Statistik

Stephan I. Böttcher

Institut für Experimentelle und Angewandte Physik
Christian Albrechts Universität zu Kiel

Sommersemester 2016

Physik ist am einfachsten, wenn sie linear ist. Wenn man eine Funktion genügend vergrößert, dann sieht sie immer aus wie eine Gerade. Wenn man die Funktion durch diese Gerade ersetzt, wird ein Problem vielleicht lösbar. Vorraussetzung ist, man findet die exakte Lösung für einen Punkt der Funktion im Zentrum der Dynamik, die man betrachtet.

Zur Linearisierung nimmt man z.B. die Taylorreihe oder die binomische Reihe.

Die binomische Formel ist

$$(a+b)^c = \sum_{k=0}^{\infty} \binom{c}{k} a^k b^{c-k}. \tag{1}$$

Der Binomialkoeffizient ist

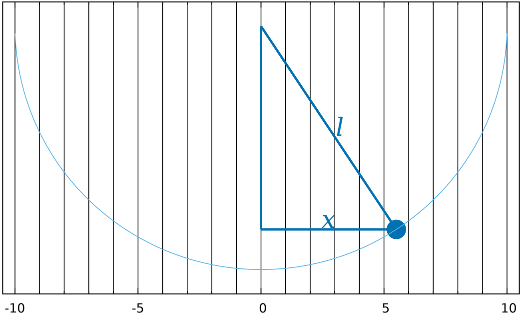
$$\binom{c}{k} = \prod_{l=1}^{k} \frac{c-l+1}{l} = \frac{c(c-1)\ldots(c-k+1)}{k!}. \tag{2}$$

Für natürliche $c \in \mathbb{N}_0$ bricht die binomische Reihe nach $k = c$ ab.

Wir klammern a^c aus, $x = b/a \ll 1$

$$(1+x)^c = \sum_{k=0}^{\infty} \binom{c}{k} x^k = 1 + cx + \mathcal{O}(x^2). \tag{3}$$

Beispiel: Potenzial eines Pendels



$$V(x) = mg(l - \sqrt{l^2 - x^2}) = mgl\left(1 - \sqrt{1 - \frac{x^2}{l^2}}\right) = mg\frac{x^2}{2l}. \quad (4)$$

Damit haben wir ein harmonisches Potenzial.

Näherung der Funktion $f(x)$ durch die Tangente am Punkt $x = x_0$

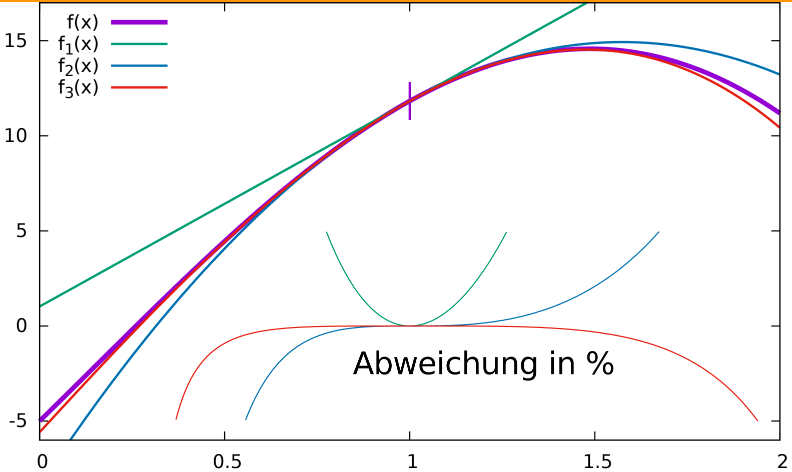
$$f(x) = f(x_0) + \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=x_0} (x - x_0). \tag{5}$$

Das sind die ersten beiden Glieder der Taylorreihe

$$f(x) = \sum_{n=0}^{\infty} \left.\frac{\mathrm{d}^n f}{\mathrm{d}x^n}\right|_{x=x_0} \frac{(x-x_0)^n}{n!}. \tag{6}$$

Beispiel:

$$e^x = 1 + x + \mathcal{O}(x^2), \qquad \ln(1+x) = x + \mathcal{O}(x^2). \tag{7}$$



$$f(x) = 20\sin(x) - x + 2x^2 - x^3 - 5 \tag{8}$$

Die Taylorreihe der Lösung der Differentialgleichung

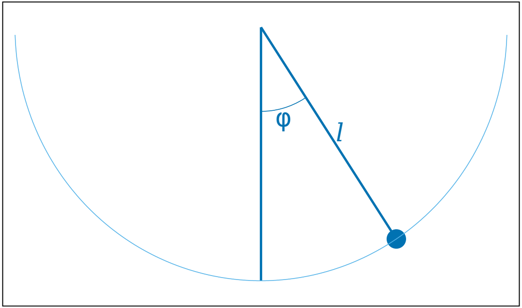
$$\frac{\mathrm{d}f}{\mathrm{d}x} = f, \qquad f(0) = 1 \tag{9}$$

entwickelt um den Punkt $x_0 = 0$ ist

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \tag{10}$$

da alle Ableitungen an der Stelle $x = 0$ den Wert 1 annehmen. Die Lösung ist natürlich die Exponentialfunftion, $f(x) = \exp(x)$.

Beispiel: Potenzial eines Pendels



$$V(\varphi) = mg(l - l\cos\varphi) = mgl(1-\cos\varphi) = mgl\frac{\varphi^2}{2}. \qquad (11)$$

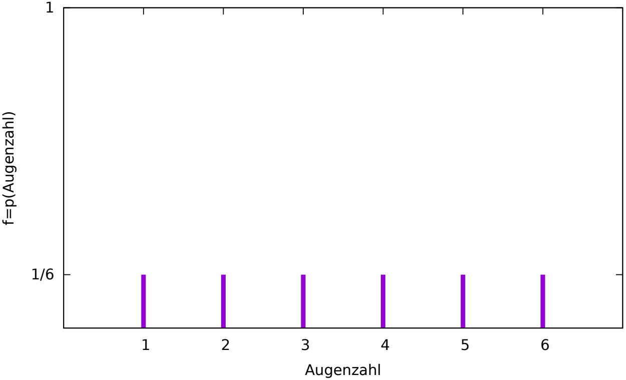
Die Taylorreihe für $\cos\varphi$ ist

$$\cos\varphi = 1 - \tfrac{1}{2}\varphi^2 + \tfrac{1}{24}\varphi^4 + \mathcal{O}(\varphi^6). \qquad (12)$$

Eine Wahrscheinlichkeitsverteilung $f(x)$ ist eine Funktion einer Zufallsvariablen x, die für jeden möglichen Wert von x angibt, wie wahrscheinlich es ist, daß dieser Wert eintritt.

Für diskrete Zufallsgrößen gibt die Wahrscheinlichkeitsverteilung für jeden möglichen Wert von x die Wahrscheinlichkeit dafür an, daß dieser Wert eintritt.

Beispiel: Würfel



Für kontinuierliche Zufallsgrößen ist $f(x)$ die Wahrscheinlichkeitsdichte

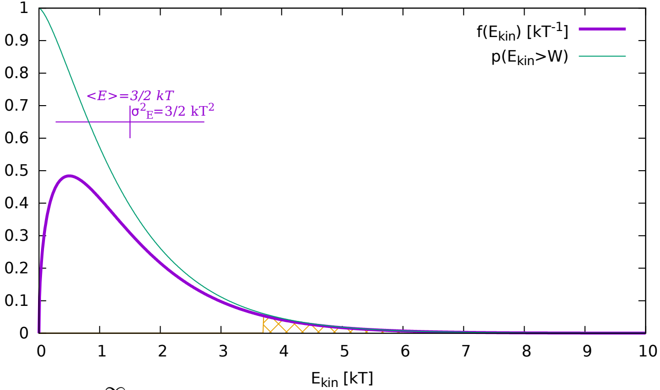
$$f(x) = \frac{\mathrm{d}p(x)}{\mathrm{d}x}. \tag{13}$$

$p(x)$ ist die Wahrscheinlichkeit, daß der Wert kleiner oder gleich x eintritt. Die Wahrscheinlichkeit, daß x in einen gegebenen Bereich (x_1, x_2) fällt ist

$$p(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x. \tag{14}$$

Für mehrdimensionale $\vec{x}$ gelten entsprechende Verallgemeinerungen.

Beispiel: Maxwell-Boltzmann Verteilung



$$p(E > W) = \int_W^\infty \frac{2}{\sqrt{\pi}}\sqrt{\frac{E}{T}}e^{-\frac{E}{T}}\ \frac{\mathrm{d}E}{T} = 1 - \mathrm{erf}(\sqrt{\frac{W}{T}}) + \frac{2}{\sqrt{\pi}}\sqrt{\frac{W}{T}}e^{-\frac{W}{T}} \tag{15}$$

Zur Transformation der Maxwell-Boltzmann Verteilung

$$\mathrm{d}p = \sqrt{\frac{2}{\pi}}\frac{mv^2}{2kT}e^{-\frac{mv^2}{2kT}}\frac{\mathrm{d}v}{\sqrt{\frac{2kT}{m}}} \tag{16}$$

in die Verteilung der kinetischen Energie, wird die Substitution

$$E = \tfrac{1}{2}mv^2, \qquad \mathrm{d}E = mv\,\mathrm{d}v = \sqrt{2Em}\,\mathrm{d}v \tag{17}$$

durchgeführt

$$\mathrm{d}p = \frac{2}{\sqrt{\pi}}\sqrt{\frac{E}{kT}}e^{-\frac{E}{kT}}\frac{\mathrm{d}E}{kT}. \tag{18}$$

Unter den Momenten der Ordnung n einer Verteilung $f(x)$ versteht man die Integrale

$$M_n = \int x^n f(x) \, \mathrm{d}x \tag{19}$$

über den gesammten Wertebereich von x. Für diskrete Zufallsvariablen wird das Integral durch entsprechende Summen ersetzt.

Der Begriff hat seinen Ursprung in der Betrachtung von Masseverteilungen $\varrho(\vec{x})$.

Das nullte Moment der Masseverteilung ist die Gesammtmasse

$$m = M_0 = \int \varrho(\vec{x})\,\mathrm{d}^3V. \tag{20}$$

Der Schwerpunkt ist das normierte erste Moment der Masseverteilung

$$\vec{S} = \frac{M_1}{M_0} = \frac{1}{m}\int \vec{x}\varrho(\vec{x})\,\mathrm{d}^3V. \tag{21}$$

Das Trägkeitsmoment ist das zweite Moment der Masseverteilung um die Drehachse

$$M = M_2 = \int (x^2 + y^2)\varrho(\vec{x})\,\mathrm{d}^3V. \tag{22}$$

Das nullte Moment einer Wahrscheinlichkeitsverteilung ist eins:

$$1 = M_0 = \int_{-\infty}^{\infty} f(x)\,dx. \tag{23}$$

Das erste Moment ist der Erwartungswert $\langle x \rangle$

$$\langle x \rangle = M_1 = \int_{-\infty}^{\infty} x\,f(x)\,dx. \tag{24}$$

Das zweite zentrale Moment ist die Varianz σ_x^2

$$\sigma_x^2 = M_2 - M_1^2 = \int_{-\infty}^{\infty} (x - \langle x \rangle)^2\,f(x)\,dx. \tag{25}$$

Der Erwartungswert eines Wurfs mit dem Würfel ist

$$\langle n\rangle = \tfrac{1}{6}\sum_{i=1}^{6} i = \tfrac{7}{2} = 3.5, \tag{26}$$

und die Varianz

$$\sigma_n^2 = \tfrac{1}{6}\sum_{i=1}^{6}\left(i-\tfrac{7}{2}\right)^2 = \tfrac{35}{12} = 2.91\bar{6} = 1.71^2 \tag{27}$$

Die Momente der Maxwell-Boltzmann Verteilung (hier wieder als Funktion der kinetischen Energie, nicht der Geschwindigkeit, mit $k = 1$) sind

$$\langle E \rangle = \int_0^\infty \frac{2}{\sqrt{\pi}} \sqrt{\frac{E}{T}} e^{-\frac{E}{T}} \, \frac{\mathrm{d}E}{T} \, E = \frac{3}{2} T \tag{28}$$

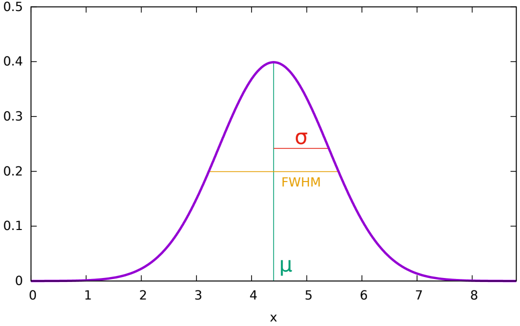
und

$$\sigma_E^2 = \int_0^\infty \frac{2}{\sqrt{\pi}} \sqrt{\frac{E}{T}} e^{-\frac{E}{T}} \, \frac{\mathrm{d}E}{T} \, (E - \langle E \rangle)^2 = \frac{3}{2} T^2 \tag{29}$$

Die wichtigste Wahrscheinlichkeitsverteilung ist die Normalverteilung, auch Gaussverteilung.

Mit Erwartungswert μ und Varianz σ^2 ist die Dichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{30}$$



Eine normalverteilte Größe mit Erwartungswert $\mu = 0$ und Varianz $\sigma^2 = 1$ ist standardnormalverteilt.

Das Integral einer Wahrscheinlichkeitsdichte heißt Verteilungsfunktion

$$F(x) = \int_{-\infty}^{x} f(\xi)\mathrm{d}\xi. \tag{31}$$

Die Wurzel der Varianz heißt Standardabweichung σ.

Intergrale der Gaussverteilung können nicht geschlossen gelöst werden. Deshalb hat man die Lösung einfach als Funktion definiert. Die Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung ist

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \tag{32}$$

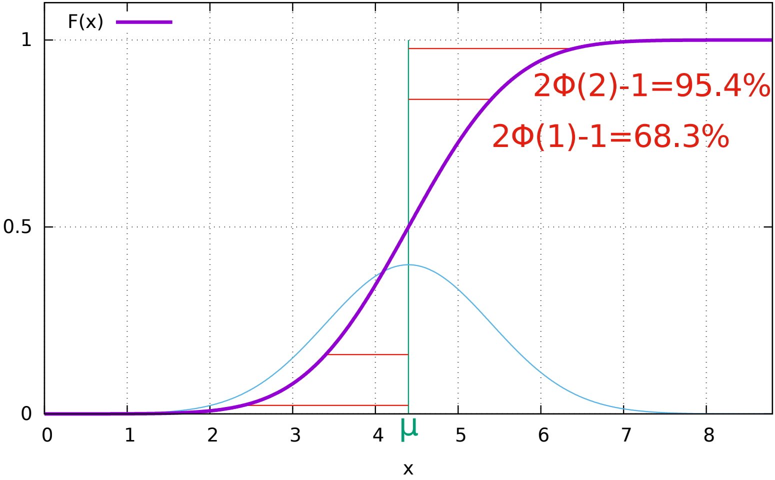
mit der Fehlerfunktion $\operatorname{erf}(x)$

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int\limits_{0}^{x} e^{-\tau^2}\,\mathrm{d}\tau. \tag{33}$$

Die Verteilungsfunktion der Normalverteilung mit Erwartungswert μ und Varianz σ^2 ist

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{34}$$

Gaußsche Normalverteilung



Die Summe von normalverteilten Zufallsgrößen ist auch Normalverteilt. Sei $x = a + b$ mit Verteilungen

$$f_a(a) = \frac{1}{\sqrt{2\pi\sigma_a^2}} e^{-\frac{(a-\mu_a)^2}{2\sigma_a^2}}, \qquad f_b(b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} e^{-\frac{(b-\mu_b)^2}{2\sigma_b^2}} \tag{35}$$

Die Verteilung von x ist

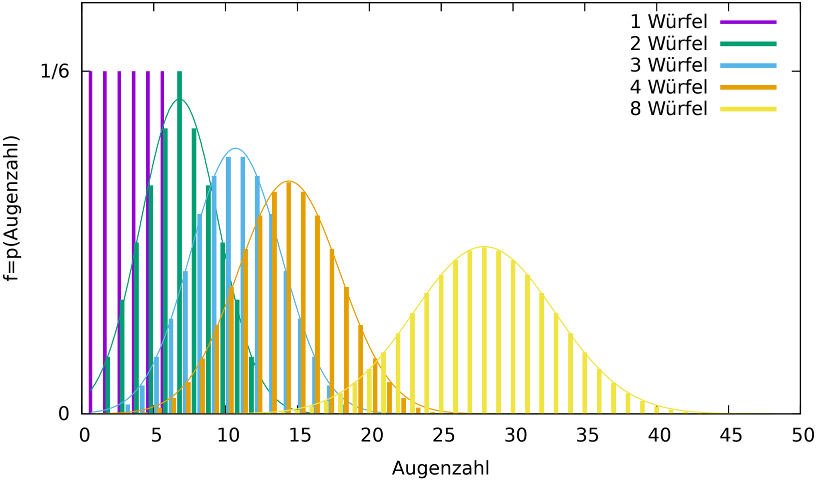
$$f(x) = \int_{-\infty}^{\infty} f_a(a) f_b(x-a)\,\mathrm{d}a = \frac{1}{\sqrt{2\pi(\sigma_a^2+\sigma_b^2)}} e^{-\frac{(x-\mu_a-\mu_b)^2}{2(\sigma_a^2+\sigma_b^2)}} \tag{36}$$

x ist normalverteilt mit Erwartungswert $\mu_x = \mu_a + \mu_b$ und Varianz $\sigma_x^2 = \sigma_a^2 + \sigma_b^2$.

Der zentrale Grenzwertsatz (central limit theorem) besagt etwa:

> *Wenn man die Summe von genügend vielen unabhängigen weitgehend beliebig verteilten Zufallszahlen bildet, dann erhält man eine annähernd normalverteilte Zufallszahl.*

Zentraler Grenzwertsatz



Wenn man unabhängige Zufallsereignisse zählt, dann erhält man eine Poisson verteilte Zufallszahl. Die Wahrscheinlichkeit bei der nächsten Zählung n Ereignisse zu zählen ist

$$P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda}. \tag{37}$$

Der Erwartungswert der Poissonverteilung ist λ

$$\sum_{n=0}^{\infty} n P_\lambda(n) = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{(n-1)}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda. \tag{38}$$

Die Varianz der Poissonverteilung ist auch $\sigma^2 = \lambda$

$$\sum_{n=0}^{\infty}(n-\lambda)^2 P_\lambda(n) = e^{-\lambda}\sum_{n=0}^{\infty}(n^2 - 2n\lambda + \lambda^2)\frac{\lambda^n}{n!} = \lambda \tag{39}$$

mit

$$\sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda}\sum_{n=0}^{\infty}(n+1)\frac{\lambda^{n+1}}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\left(\lambda^2 \frac{\lambda^{n-1}}{(n-1)!} + \lambda\frac{\lambda^n}{n!}\right) = \lambda^2+\lambda \tag{40}$$

Die Standardabweichung einer Zählung ist immer die Wurzel der Anzahl der gezählten Ereignisse $\sigma = \sqrt{\lambda}$. Das gilt z.B. für die Fehlerbalken an den *bins* eines Histogramms.

Die Summe von poissonverteilten Zufallsgrößen ist poissonverteilt:

$$\sum_{k=0}^{n} P_{\lambda_1}(k) P_{\lambda_2}(n-k) = \sum_{k=0}^{n} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2} \tag{41}$$

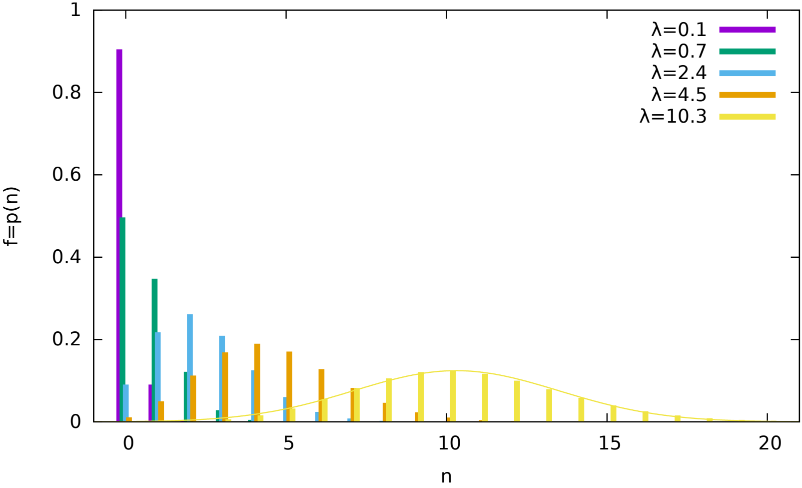
$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \tag{42}$$

$$= \frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)} \tag{43}$$

Wenn Zählungen addiert werden ist das equivalent einer längeren Zählung.

Nach dem zentralen Grenzwertsatz sind Poissonverteilungen für große λ annähernd normal verteilt.

Possionverteilung



Wir nehmen eine Folge von n normalverteilten Zufallszahlen x_i, mit Erwartungswerten μ_i und Varianzen σ_i^2. Daraus bilden wir die Summe

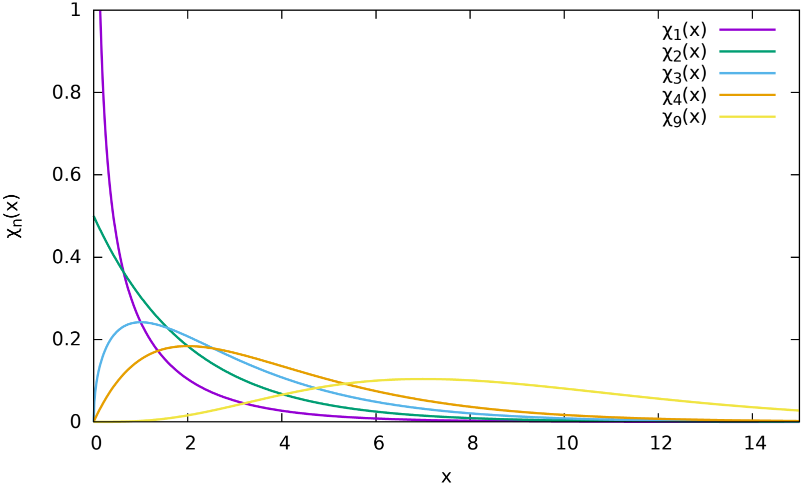
$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}. \tag{44}$$

Die Wahrscheinlichkeitsverteilung dieser Summe heißt χ^2-Verteilung mit n Freiheitsgraden.

$$\chi_n^2(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}, \tag{45}$$

mit Erwartungswert $\langle x \rangle = n$ und Varianz $\sigma_x^2 = 2n$.

χ^2-Verteilung



Das Resultat der Messung einer physikalischen Größe ist eine Zufallszahl. Ziel der Messung ist, die Parameter oder die Momente der Wahrscheinlichkeitsverteilung dieser Zufallszahlen zu bestimmen.

Meist wird eine Schätzung des Erwartungswertes und eine Schätzung der Standardabweichung (Quadratwurzel der Varianz) als Messgröße und Fehlerangabe angegeben.

Der zentrale Grenzwertsatz sagt uns, da eine Messung von vielen kleinen Fehlerquellen beeinflusst wird, daß die Messgrößen oft als normalverteilt angenommen werden können.

Beliebige Funktionen $M(a, b, \dots)$ von unabhängigen normalverteilten Zufallszahlen ergeben im Allgemeinen keine normalverteilten Werte. Wenn die Fehler nicht zu groß sind, dann können wir die Funktion in den Fehlern linearisieren, indem wir die normalverteilten Funktionsargumente $a, b, \dots$ durch die Abweichung vom Erwartungswert ersetzen

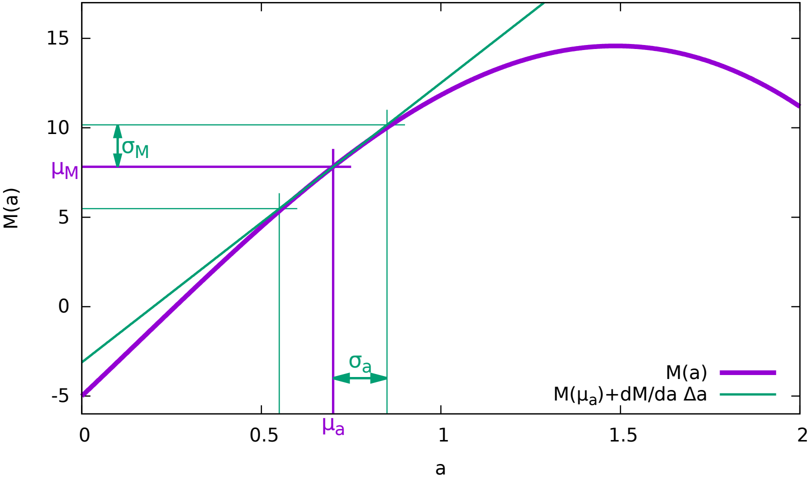
$$a = \mu_a + \Delta a, \quad b = \mu_b + \Delta b, \quad \dots \tag{46}$$

Die Zufallszahlen $\Delta a, \Delta b, \dots$ haben die gleiche Varianz wie die ursprünglichen Messgrössen und Erwartungswert 0.

$$M(a, b, \dots) = M(\mu_a, \mu_b, \dots) + \frac{\partial M}{\partial a}\Delta a + \frac{\partial M}{\partial b}\Delta b + \dots \tag{47}$$

Der erste Term ist der Erwartungswert von M, und die Varianz ist

$$\sigma_M^2 = \left(\frac{\partial M}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial M}{\partial b}\right)^2 \sigma_b^2 + \dots \tag{48}$$


15
10
5
0
-5
M(a)
σM
μM
σa
μa
0
0.5
1
1.5
2
a
M(a)
M(μa)+dM/da Δa

Die Bildung eines Mittelwertes erlaubt eine einfache Schätzung des Erwartungswertes und der Varianz einer Messmethode. Durch Wiederholung der Messung werden n Messwerte $a_i, i = 1 \dots n$ gewonnen, die der gleichen Wahrscheinlichkeitsverteilung unterliegen. Wir nehmen natürlich an, daß die Messwerte normalverteilt sind. Der Mittelwert

$$\bar{a} = \frac{\sum a_i}{n} \tag{49}$$

ist ein Schätzwert des Erwartungswertes dieser Verteilung.

Die Varianz der Verteilung der Messwerte könnte mit

$$\sigma_a^2 = \frac{\sum(a_i - \mu_a)^2}{n} \tag{50}$$

abgeschätzt werden, wenn wir den Erwartungswert μ_a kennen würden. Statt dessen müssen wird den Mittelwert $\bar{a}$ nehmen. Dabei geht uns ein Freiheitsgrad im Zähler des Bruchs verloren

$$\sigma_a^2 = \frac{\sum(a_i - \bar{a})^2}{n-1}. \tag{51}$$

Der Mittelwert $\bar{a}$ ist eine Summe normalverteilter Größen, und damit auch normalverteilt, mit Erwartungswert μ_a und Varianz

$$\sigma_{\bar{a}}^2 = \sum_{i=1}^{n} \frac{\sigma_a^2}{n^2} = \frac{\sigma_a^2}{n} \tag{52}$$

Mit dem Schätzwert für die Varianz σ_a^2

$$\sigma_{\bar{a}}^2 = \frac{\sum (a - \bar{a})^2}{n(n-1)}. \tag{53}$$

Das Messergebnis wird angegeben als

$$a = \bar{a} \pm \sqrt{\sigma_{\bar{a}}^2}. \tag{54}$$

Mit den Momenten A_k der Verteilung der Meßwerte a_i

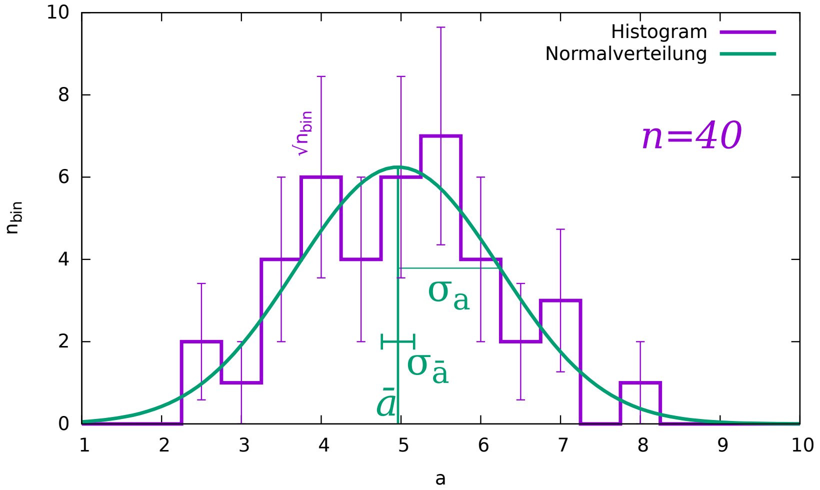
$$A_k = \sum_{i=1}^{n} a_i^k \tag{55}$$

ist der Mittelwert

$$\bar{a} = \frac{A_1}{A_0} \tag{56}$$

und die Varianz

$$\sigma_a^2 = \frac{\sum(a_i - \bar{a})^2}{n-1} = \frac{A_2 - 2A_1\frac{A_1}{A_0} + A_0\frac{A_1^2}{A_0^2}}{n-1} = \frac{A_2 - \frac{A_1^2}{A_0}}{n-1} = \frac{\sum a_i^2 - \frac{(\sum a_i)^2}{n}}{n-1}. \tag{57}$$


Histogram
Normalverteilung
n=40
√n_bin
n_bin
a
σ_a
σ_ā
ā
0
2
4
6
8
10
1
2
3
4
5
6
7
8
9
10

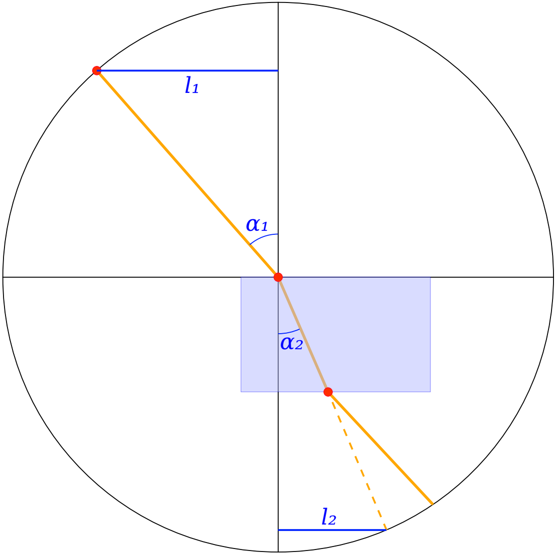
Oft wird eine Größe mit verschiedene Methoden messen, oder eine Methode mit verschiedenen Parametern anwenden. Dann werden die einzelnen Werte mit verschiedenen Gewichten g_i im Mittelwert berücksichtigt.

$$\bar{a} = \frac{\sum_{i=1}^{n} g_i a_i}{\sum_{i=1}^{n} g_i}. \tag{58}$$

Mit $g_i = 1$ ergibt sich der gewöhnliche Mittelwert. Wenn die Fehler σ_i der Meßwerte a_i bekannt sind, setzt man

$$g_i = \frac{1}{\sigma_i^2}. \tag{59}$$

Sonst wählt man die Gewichte entsprechend der Genauigkeit, die sich aus den Parametern der Messung ergeben.



Brechungsindex

$$n = \frac{\sin\alpha_1}{\sin\alpha_2} = \frac{l_1}{l_2}$$

Der Brechungsindex ist

$$n = \frac{\sin \alpha_1}{\sin \alpha_2} = \frac{l_1}{l_2} \tag{60}$$

Wenn die Längen l_j mit der gleichen Genauigkeit σ_l gemessen werden, ist die Varianz

$$\sigma_n = \left(\left(\frac{\partial n}{\partial l_1} \right)^2 + \left(\frac{\partial n}{\partial l_2} \right)^2 \right) \sigma_l^2 = \frac{(1+n)^2}{l_2^2} \sigma_l^2 = \frac{n^2(1+n)^2}{l_1^2} \sigma_l^2 . \tag{61}$$

Als Gewichte zur Mittelung der Messungen n_i des Brechungsindex bieten sich an

$$g_i = l_{1(i)}^2, \quad \text{oder} \quad g_i = l_{2(i)}^2 . \tag{62}$$

Wenn die Gewichte umgekehr proportional zu den Varianzen der Messwerte gewählt werden

$$g_i = \frac{c}{\sigma_i^2}, \tag{63}$$

dann ist

$$\sum_{i=1}^{n} g_i(a_i - \bar{a})^2 = c \sum_{i=1}^{n} \frac{(a_i - \bar{a})^2}{\sigma_i^2} = c\chi^2. \tag{64}$$

χ^2 ist $\chi^2_{(n-1)}$-verteilt, mit Erwartungswert $n-1$. Der Gewichtsfaktor c kann also mit

$$c = \frac{\sum g_i(a_i - \bar{a})^2}{n-1} \tag{65}$$

geschätzt werden.

Die Varianz des gewichteten Mittelwertes ist

$$\sigma_{\bar{a}}^2 = \sum_{i=1}^{n} \left(\frac{\partial \bar{a}}{\partial a_i}\right)^2 \sigma_i^2 = \sum_{i=1}^{n} \left(\frac{g_i}{\sum_j g_j}\right)^2 \frac{c}{g_i} = \frac{\sum_i g_i (a_i - \bar{a})^2}{(n-1) \sum_i g_i}. \qquad (66)$$

Für Messwerte a_i mit bekannten Fehlern σ_i ist

$$\sigma_{\bar{a}}^2 = \frac{\sum \frac{(a_i - \bar{a})^2}{\sigma_i^2}}{(n-1) \sum \frac{1}{\sigma_i^2}} = \frac{\chi^2}{(n-1) \sum \frac{1}{\sigma_i^2}} = \frac{1}{\sum \frac{1}{\sigma_i^2}}. \qquad (67)$$

Mit den Momenten

$$A_k = \sum_{i=1}^{n} g_i a_i^k \tag{68}$$

ergeben sich die gleichen Formeln wie für den gewöhnlichen Mittelwert

$$\bar{a} = \frac{A_1}{A_0} \tag{69}$$

und die Varianz

$$\sigma_{\bar{a}}^2 = \frac{\frac{A_2}{A_0} - \left(\frac{A_1}{A_0}\right)^2}{n-1}. \tag{70}$$

Nur gilt hier $n \neq A_0$.

Fehlerquellen, die man durch wiederholen der Messung und Mittelwertbildung reduzieren kann sind statistische Fehler. Solche Fehlerquellen sind Rauschen, Zählstatistik und andere nicht-reproduzierbare Einflüsse, wie z.B. Ableseblickwinkel.

Systematische Fehler können durch Mittelwertbildung nicht reduziert werden. Diese resultieren aus unvollständiger Kenntnis der Messanordnung. Wenn die Skala des verwendeten Zollstocks ungenau ist, dann nützt es nichts, die Messung mit dem selben Zollstock zu wiederholen.

Mann kann die Messung mit verscheidenen Zollstöcken verschiedener Hersteller wiederholen. Dann wird aus dem systematischen Fehler ein statistischer. Zumindest wenn weitere Korrelationen der Fehler ausgeschlossen werden können.

Den relative Fehler erhält man, wenn man den absoluten Fehler durch den Messwert teilt. Fehlerfortpflanzung ist besonders einfach mit relativen Fehlern, wenn die Funktion $M(a, b, \dots)$ proportional zu Potenzen der eingesetzten Messwerte ist

$$M(a, b) = k a^n b^m. \tag{71}$$

Der absoluter Fehler ist

$$\sigma_M = \sqrt{(nka^{n-1}b^m)^2 \sigma_a^2 + (mka^n b^{m-1})^2 \sigma_b^2}. \tag{72}$$

Der relative Fehler ist

$$\frac{\sigma_M}{M} = \sqrt{n^2 \left(\frac{\sigma_a}{a}\right)^2 + m^2 \left(\frac{\sigma_b}{b}\right)^2}. \tag{73}$$

Wir haben eine Theorie oder ein Modell $R(\vec{m};\vec{x}) = 0$, mit Modellparametern $\vec{m}$, dem unsere Messwerte $\vec{x}_i, i = 1 \ldots n$ gehorchen sollten.

Zum Beispiel, die Wertepaare $\vec{x}_i = (x_i, y_i)$ sollten auf einer Geraden liegen, geben durch die Parameter $\vec{m} = (a, b)$:

$$R(a, b; x, y) = y - ax - b = 0. \tag{74}$$

Unser Ziel ist nun, die Parameter $\vec{m}$ zu bestimmen, die am besten zu den Messwerten $\vec{x}_i$ passen.

Und wir wollen prüfen, ob das Modell die Messdaten erklären kann.

Zu jedem Messwert $\vec{x}_i$ und gegebenen Satz Parameter $\vec{m}$ kann die Wahrscheinlichkeit $p(\vec{m}; \vec{x}_i)$ angegeben werden, daß der Wert $\vec{x}_i$ im Rahmen der Messgenauigkeit zum Modell passt. Nehmen wir an, die Modellfunktion $R(\vec{m}; \vec{x})$ sei so normiert, daß diese Wahrscheinlichkeit nur eine Funktion der Residuen R ist

$$p(\vec{m}; \vec{x}) = p(R(\vec{m}; \vec{x})). \tag{75}$$

Zum Beispiel, die y_i-Werte der Gerade seien normalverteilt mit der Varianz σ_i^2, dann sind die Funktionswerte von

$$R(a, b; x_i, y_i, \sigma_i) = \frac{y_i - a x_i - b}{\sigma_i} \tag{76}$$

standardnormalverteilt.

Wir suchen nun die Werte der Parameter $\vec{m}$, für die unsere Messwerte die größte Wahrscheinlichkeit ergeben

$$p = \prod_{i=1}^{n} p(\vec{m}; \vec{x}_i) \quad \text{maximal.} \tag{77}$$

Einfacher umzugehen ist damit, wenn wir den Logarithmus nehmen

$$\log p = \sum_{i=1}^{n} \log p(\vec{m}; \vec{x}_i) \quad \text{maximal.} \tag{78}$$

Das ist die Suche nach der *maximum log likelihood*.

Wenn die Residuen der Modelfunktion $R(\vec{m}; \vec{x})$ standardnormalverteilt sind, dann ist die *likelihood* eines Messpunktes $\vec{x}_i$

$$p(\vec{m}; \vec{x}_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{R(\vec{m};\vec{x}_i)^2}{2}}. \tag{79}$$

Die *log likelihood* ist dann

$$\log p = \sum_{i=1}^{n} \log p(\vec{m}; \vec{x}_i) = -\frac{1}{2} \sum_{i=1}^{n} R(\vec{m}; \vec{x}_i)^2 = -\frac{1}{2}\chi^2. \tag{80}$$

Die Summe

$$\chi^2 = \sum_{i=1}^{n} R(\vec{m}; \vec{x}_i)^2 \tag{81}$$

ist χ^2_{n-m}-verteilt. m ist die Anzahl der Parameter.

Es gilt, Parameterwerte $\vec{m}$ für das kleinste χ^2 zu finden. Das ist die *least squares* Methode.

Um das kleinste χ^2 zu finden, müssen wir das Gleichungssystem

$$\frac{\partial \chi^2}{\partial m_j} = 0, \qquad j = 1 \ldots m, \tag{82}$$

mit den Unbekannen $\vec{m} = (m_1, m_2, \ldots, m_m)$ lösen.

Wenn das Modell linear in den Parametern ist,

$$R = \sum_{i=1}^{n} \left(r_0(\vec{x}_i) + \vec{m} \cdot \vec{r}(\vec{x}_i) \right), \tag{83}$$

dann ist auch das Gleichungssystem linear.

$$\frac{1}{2} \frac{\partial \chi^2}{\partial m_j} = \sum_{i=1}^{n} r_0(\vec{x}_i) r_j(\vec{x}_i) + \sum_{k=1}^{m} m_k \sum_{i=1}^{n} r_j(\vec{x}_i) r_k(\vec{x}_i) = 0. \tag{84}$$

Die Summe χ^2 der Quadrate der Residuen ist χ^2_{n-m}-verteilt, wenn die Residuen R unabhängig und annähernd standardnormalverteilt sind, die Fehlerbalken an den Datenpunkten also bekannt und korrekt berücksichtigt sind.

Dabei ist n die Anzahl der Datenpunkte und m ist die Anzahl der Parameter, mir denen das Modell an die Datenpunkte angepasst wurde.

Anhand von χ^2 kann man abschätzen, ob das Model stimmt. Wenn der Wert $\chi^2 \gg n - m$, dann ist entweder das Modell falsch, oder die Fehlerbalken, oder es sind Korrellationen vorhanden.

Ein zu kleines χ^2 deutet u.U. auf eine Überschätzung der Fehler hin.

Das Modell für die Residuen einer Geraden ist

$$R(a,b;x_i,y_i,\sigma_i) = \tfrac{y_i - a x_i - b}{\sigma_i}, \tag{85}$$

$$\chi^2 = \sum_{i=1}^{n} \tfrac{(y_i - a x_i - b)^2}{\sigma_i^2} = {\textstyle\sum}_{y^2} + a^2 {\textstyle\sum}_{x^2} + b^2 {\textstyle\sum}_{1} + 2ab {\textstyle\sum}_{x} - 2a {\textstyle\sum}_{xy} - 2b {\textstyle\sum}_{y}, \tag{86}$$

mit ${\textstyle\sum}_{\xi} = \sum_{i=1}^{n} \frac{\xi_i}{\sigma_i^2}$.

$$\frac{1}{2}\frac{\partial\chi^2}{\partial a} = \quad a{\textstyle\sum}_{x^2} + b{\textstyle\sum}_{x} - {\textstyle\sum}_{xy} \quad = 0 \tag{87}$$

$$\frac{1}{2}\frac{\partial\chi^2}{\partial b} = \quad a{\textstyle\sum}_{x} + b{\textstyle\sum}_{1} - {\textstyle\sum}_{y} \quad = 0 \tag{88}$$

Steigung, Regressionskoeffizient$_x$

$$a = \frac{\sum xy \sum 1 - \sum x \sum y}{\sum x^2 \sum 1 - \left(\sum x\right)^2} = \frac{\sigma_{xy}}{\sigma_x^2}, \tag{89}$$

y-Achsenabschnitt

$$b = \frac{\sum x^2 \sum y - \sum x \sum xy}{\sum x^2 \sum 1 - \left(\sum x\right)^2}, \tag{90}$$

Korrelationskoeffizient

$$r = \frac{\sum xy \sum 1 - \sum x \sum y}{\sqrt{\left(\sum x^2 \sum 1 - \left(\sum x\right)^2\right)\left(\sum y^2 \sum 1 - \left(\sum y\right)^2\right)}} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}, \tag{91}$$

Die Varianzen der x- und y-Werte, sowie die Kovarianz sind

$$\sigma_x^2 = \frac{\sum_{i=1}^{n} g_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} g_i} = \frac{\sum x^2}{\sum 1} - \left(\frac{\sum x}{\sum 1}\right)^2, \tag{92}$$

$$\sigma_y^2 = \frac{\sum_{i=1}^{n} g_i (y_i - \bar{y})^2}{\sum_{i=1}^{n} g_i} = \frac{\sum y^2}{\sum 1} - \left(\frac{\sum y}{\sum 1}\right)^2, \tag{93}$$

$$\sigma_{xy} = \frac{\sum_{i=1}^{n} g_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} g_i} = \frac{\sum xy}{\sum 1} - \frac{\sum x \sum y}{(\sum 1)^2}. \tag{94}$$

Für die Fehlerrechnung der Parameter a und b nehmen wir an, die Verteilung der Residuen σ_i spiegelt sich in den Fehlern der y_i-Werte

$$\sigma_a^2 = \sum_{i=1}^{n} \left(\frac{\partial a}{\partial y_i}\right)^2 \sigma_i^2 = \frac{1}{\sigma_x^2 \sum 1} = \frac{\sigma^2}{\sigma_x^2} \tag{95}$$

und

$$\sigma_b^2 = \sum_{i=1}^{n} \left(\frac{\partial b}{\partial y_i}\right)^2 \sigma_i^2 = \frac{\sum x^2}{\sigma_x^2 (\sum 1)^2} = \frac{\bar{x^2}\sigma^2}{\sigma_x^2} = \mu_{x^2}\sigma_a^2 = (\sigma_x^2 + \mu_x^2)\sigma_a^2, \tag{96}$$

mit

$$\sigma^2 = \frac{1}{\sum\limits_{i=1}^{n} \frac{1}{\sigma_i^2}}. \tag{97}$$

Wenn wir die Fehler der y_i-Werte nicht kennen und annehmen, sie seien alle gleich verteilt, dann rechnen wir mit $g_i = 1$, also

$$\chi^2 = \sum_{i=1}^{n} (y_i - a x_i - b)^2 \tag{98}$$

mit dem Erwartungswert

$$\left\langle \chi^2 \right\rangle = \sigma_i^2 (n-2). \tag{99}$$

Damit können wir abschätzen

$$\sigma_a^2 = \frac{\chi^2}{n(n-2)\sigma_x^2}, \qquad \sigma_b^2 = \frac{\chi^2}{n(n-2)} \left(1 + \frac{\mu_x^2}{\sigma_x^2} \right). \tag{100}$$

Wenn die Residuen linear in den Parametern sind, dann kann man das Gleichungssystem ohne Anfangswerte und Iteration lösen. Die Gerade ist natürlich linear. Es sei denn, die Fehler hängen von den Parametern ab, z.B.

$$\sigma_i^2 = \sigma_{y_i}^2 + a\sigma_{x_i}^2, \tag{101}$$

und

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - a x_i - b)^2}{\sigma_{y_i}^2 + a\sigma_{x_i}^2}. \tag{102}$$

In diesem Fall kann man die Ableitung des Nenners vernachlässigen und direkt lösen, muss dann aber iterieren, um die Fehler korrekt zu skalieren.

Wenn die Residuen nicht linear sind in den Parametern $\vec{m}$, dann verwenden wir statt dessen die Taylorentwicklung in erster Ordnung

$$R(\vec{m};\vec{x}_i) = R(\vec{m}_0;\vec{x}_i) + \vec{\nabla}_{\vec{m}} R(\vec{m}_0;\vec{x}_i)\Delta\vec{m} + \mathcal{O}(\Delta\vec{m}^2). \tag{103}$$

Dazu brauchen wir gute Anfangswerte $\vec{m}_0$ für die Parameter und müssen die Lösung iterieren, bis wir mit dem Ergebnis zufrieden sind. Die Lösung

$$\vec{m} = \vec{m}_0 + \Delta\vec{m} \tag{104}$$

wird der Anfangswert für die nächste Iteration.

Das zu lösende Gleichungssystem ist

$$\mathbf{A}\Delta\vec{m} = \vec{b}, \tag{105}$$

mit der Matrix

$$\mathbf{A}_{jk} = \sum_{i=1}^{n} \frac{\partial R(\vec{m}_0;\vec{x}_i)}{\partial m_j} \frac{\partial R(\vec{m}_0;\vec{x}_i)}{\partial m_k}, \tag{106}$$

und

$$\vec{b}_k = -\sum_{i=1}^{n} \frac{\partial R(\vec{m}_0;\vec{x}_i)}{\partial m_k} R(\vec{m}_0;\vec{x}_i). \tag{107}$$

Dies ist die *Gauss-Newton* Methode.

Wenn die Anfangswerte nicht wirklich gut gewählt werden, dann kann es sein, daß die Iteration nicht konvergiert. Kenneth Levenberg hat vorgeschlagen die Iterationen zu dämpfen, mit einem Dämpfungsfaktor λ

$$(\mathbf{A} + \lambda \mathbf{I}) \Delta \vec{m} = \vec{b}. \tag{108}$$

Wenn das χ^2 schnell abnimmt, dann kann λ kleiner gemacht werden, wenn es nicht voran geht, wird λ vergrößert.

Donald Marquardt meint, es sei besser, die Einheitsmatrix $\mathbf{I}$ durch die Diagonalmatrix $\mathrm{diag}(\mathbf{A})$ zu ersetzen

$$(\mathbf{A} + \lambda \, \mathrm{diag}(\mathbf{A})) \, \Delta \vec{m} = \vec{b}. \tag{109}$$

Es gibt verschiedene Ansichten, wie der Dämpfungsfaktor λ gewählt werden sollte. Marquardt schlug vor mit einem λ_0 anzufangen und auszuprobieren, ob χ^2 damit kleiner wird. Sonst wird λ solange mit einem konstanten Faktor ν multipliziert bis es klappt, $\lambda = \nu^k \lambda_0$.

Dann wird jede Iteration mit λ/ν ausprobiert. Wenn das zu einem kleineren χ^2 führt wird λ durch λ/ν ersetzt und iteriert. Sonst wird wieder $\nu^k \lambda, k = 0 \ldots$ ausprobiert, bis sich ein Fortschritt einstellt.

Bei komplexen Spektrometern müssen wir oft inverse Probleme lösen, um aus einer Messung von Zählraten $\vec{c}$ auf den spektralen Fluss $\vec{f}$ zurückzurechnen

$$\vec{c} = \mathbf{A}\vec{f}, \tag{110}$$

wobei die Geometriefaktoren $\mathbf{A}$ meist mit Monte-Carlo Rechnungen gewonnen wurden, zum Beispiel mit GEANT4.

In dieser Rechnung in Vorwärtsrichtung wirkt $\mathbf{A}$ als Tiefpass. Die Inverse Matrix $\mathbf{A}^{-1}$ ist also ein Hochpass. Naïve lineare Algebra $[\vec{f} = \mathbf{A}^{-1}\vec{c}]$ führt hier nicht zum Ziel.

Die Zählraten c_i sind meist Poisson-verteilt. Die Elemente der Matrix $\mathbf{A}$ auch. Der Vektor $\vec{f}$ folgt einer *a priori* Wahrscheinlichkeitsverteilung, die unsere Erwartungen widerspiegelt, wie so ein Spektrum auszusehen hat.

Es steckt also jede Menge Statistik in dem Problem.