

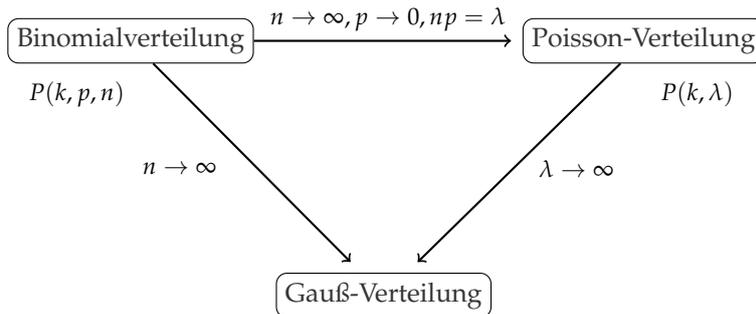
Erwartungswert

Varianz

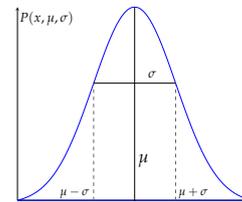
3.3 Gauß-Verteilung

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 basiert sowohl die Fehlerbeschreibung als 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 Nationen, Mathematiker, Astronom, Geodät und Physiker



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

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

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

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

Gauß-Verteilung

Ist x eine normalverteilte Messgröße, die um den wahren Wert μ streut, so ist Wahrscheinlichkeit, x in einem Intervall $\pm n\sigma$ um μ zu finden:

$[\mu - \sigma, \mu + \sigma]$	$[\mu - 2\sigma, \mu + 2\sigma]$	$[\mu - 3\sigma, \mu + 3\sigma]$
68,3 %	95,4 %	99,7 %

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 Grenzfall $n \rightarrow \infty$ 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:

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

Die Konturen konstanter Fehlerwahrscheinlichkeit sind Ellipsen. Die durch $[-\sigma_x, +\sigma_x]$ and $[-\sigma_y, +\sigma_y]$ eingeschlossene Ellipse ist gegeben durch

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

In Vektordarstellung ausgedrückt wird die Ellipsengleichung 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

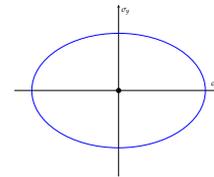
$$\vec{x}^T \mathbf{B} \vec{x} = 1.$$

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:

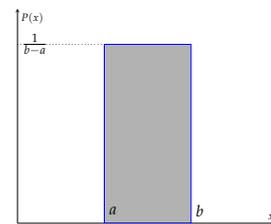
$$P(x) = \begin{cases} c = \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Sie besitzt die folgenden Eigenschaften:

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

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

Die Standardabweichung einer Gleichverteilung insbesondere beträgt $\sigma = \frac{b-a}{\sqrt{12}}$.



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 Spezialfall 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 und sind damit undefiniert, da die Funktion nicht schnell genug abfällt. Daher wird für die Breit-Wigner-Verteilung die Breite durch das Full-Width-Half-Maximum (FWHM) angegeben:

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

⁵ nach Gregory BREIT, Ukraine, Physiker

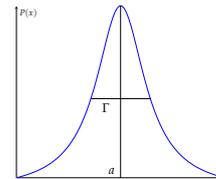
⁶ nach Eugene WIGNER, Königreich Ungarn, Physiker

⁷ Hendrik Antoon LORENTZ, Niederlande, Mathematiker und Physiker

⁸ nach Augustin-Louis CAUCHY, Frankreich, Mathematiker

Erwartungswert

Varianz



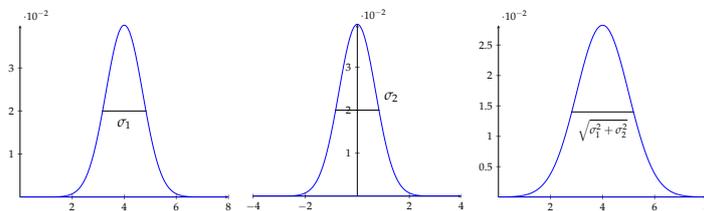
3.6 Faltung von Verteilungen

Eine **Faltung** oder auch **Konvolution**⁹ beschreibt die Auswirkung des Auflösungsvermögens einer Apparatur auf eine Observable. Seien die Wahrscheinlichkeitsdichten der Observablen $f(x)$ und die der Messunsicherheit $g(y)$ und ergebe sich der Messwert $z = x + y$, dann beschreibt das Faltungsintegral die Wahrscheinlichkeitsdichte $h(z)$

$$h(z) = \int_{-\infty}^{\infty} f(t)g(z-t) dt = \int_{-\infty}^{\infty} f(z-t)g(t) dt.$$

Ein wichtiger Fall ist die Faltung zweier Gauß-Verteilungen $N(x; \mu_1, \sigma_1)$ und $N(y; \mu_2, \sigma_2)$. Dies ergibt wieder eine Gauß-Verteilungen $N(z; \mu, \sigma)$ mit

$$\mu = \mu_1 + \mu_2, \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}.$$



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

⁹ von lat. *convolvere* = zusammenrollen

4 Fehlerbehandlung

Vorbemerkung: Für Messwerte, wie sie im statistischen Sinne aufgefasst werden, wird häufig angenommen, dass deren Streuung um den wahren Wert durch eine Gauß-Verteilung beschrieben wird. Daher wird die Messunsicherheit s mit der Standardabweichung σ einer Gauß-Verteilung identifiziert.

4.1 Unsicherheit 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

Die Unsicherheit des arithmetischen Mittelwertes ergibt sich nach

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

Haben die Messwerte unterschiedlich Unsicherheiten 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

Die Unsicherheit 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

Die Messgrößen x und y seien unabhängig und normalverteilt um die wahren Werte μ_x und μ_y mit den Standardabweichungen σ_x bzw. σ_y . Wir interessieren uns nun für die Streuung einer abhängigen Größe $z = f(x, y)$. Mit der **Gauß'schen Fehlerfortpflanzung** schätzt man die Varianz σ_z^2 durch eine Taylorentwicklung erster Ordnung ab¹:

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

In dieser Näherung gilt für den Erwartungswert $E[z] = E[f(x, y)]$

$$E[f(x, y)] \approx f(\mu_x, \mu_y)$$

Die Varianz lässt sich schreiben als

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

Die Kovarianz $\text{cov}(x, y) \equiv E[(x - \mu_x)(y - \mu_y)]$ verschwindet für unabhängige x und y und wir erhalten

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

Somit lässt sich die Unsicherheit von $z = f(x, y)$ schreiben als

$$\sigma_z = \sqrt{\left(\left. \frac{\delta f}{\delta x} \right|_{\mu_x, \mu_y} \right)^2 \sigma_x^2 + \left(\left. \frac{\delta f}{\delta y} \right|_{\mu_x, \mu_y} \right)^2 \sigma_y^2}.$$

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

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

Die Kovarianz ergibt sich dabei entweder

¹ Unter den Annahmen, dass:

- Terme höherer Ordnung vernachlässigbar sind
- Ableitungen bei μ_x, μ_y durch Ableitungen bei den Messwerten x, y genähert werden können

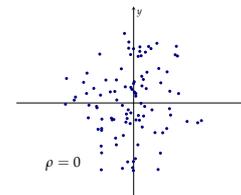
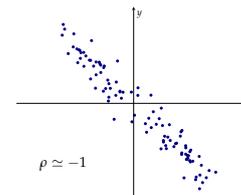
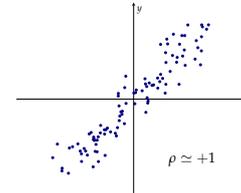
- aus dem experimentellen Aufbau. Eine Arrangement aus zwei Detektoren etwa, bei dem ein Ereignis entweder im einen oder zwingend im anderen Detektor erfasst wird, bedeutet eine Korrelation von $\rho = -1$ auf die Zählraten.
- aus einem bekannten funktionalen Zusammenhang. Dies ist insbesondere dann 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 von den Größen x_1, \dots, x_n abhängt, 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 die Unsicherheit von 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 von m Funktionen $\vec{y} = (f_1, \dots, f_m)$ der Variablen x_1, \dots, x_n ergibt sich für jedes Paar f_k, f_l eine (Ko-)Varianz, sodass die Unsicherheit $\sigma^2(f(x, y))$ 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:

$$\mathbf{E} = \mathbf{G} \cdot \mathbf{C} \cdot \mathbf{G}^T.$$

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

4.5 Kovarianzmatrix und systematische Unsicherheiten

Besitzen zwei Messgrößen x_1, x_2 statistische Unsicherheiten $\sigma_{x_1}, \sigma_{x_2}$ sowie eine gemeinsame systematische Unsicherheit s

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

so sind beide Größen korreliert:

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

Die entsprechende Kovarianzmatrix ist

$$\mathbf{C} = \begin{pmatrix} \sigma_{x_1}^2 + s^2 & s^2 \\ s^2 & \sigma_{x_2}^2 + s^2 \end{pmatrix}.$$

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. Aus der Stichprobe sind der mit einer Unsicherheit behaftete Schätzwert $\vec{\hat{\lambda}}$ für $\vec{\lambda}$ und dessen Varianz $\sigma(\vec{\hat{\lambda}})$ zu bestimmen. Die Anforderungen an einen guten Schätzer $S(x_1, \dots, x_n)$ sind, dass er folgende Eigenschaften besitzt:

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

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

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

- **konsistent**: Im Limes einer unendlichen Anzahl von Messungen soll der Schätzer gegen den wahren Wert konvergieren

$$\lim_{n \rightarrow \infty} \hat{\lambda} = \lambda.$$

- **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.

Schätzer für den Mittelwert einer Gauß-Verteilung

Für eine Gauß-Verteilung ist der Mittelwert $\bar{x} = 1/n \sum x_i$ ein erwartungstreuer 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.

Schätzer für die Varianz einer 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.

4.7 χ^2 -Verteilung

Die χ^2 -Verteilung lässt sich wie folgt definieren: Folgen die Zufallsvariablen x_1, \dots, x_n einer Gaußverteilung mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$, so folgt die Summe z der n Quadrate

$$z = \chi^2 = x_1^2 + \dots + x_n^2$$

einer χ^2 -Verteilung mit n Freiheitsgraden. Die χ^2 -Verteilung ist gegeben durch

$$f(z; n) = \frac{z^{n/2-1} e^{-z/2}}{2^{n/2} \Gamma(\frac{n}{2})} \quad (z \geq 0)$$

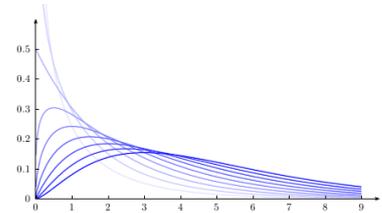
wobei $\Gamma(x)$ die Gamma-Funktion ist. Mittelwert und Varianz sind

$$E(z) = n, \quad \text{Var}(z) = 2n.$$

Die χ^2 -Verteilung spielt eine wichtige Rolle als Maß für die Güte der Beschreibung von Messdaten durch ein Modell. Hat man einen Datensatz aus n normalverteilten Messwerten y_i an den Orten x_i mit Standardabweichungen σ_i und eine Modellvorhersage $f(x; \vec{\lambda})$ mit vorher festgelegten Parametern, so folgt für wiederholte Messungen des Datensatzes die Größe

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - f(x_i; \vec{\lambda}))^2}{\sigma_i^2}$$

einer χ^2 -Verteilung mit n Freiheitsgraden, wenn das Modell die Erwartungswerte $\langle y_i \rangle$ korrekt beschreibt. Sind die m Parameter durch eine Anpassung an die Daten bestimmt worden, so folgt das so definierte χ^2 einer χ^2 -Verteilung mit $n - m$ Freiheitsgraden.



Anzahl der Freiheitsgrade n_F bei Anpassung eines Modells an Daten:
Anzahl der Messwerte n minus
Anzahl der Parameter m : $n_F = n - m$

Anwendung: Hypothesentest

Unter der Annahme, dass das Modell korrekt ist und rein statistische Unsicherheiten vorliegen, sollte der χ^2 -Wert des Fits bei wiederholten Messungen des Datensatzes einer χ^2 -Verteilung folgen. Die Wahrscheinlichkeit P , für eine gegebene Anzahl n an Freiheitsgraden einen schlechteren (also größeren) Wert als das beobachtete χ^2 zu erhalten, wird p -Wert genannt. Er ergibt sich aus der Verteilungsfunktion:

$$p\text{-Wert} = P(\chi_{n_F}^2) = 1 - F(\chi_{n_F}^2).$$

Der p -Wert ist also die Wahrscheinlichkeit, einen größeren als den beobachteten χ^2 -Wert zu erhalten, unter der Annahme, dass das verwendete Modell korrekt ist. Ist der p -Wert klein, kann die Hypothese, dass das verwendete Modell die Daten beschreibt, verworfen werden. Ein willkürliches, aber häufig verwendetes Kriterium, ist $p\text{-Wert} < 0.05$. Man erwartet, dass jeder Freiheitsgrad das χ^2 um 1 erhöht:

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

Der Wert χ^2/doF^2 , auch χ_{red}^2 , reduziertes Chi² genannt, 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:

² dof = degrees of freedom

- das verwendete Modell ist falsch oder unzureichend,
- die Gauß-Statistik für die Verteilung der Unsicherheiten ist eine falsche Annahme,
- es gibt unberücksichtigte systematische Unsicherheiten,
- die angenommene Standardabweichung σ ist zu groß (χ^2 wird zu klein) oder zu klein (χ^2 wird zu groß).

Allerdings besitzt auch der Wert χ^2/doF eine Unsicherheit, welche mit $1/\sqrt{n}$ skaliert. Daher ist $\chi_{\text{red}}^2 = 1.5$ bei $n = 5$ hinreichend akzeptabel, bei $n = 100$ nicht.

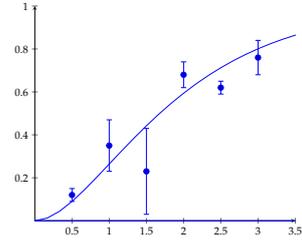
4.8 Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate soll einen unverzerrten und konsistenten Schätzer liefern. Für n gegebene Datenpunkte $x_i, y_i \pm \sigma_i$ wird die Funktion $f(x; \vec{\lambda})$ bestimmt, indem χ^2 **minimiert** wird. Dies liefert den besten Schätzwert für die Parameter $\vec{\lambda}$:

$$\arg \min_{\vec{\lambda}} (\chi^2) = \arg \min_{\vec{\lambda}} \left(\sum_{i=1}^n \frac{(y_i - f(x_i; \vec{\lambda}))^2}{\sigma_i^2} \right).$$

Für m Parameter $\lambda_1, \dots, \lambda_m$ ergeben sich m Gleichungen:

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



Beispiel: **Ausgleichsgerade**

$$f(x; \vec{\lambda}) = f(x; a_1, a_0) = a_1 x + a_0$$

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - a_1 x_i - a_0)^2}{\sigma_i^2}.$$

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

$$\frac{d\chi^2}{da_0} = \frac{1}{\sigma^2} (-2) \sum (y_i - a_1 x_i - a_0) = 0,$$

$$\frac{d\chi^2}{da_1} = \frac{x_i}{\sigma^2} (-2) \sum (y_i - a_1 x_i - a_0) = 0,$$

als Schätzer

$$\hat{a}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{cov}(x, y)}{\sigma^2(x)}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$$

mit den Unsicherheiten

$$\sigma^2(\hat{a}_1) = \sum \left(\frac{\delta \hat{a}_1}{\delta x_i} \right)^2 \sigma_i^2 \rightarrow \sigma^2(\hat{a}_1) = \frac{\sigma^2}{n(\overline{x^2} - \bar{x}^2)},$$

$$\sigma^2(\hat{a}_0) = \frac{\sigma^2 \bar{x}^2}{n(\overline{x^2} - \bar{x}^2)},$$

$$\text{cov}(\hat{a}_1, \hat{a}_0) = -\frac{\sigma^2 \bar{x}}{n(\overline{x^2} - \bar{x}^2)}.$$

Wichtig: Die Kovarianz von \hat{a}_1 und \hat{a}_0 hängt vom Erwartungswert \bar{x} ab. Für eine Gerade der form

$$f(x; a_1, a_0) = a_1 (x - \bar{x}) + a_0$$

ergeben sich die unkorrelierten Lösungen

$$\hat{a}_1 = \frac{\overline{xy}}{\overline{x^2}}, \quad \sigma^2(\hat{a}_1) = \frac{\sigma^2}{n\overline{x^2}},$$

$$\hat{a}_0 = \bar{y}, \quad \sigma^2(\hat{a}_0) = \frac{\sigma^2}{n}.$$

**Verallgemeinerung:
beliebige Funktionen und beliebige Fehlermatrizen**

Gegeben seien Messwerte y_i an den Stellen x_i

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

welche nicht unabhängig sind, also die Kovarianzmatrix \mathbf{C} nicht verschwindende Nebendiagonalelemente $\text{cov}(y_i, y_j)$ besitzt. Die Messwerte y_i sollen durch eine Funktion $f(x; \vec{\lambda})$ beschrieben werden, die an den Stellen x_i die Werte $\mu_i = f(x_i; \vec{\lambda})$ annimmt. Die χ^2 -Funktion lautet dann:

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

mit $\vec{\mu} = (\mu_1, \dots, \mu_n)$. 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 .

Im Allgemeinen müssen numerische Minimierungsverfahren zur Bestimmung des χ^2 -Minimum verwendet werden. Für Funktionen, die linear in den Parametern sind, lässt sich jedoch eine Lösung in geschlossener Form angeben. Eine Funktion dieser Art lässt sich schreiben als

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

dann lässt sich abkürzend schreiben

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

sodass die χ^2 -Funktion geschrieben werden kann als

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

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

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

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

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

Somit erhält man

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

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

$$\vec{\mu} = \begin{pmatrix} f(x_1; \vec{\lambda}) \\ \vdots \\ f(x_n; \vec{\lambda}) \end{pmatrix}$$

Kovarianzmatrix \mathbf{C} :

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

Parameter

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

Beispiel einer Funktion, die linear in den Parametern ist:

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

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

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

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

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 frequentistische Ansatz³:

Dieser Ansatz basiert auf der frequentistischen Definition der Wahrscheinlichkeit. Es wird vorab der Ereignisraum Ω festgelegt. Die Wahrscheinlichkeit eines Ereignisses E wird als Limes der relativen Häufigkeit n_E/n_G festgelegt, wobei n_E die beobachtete Anzahl des Ereignisses E und n_G die Gesamtzahl der Beobachtungen ist:

$$P(E) = \lim_{n_G \rightarrow \infty} \frac{n_E}{n_G}.$$

Im frequentistische Ansatz ist $\hat{\lambda} \pm \sigma(\hat{\lambda})$ eine Aussage über das konstruierte Intervall, die so zu verstehen ist, dass bei wiederholter Durchführung der Messung im Mittel ein Anteil P der konstruierten Intervalle den wahren Wert enthält. In anderen Worten, die Aussage $\lambda \in [\hat{\lambda} - \sigma, \hat{\lambda} + \sigma]$ ist mit einer Wahrscheinlichkeit von 68% richtig. Das Konzept einer Wahrscheinlichkeit für den Wert eines Parameters existiert im frequentistischen Ansatz nicht.

Der Bayessche Ansatz⁴:

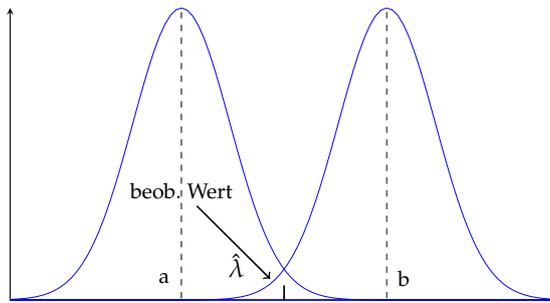
Dieser Ansatz basiert auf dem Bayesschen Wahrscheinlichkeitsbegriff, nach dem $P(H)$ den Grad der Überzeugung eines Subjekts in eine Hypothese H angibt. Im Falle eines zu bestimmenden Parameters λ wird der Grad der Überzeugung für einen bestimmten Wert von λ durch eine Wahrscheinlichkeitsdichte $P(\lambda)$ angegeben. Um ein Ergebnis interpretieren zu können, muss dieses immer unter seiner vorab feststehenden Wahrscheinlichkeit, der *A-priori*-Wahrscheinlichkeitsdichte, betrachtet werden. Diese wird durch den Experimentator festgesetzt und unterliegt damit seiner subjektiven Einschätzung. Eine Messung eines Parameters λ zusammen mit der *A-priori*-Verteilung führt zu einer neuen Wahrscheinlichkeitsverteilung $P'(\lambda)$. Die scheinbare Willkür bei der Wahl der *A-priori*-Verteilung wird häufig als Kritikpunkt am Bayesschen Ansatz angeführt.

³ von engl. 'Häufigkeit'

⁴ nach Thomas BAYES, England, Mathematiker und Pfarrer

Konfidenzintervalle:

Im frequentistischen Ansatz verwendet man zur Bestimmung von Vertrauensintervallen die Wahrscheinlichkeitsdichte der Messwerte bei festem Parameter. Es werden bei dieser Methode die Grenzen bestimmt, in welchen der Messwert mit der vorgegebener Wahrscheinlichkeit liegt.



Das Konfidenzintervall $[a, b]$ wird so festgelegt, dass

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

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

für die vorgegebenen Ausschlusswahrscheinlichkeiten α, β gilt. Das Intervall $[a, b]$ beinhaltet den wahren Wert mit einer Wahrscheinlichkeit von $P = 1 - \alpha - \beta$. Die Wahrscheinlichkeit P wird Konfidenzniveau genannt.

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

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

Likelihood-Funktion

Bei der Likelihood-Funktion betrachtet man die gemessenen Werte als fest und die Parameter als Variablen. Das Maximum-Likelihood-Prinzip besagt nun, dass die beste Schätzung $\vec{\lambda}_{ML}$ für die Parameter $\vec{\lambda}$ diejenige ist, die die Likelihood-Funktion maximiert:

$$\vec{\lambda}_{ML} = \arg \max_{\vec{\lambda}} L(x, \vec{\lambda})$$

Numerisch ist es häufig vorteilhaft, den Logarithmus der Likelihood-Funktion, die **Log-Likelihood-Funktion** zu betrachten:

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

Der beste Schätzer $\vec{\lambda}$ ergibt sich als Lösung der Gleichungen

$$\frac{\delta}{\delta \lambda_j} L(x_i, \lambda_j) = 0, \quad j = 1, \dots, m$$

wobei m die Anzahl der Parameter ist.

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 $\hat{\tau}$:

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

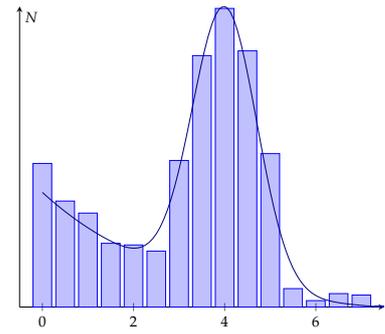
Der Maximum-Likelihood-Schätzer für die Lebensdauer beim radioaktiven Zerfall ist gerade das arithmetische Mittel der gemessenen Zeiten.

Maximum Likelihood für Histogramme

Für geringe Stichprobengrößen, wie sie etwa in Histogrammen mit **wenigen Einträgen pro Bin** vorkommen, kann man nicht von einer Normalverteilung für die Anzahl der Einträge pro Bin ausgehen. Es muss entweder eine Binomialverteilung oder eine Poisson-Verteilung angenommen werden, siehe Kapitel 3. Die χ^2 -Methode kann somit nicht angewendet werden und die Maximum-Likelihood-Methode ist dann das Mittel der Wahl.

Die Wahrscheinlichkeit, dass ein Ereignis in Bin i fällt ist

$$p_i(\vec{\lambda}) = \int_{x_i - \Delta x_i / 2}^{x_i + \Delta x_i / 2} f(x; \vec{\lambda}) dx \approx f(x_i) \Delta x_i$$



Es ist häufig eine gute Wahl, Poisson-Statistik anzunehmen. Die Gesamtzahl der Ereignisse fluktuiert damit entsprechend der Poisson-Verteilung $P(N_{\text{tot}}; \bar{N}_{\text{tot}})$ um den Mittelwert \bar{N}_{tot} , der eine Vorhersage des Modells oder ein freier Parameter sein kann. Der vom Modell vorhergesagte Erwartungswert für die Anzahl der Ereignisse in Bin i ist $\mu_i(\vec{\lambda}) = \bar{N}_{\text{tot}} p_i$. Die Likelihood-Funktion lässt sich dann schreiben als

$$L(N_1, \dots, N_n; \vec{\lambda}) \equiv L(\vec{\lambda}) = \prod_{i=1}^n P(N_i; \mu_i(\vec{\lambda})) = \prod_{i=1}^n \frac{\mu_i(\vec{\lambda})^{N_i}}{N_i!} e^{-\mu_i(\vec{\lambda})}$$

wobei N_i die beobachtete Anzahl von Ereignissen in Bin i ist. Geht man nun zur Log-Likelihood-Funktion über und lässt Terme, die nicht von den Parameters abhängen weg, ergibt sich die Funktion

$$\tilde{\mathcal{L}}(\vec{\lambda}) = \sum_{i=1}^n N_i \ln \mu_i(\vec{\lambda}) - \mu_i(\vec{\lambda}) = -\bar{N}_{\text{tot}} + \sum_{i=1}^n N_i \ln \mu_i.$$

Die Parameter, die die Funktion maximieren, stellen die Maximum-Likelihood-Schätzer dar.

Varianz des Maximum-Likelihood-Schätzers

Um die Varianz des Maximum-Likelihood-Schätzers $\hat{\lambda}$ zu bestimmen, kann man die Log-Likelihood-Funktion um ihr Maximum entwickeln:

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

Falls ein Maximum-Likelihood-Schätzer existiert, nähert sich $L(x_1, \dots, x_n, \vec{\lambda})$ im Grenzfall großen Anzahl n von Messungen einer Normalverteilung an. Entsprechend ist dann $\mathcal{L}(x_1, \dots, x_n, \vec{\lambda})$ eine quadratische Funktion in $\vec{\lambda}$. Dies wird *asymptotische Normalität* genannt. Es ergibt sich somit 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

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

Für den Spezialfall nur eines Parameters reduziert sich dies zu

$$\sigma_{\lambda}^2 = -\frac{1}{\frac{\delta^2 \mathcal{L}}{\delta^2 \lambda} \Big|_{\lambda=\hat{\lambda}}}.$$

Die Unsicherheit von λ kann alternativ über die Werte $\hat{\lambda} - \sigma_{\lambda}^-$ und $\hat{\lambda} + \sigma_{\lambda}^+$ abgeschätzt werden, für die die Log-Likelihood-Funktion den Wert $\ln L_{\max} - \frac{1}{2}$ annimmt:

$$\ln L(\hat{\lambda} \pm \sigma_{\lambda}) = \ln L_{\max} - \frac{1}{2}$$

Dies liefert i.A. asymmetrische Unsicherheiten. Im Limes eines großen Datensamples gilt jedoch $\sigma_{\lambda} = \sigma_{\lambda}^- = \sigma_{\lambda}^+$.

4.11 Bayessche Parameterschätzung

In der Bayesschen Parameterschätzung ist die gesamte Information über die zu schätzenden Parameter $\vec{\lambda} = (\lambda_1, \dots, \lambda_n)$ in der Posterior-Wahrscheinlichkeitsfunktion $P(\vec{\lambda}|x)$ enthalten. Unter Verwendung des Bayesschen Theorems lässt sich diese schreiben als

$$P(\vec{\lambda}|x) = \frac{L(x|\vec{\lambda})\pi(\vec{\lambda})}{\int L(x|\vec{\lambda})\pi(\vec{\lambda}) d\vec{\lambda}}$$

wobei wie schon vorher die Likelihood-Funktion durch $L(x|\vec{\lambda}) = \prod_{i=1}^n f(x_i, \vec{\lambda})$ gegeben ist. Die A-priori-Wahrscheinlichkeitsverteilung $\pi(\vec{\lambda})$ beschreibt das Wissen über die zu schätzenden Parameter vor der Messung. Die Formel für $P(\vec{\lambda}|x)$ gibt dann an, wie dieses Wissen auf Basis der Messwerte x_i aktualisiert wird. Der Nenner $\int L(x|\vec{\lambda})\pi(\vec{\lambda}) d\vec{\lambda}$ hängt nicht von $\vec{\lambda}$ und kann damit als reiner Normierungsfaktor betrachtet werden, der durch die Normierung der Posterior-Wahrscheinlichkeitsverteilung auf 1 bestimmt werden kann. Wird eine konstante A-priori-Verteilung gewählt, so entspricht der Parametervektor $\vec{\lambda}$, für den die Posterior-Verteilung ein Maximum hat, dem Maximum-Likelihood-Schätzer. Eine konstante Wahrscheinlichkeitsverteilung lässt sich nicht normieren und ist damit eine sog. uneigentliche A-priori-Verteilung.

Als Beispiel betrachten wir zwei Messungen eines Parameters λ . Bei gegebenem λ sei die Wahrscheinlichkeitsdichte dafür, einen bestimmten Wert x_1 zu messen, durch eine Normalverteilung $N(x_1; \lambda, \sigma_1)$ mit Mittelwert λ und Standardabweichung σ_1 gegeben. Betrachtet man diese Verteilung im Bayesschen Ansatz nun als Funktion von λ bei festem x_1 , so ergibt sich als A-priori-Verteilung

$$\pi(\lambda) = N(\lambda; x_1, \sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(\lambda-x_1)^2}{2\sigma_1^2}}.$$

Berücksichtigt man nun auch die zweite Messung x_2 von λ (mit Standardabweichung σ_2), so ergibt sich

$$\begin{aligned} P(\lambda|x_2) &\propto L(\lambda|x_2)\pi(\lambda) = N(\lambda; x_2, \sigma_2)\pi(\lambda) \\ &= N(\lambda; x_2, \sigma_2)N(\lambda; x_1, \sigma_1). \end{aligned}$$

Das Produkt zweier Normalverteilungen ist proportional zu einer Normalverteilung und es ergibt sich als Posterior-Verteilung

$$P(\lambda|x_2) = N(\lambda; \mu, \sigma)$$

mit

$$\mu = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \left(\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} \right) \quad \text{und} \quad \sigma = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}.$$

Der Bayessche Ansatz liefert in diesem Fall also die Formel für den gewichteten Mittelwert aus Abschnitt 4.1.

5 Durchführung

Die praktischen Aufgaben werden in Form von **Jupyter-Notebooks** gestellt. Diese Notebooks sollen mit Text und Code ergänzt werden. Die fertigen Notebook mit vollständiger Lösung werden dann an den/die Tutor/in geschickt. Als Programmiersprache soll Python mit entsprechenden Paketen (numpy, matplotlib, ...) verwendet werden.

Die zu bearbeitenden praktischen Aufgaben werden vom Tutor / von der Tutorin ausgewählt. Hier einige mögliche Aufgaben:

Fehlerfortpflanzung

1. In dieser Aufgabe soll Sympy verwendet werden, um analytische Ausdrücke für die Unsicherheit einer abhängigen Größe $y = f(x_1, \dots, x_n)$ zu bestimmen, wobei die Messwerte x_i korreliert sein können
(S01_error_prop_01.ipynb).

Methode der kleinsten Quadrate

1. Linearer χ^2 -Fit
Ist die Fitfunktion linear in den Fitparametern, lässt sich die Anpassung analytisch bestimmen. Dies soll in dieser Aufgabe gemacht werden
(S01_least_squares_01.ipynb)
2. Simultaner χ^2 -Fit
Hier wird ein Modell simultan an unterschiedliche Messdaten angepasst
(S01_least_squares_02.ipynb).

Maximum-Likelihood-Methode

1. Mittlere Lebensdauer beim exponentiellen Zerfall
Hier wird die Maximum-Likelihood-Methode am Beispiel des exponentiellen Zerfall genauer untersucht.
(S01_ml_01.ipynb).
2. Ungebinnter Maximum-Likelihood-Fit
Eine einfacher Maximum-Likelihood-fit mit einem Parameter
(S01_ml_02.ipynb).

Bayessche Parameterschätzung

1. Bestimmung der Posterior-Verteilung für den Parameter p einer Binomial-Verteilung
(S01_bayes_01.ipynb).

Für diesen Versuch sind Grundkenntnisse in Python notwendig. Der [AP Python Einführungskurs](#) bietet einen guten Einstieg. Zur numerischen Minimierung bei χ^2 - und Maximum-Likelihood-Fits bietet sich das [iminuit-Paket](#) an. Die Jupyter-Notebooks können auf dem eigenen Rechner oder auf dem [KIP-Jupyter-Server](#) bearbeitet werden.