

Interner Bericht
DESY F14-81/01
Juli 1981

METHODEN DER DATENANALYSE IN DER HOCHENERGIEPHYSIK

von

Volker Blobel

Eigentum der Property of	DESY	Bibliothek library
Zugang Accession	14. JUL 1988	
Leihfrist Loan period:	7	Tage days

DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.

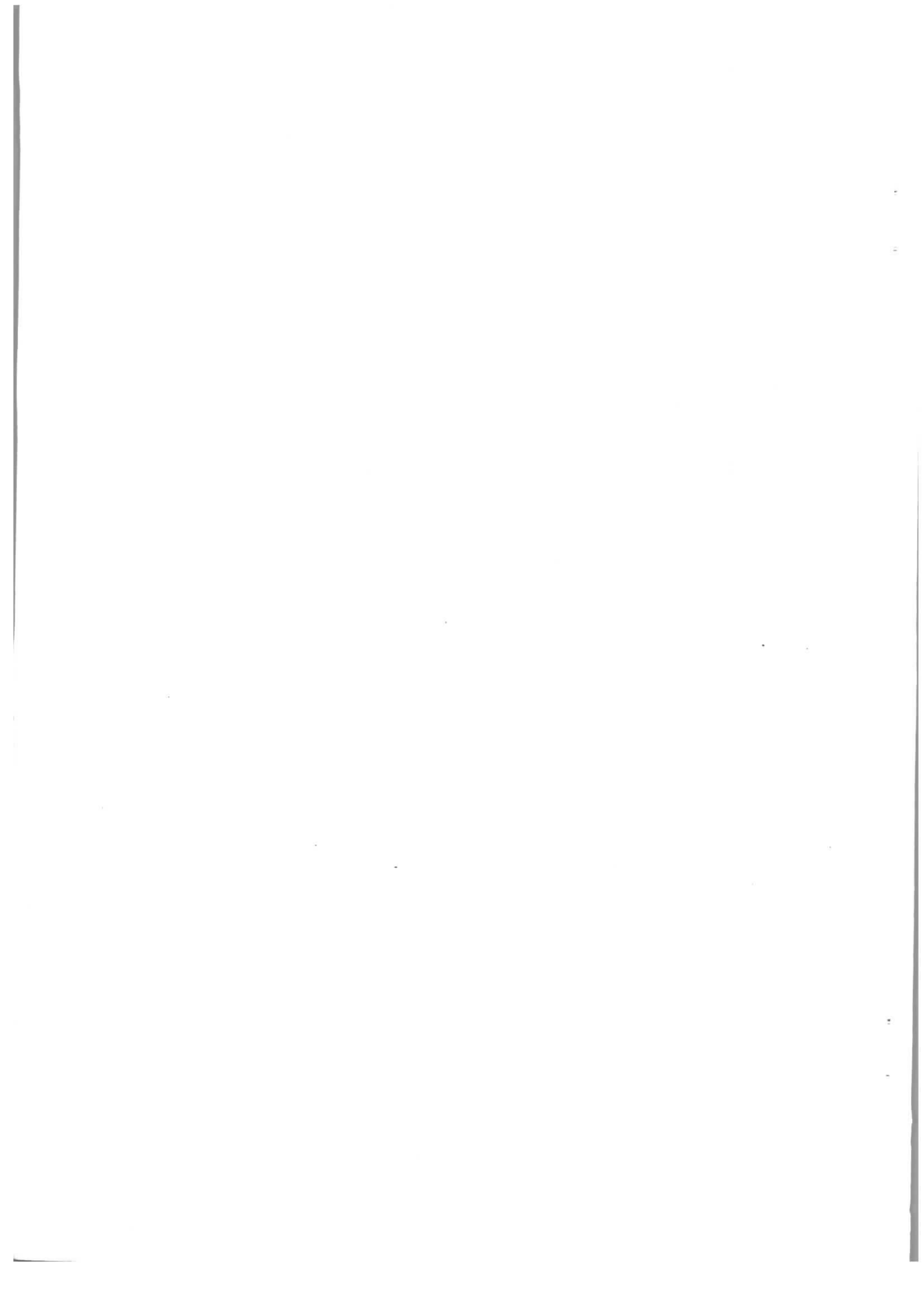
DESY reserves all rights for commercial use of information included in this report, especially in case of filing application for or grant of patents.

“Die Verantwortung für den Inhalt dieses
Internen Berichtes liegt ausschließlich beim Verfasser“

Methoden der Datenanalyse in der Hochenergiephysik

von Volker Blobel

Notizen zu einer Vorlesung an der Universität Hamburg WS 78/79



Inhaltsverzeichnis

	Seite
A. <u>Einführung in die Statistik</u>	1
1. Einleitung	1
2. Zufallsvariablen	7
2.1 Kontinuierliche und diskrete Zufallsvariable	7
2.2 Erwartungswerte	9
2.3 Mehrkomponentige Zufallsvariable	14
2.4 Transformationen	22
3. Spezielle Verteilungen	31
3.1 Die Normalverteilung	31
3.2 Binomial- und Poisson-Verteilung	33
3.3 χ^2 -Verteilung	39
4. Bestimmung von Parametern	41
4.1 Eigenschaften der Methoden zur Parameterbestimmung	41
4.1 Momentenmethode	46
4.3 Maximum-Likelihood Methode	49
4.4 Methode der kleinsten Quadrate	51
5. Konfidenzintervalle	66
5.1 Normalverteilte Größen	66
5.2 Der allgemeine Fall	71
5.3 Mehrdimensionale Zufallsvariablen	78
6. Testen von Hypothesen	82
6.1 Einführung	82
6.2 Parametertests	87
6.3 Anpassungstests	89
B. <u>Anwendung der Maximum-Likelihood Methode</u>	93
1. Problemstellung	93
2. Eigenschaften der Maximum-Likelihood Methode	101
3. Numerische Durchführung	108
4. Beispiel	120

	Seite
<u>C. Methode der kleinsten Quadrate</u>	123
1. Problemstellung der vermittelnden Ausgleichung	123
2. Lineare Probleme	126
3. Lösungseigenschaften und Folgerungen	130
4. Nichtlineare Probleme	133
5. Bedingte Ausgleichung	137
6. Direkte Ausgleichung mit nichtlinearen Bedingungsgleichungen	142
7. Programm Package CONLES	155
<u>D. Interpolation und Approximation</u>	159
1. Glätten von Daten	159
2. Ausgleichende Interpolation mit orthogonalen Polynomen	166
3. Spline Funktionen	171

Tabellen

Literatur

A. Einführung in die Statistik

1. Einleitung

Experimentalphysik: Messung von physikalischen Größen
↑
exakte Bestimmung i.a. nicht möglich

↳ Statistische Methoden notwendig

- zur Interpretation der Ergebnisse
- beim Vergleichen und Kombinieren von Messergebnissen

Beispiel für Messung einer physikalischen Größe:

1. Messung: 27.62

2. Messung: 28.57

weitere Messungen: 27.19, 29.12, 24.41, 26.90;
29.80, 26.90, 26.99, 28.72

Mögliche Ursachen der Variabilität:

- ① systematischer Effekt (Wackeln der Apparatur, zeit- oder temperaturabhängiger Einfluss) → im Prinzip kontrollierbar

→ ② statistische Streuung = Element der Variabilität, das prinzipiell nicht kontrolliert werden kann

Welche Aussagen sind trotz der Variabilität noch möglich?

Das x -Intervall $27 < x \leq 28$ wird betrachtet:

Von den $N=10$ Messungen fallen $n=2$ in dieses Intervall

$$N = 10 \quad \frac{n}{N} = 20\%$$

$$N = 10 \quad \frac{n}{N} = 20\%$$

Weitere Messungen:

$$N = 100 \quad \frac{n}{N} = 9\%$$

$$N = 1000 \quad \frac{n}{N} = 12.9\%$$

$$N \rightarrow \infty \quad \frac{n}{N} = 12.3975\%$$

Im Prinzip kann $\frac{n}{N}$ beliebig genau bestimmt werden.

→ ermöglicht Definition einer Wahrscheinlichkeit
d.h. die unermeldbare Variabilität kann durch eine Wahrscheinlichkeitsaussage beschrieben werden.

z.B. ein einzelner Messwert x fällt mit der Wahrscheinlichkeit $p = 0.123975$ in das Intervall $27.0 < x \leq 28.0$

Ereignis mit Eigenschaft e : x fällt in $27.0 < x \leq 28.0$

Empirische Definition der Wahrscheinlichkeit:

Die Wahrscheinlichkeit p für Ereignis e ist der (empirische) Grenzwert des Verhältnisses $\frac{n}{N}$;
$$\frac{n}{N} \rightarrow p(e) \quad \text{für } N \rightarrow \infty$$

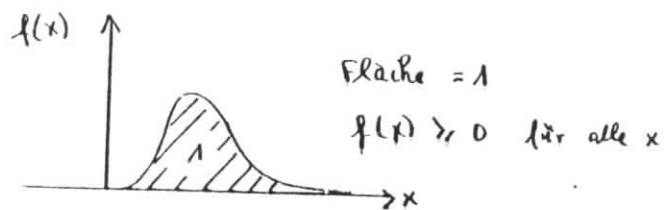
 N = Gesamtzahl der Ereignisse, davon n Ereignisse mit Eigenschaft e .

Bei einer kontinuierlichen Variablen x läßt sich diese Betrachtung für jedes Intervall anstellen. → Definition einer Funktion

$$p(e) = \int_e f(x) dx$$

mit der Eigenschaft: $\int_{-\infty}^{+\infty} f(x) dx = 1$

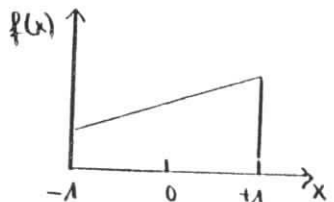
$f(x)$ = Wahrscheinlichkeitsdichte, kurz Dichte



in der Physik: $f(x)$ = "Verteilungsfunktion"
(wird in der Statistik mit anderer Bedeutung verbunden)

Physik: In der Physik wird die Verteilung der einzelnen Meßwerte x häufig durch den zu untersuchenden physikalischen Prozess selbst bestimmt.

Beispiel: Zerfall von Spin $\frac{1}{2}$ Teilchen



Zerfallswinkelverteilung $x = \cos \vartheta$
 $-1 \leq x \leq +1$
Dichte $f(x) = \frac{1+ax}{2}$ a = Parameter

Hier ist die Wahrscheinlichkeit

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x) dx$$

berechenbar, wenn a bekannt ist.

Häufiges Problem: Aus Einzelwerten x_1, x_2, x_3, \dots
Wert für Parameter a bestimmen, der dem "wahren" Wert a^0
möglichst nahe kommt

Physik:

Ergebnisse von Experimenten werden in folgender Form angegeben:

$$\hat{a} \pm \Delta a \quad \Delta a = \text{Standardabweichung}$$

Was ist damit gemeint?

i. a. folgendes:

$$P(\hat{a} - \Delta a < a^0 < \hat{a} + \Delta a) = 68.3\%$$

d.h. man macht eine Aussage über den "wahren" Wert a^0 .

(inverse Wahrscheinlichkeit)

(In der Sprache der Statistik wie diese Wahrscheinlichkeit)
 $P = 0$ oder 1 ; unterschiedlicher Sprachgebrauch

Warum 68.3%? gilt bei Normalverteilung.

bei $\Delta a = \text{Standardabweichung}$

Wenn bei einem physikalischen Ergebnis keine weiteren Angaben (außer $\hat{a} \pm \Delta a$) erfolgen, sollte man den o.a. "Wahrscheinlichkeitsgehalt" voraussetzen können.

→ keinen "Maximalfehler" angeben, keine Sicherheitsfaktoren wie z.B. \otimes , da dies Vergleiche und Kombinationen mit anderen Messungen unmöglich macht.

oder: "Fehler" mit expliziter Angabe der Wahrscheinlichkeit.

Moderne Definition von Wahrscheinlichkeit:

Ω = Menge aller elementaren Ereignisse e_i (exklusiv)

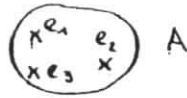
$P(e_i)$ = Wahrscheinlichkeit für Eintreten von e_i

hat die Eigenschaften:

- (1) $P(e_i) \geq 0$ für alle i
- (2) $P(e_i \text{ oder } e_j) = P(e_i) + P(e_j)$
- (3) $\sum_{\Omega} P(e_i) = 1$

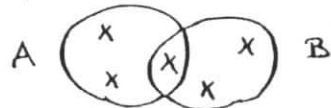
Nichtelementares Ereignis A = Menge von Elementaren Ereignissen

$P(A)$ = Wahrscheinlichkeit, daß ein $e_i \in A$ eintritt



Additionsgesetz: A und B seien sich nicht gegenseitig ausschließende Mengen von elementaren Ereignissen e_i

$$P(A \text{ oder } B) = P(A) + P(B) - P(A \text{ und } B)$$



Bedingte Wahrscheinlichkeit:

$P(A|B)$ = Wahrscheinlichkeit für Eintreten von A, wenn B eingetreten ist

$$P(A \text{ und } B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$\Rightarrow P(A|B) = P(B|A) \frac{P(A)}{P(B)} \quad P(B) \neq 0$$

(Bayes Theorem)

Unabhängigkeit:

A und B heißen unabhängig, wenn $P(A|B) = P(A)$

Beispiel: In einem Experiment soll der Zerfall eines Elementarteilchens $X \rightarrow e + \dots$ untersucht werden. Zum Nachweis der Zerfälle wird ein Cerenkov-Fähler benutzt.

$P(A)$ = Wahrscheinl. für Zerfall $X \rightarrow e + \dots$
(= 0.4, aus anderen Experiment bekannt)

$P(B)$ = Wahrscheinlichkeit für Auslösen des C-Fählers bei einem Ereignis
(= 0.25, mit Apparatur gemessen)

• Gesucht:

$P(A|B)$ = Wahrscheinlichkeit, daß ein Ereignis tatsächlich der gesuchte Elementarzerfall $X \rightarrow e + \dots$, wenn der Cerenkov-Fähler ausgelöst ist.

Aus Geometrie und Efficiency kann berechnet werden.

$P(B|A)$ = Wahrscheinlichkeit, daß ein Elementarzerfall den Cerenkov-Fähler auslöst (= 0.5)

$$\bullet P(A|B) = P(B|A) \frac{P(A)}{P(B)} = 0.5 \frac{0.4}{0.25} = 0.8$$

\Rightarrow 20% Untergrundereignisse

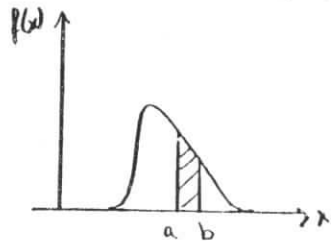
2. Zufallsvariable

2.1 Kontinuierliche und diskrete Zufallsvariable

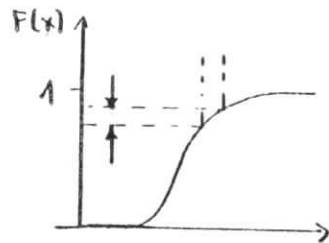
Zufälliges Ereignis = Ereignis mit mehr als einem möglichen Resultat.
 z.B. Messung einer kontinuierlichen Variablen x .
 Ω = Menge der möglichen Werte

$f(x)$ = Dichtefunktion für kont. Variable, definiert durch

- (1) $f(x) \geq 0$ für $x \in \Omega$
- (2) $\int_{\Omega} f(x) dx = 1$
- (3) $P(a < x \leq b) = \int_a^b f(x) dx$ = Wahrscheinlichkeit dafür, dass x in das Intervall $a < x \leq b$ fällt.



Dichtefunktion $f(x)$



(Kumulative) Verteilungsfunktion $F(x)$

$$F(x) = \int_{-\infty}^x f(t) dt = P(t \leq x)$$

monoton nicht fallende Funktion

$$F(-\infty) = 0 \quad F(+\infty) = 1$$

$$P(a < x \leq b) = F(b) - F(a)$$

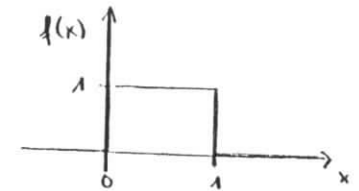
Wenn $F(x)$ differenzierbar ist:

$$f(x) = \frac{dF(x)}{dx}$$

Beispiel für Verteilung einer kontinuierlichen Variablen:
 Gleichverteilung $U(0,1)$

$$f(x) = 1 \quad 0 \leq x < 1$$

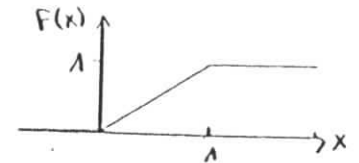
$$f(x) = 0 \quad \text{sonst}$$



$$F(x) = 0 \quad x < 0$$

$$F(x) = x \quad 0 \leq x < 1$$

$$F(x) = 1 \quad x \geq 1$$

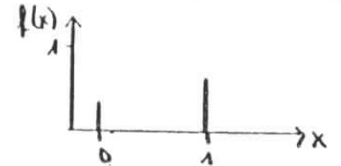


Entsprechende Begriffe für Variable, die nur diskrete Werte annehmen können: Integrale \rightarrow durch Summen ersetzt

Beispiel: Variable x kann nur Werte 0 und 1 annehmen

$$f(x) = p^x (1-p)^{1-x} \quad x=0,1$$

$$f(x) = 0 \quad \text{sonst}$$

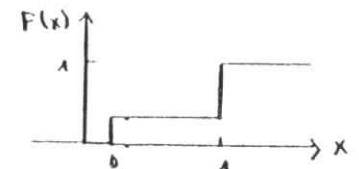


$$\Rightarrow f(0) = 1-p \quad ; \quad f(1) = p$$

$$\sum_{x=0,1} f(x) = (1-p) + p = 1$$

$$F(x) = \sum_{t \leq x} f(t)$$

(Treppenfunktion)



z.B. die Ausprägungswahrscheinlichkeit eines Fehlers kann durch die obige Dichte beschrieben werden.

2.2 Erwartungswerte

x sei Zufallsvariable mit Dichte $f(x)$.

Ist es möglich, die Verteilung von x durch eine oder mehrere Parameter zu charakterisieren?

→ Lokalisationsparameter (wo?)
Dispersionsparameter (wie "breit"?)

Mittelwert oder Erwartungswert der Zufallsvariablen x definiert durch

$$E(x) = \int_{\Omega} x \cdot f(x) dx \quad \text{kontin. Variable}$$

$$\text{bzw. } E(x) = \sum_{\Omega} x_i \cdot f(x_i) \quad \text{diskret Variable}$$

(Bezeichnung: $E(x) = \langle x \rangle$)

Anschaulich: die möglichen Werte von x werden mit der Wahrscheinlichkeit, mit der sie auftreten, gewichtet.

Beispiel:

Gleichverteilung $u(0,1)$

$$E(x) = \int_0^1 x dx = \frac{1}{2} x^2 \Big|_0^1 = \frac{1}{2}$$

Diskrete Verteilung mit $f(x) = p^x (1-p)^{1-x}$, $x = 0,1$

$$E(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

↑
Erwartungswert selbst ist kein möglicher Wert von x .

$g(x)$ = beliebige Funktion der Zufallsvariablen x

Erwartungswert von $g(x)$ definiert durch

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

Beispiel: $g(x) = x^2$ bei Gleichverteilung $u(0,1)$

$$E(x^2) = \langle x^2 \rangle = \int_0^1 x^2 dx = \frac{1}{3} x^3 \Big|_0^1 = \frac{1}{3}$$

Der Operator der Erwartung ist linear, denn

$$E(a \cdot g(x) + b \cdot h(x)) = a \cdot E(g(x)) + b \cdot E(h(x))$$

Für die spezielle Wahl $g(x) = x$ erhält man den Mittelwert $\langle x \rangle$, dies ist der einfachste Lokalisationsparameter einer Verteilung.

Einfachster Dispersionsparameter einer Verteilung

$$= \text{Erwartungswert von } g(x) = (x - \langle x \rangle)^2$$

= Varianz

$$V(x) = E[(x - \langle x \rangle)^2] = \int_{\Omega} (x - \langle x \rangle)^2 f(x) dx$$

↑

Varianz = nichtnegative Größe

Standardabweichung σ = Quadratwurzel der Varianz

Für die Varianz gilt:

$$V(x) = E[(x - \langle x \rangle)^2] = E[x^2 - 2\langle x \rangle x + \langle x \rangle^2]$$

$$= E(x^2) - 2\langle x \rangle E(x) + \langle x \rangle^2$$

$$= E(x^2) - [E(x)]^2$$

Standardabweichung σ = Maß dafür, wie stark eine Zufallsvariable um ihren Mittelwert streut.

Wenn $f(x)$ bekannt ist, kann die Wahrscheinlichkeit dafür, daß ein einzelner Wert x in das Intervall $\langle x \rangle - \sigma$ bis $\langle x \rangle + \sigma$ fällt, durch $\int_{\langle x \rangle - \sigma}^{\langle x \rangle + \sigma} f(x) dx$ berechnet werden.

Beispiel: Normalverteilung $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\langle x \rangle)^2}{2\sigma^2}}$
 $P_1 = 68.3 \%$

Beispiel: Gleichverteilung $U(0,1)$

$$V(x) = \int_0^1 (x - \frac{1}{2})^2 dx = \left[\frac{1}{3}x^3 - \frac{1}{2}x^2 + \frac{1}{4}x \right]_0^1 = \frac{1}{3} - \frac{1}{4} + \frac{1}{4} = \frac{1}{12}$$

$$\sigma^2 = \frac{1}{12} \quad \sigma = \frac{1}{\sqrt{12}} = 0.2887$$

$$P_1 = 57.74 \%$$

Ist eine entsprechende Wahrscheinlichkeitsaussage möglich, wenn von $f(x)$ nur $\langle x \rangle$ und σ bekannt sind?

Tschebyscheffsche Ungleichung:

$$P(|x - \langle x \rangle| > k\sigma) \leq \frac{1}{k^2}$$

Tschebyscheffsche Ungleichung:

$$P(|x - \langle x \rangle| > k\sigma) \leq \frac{1}{k^2}$$



Diese Aussage ist schwach, da nur wenige Voraussetzungen gemacht werden ($\langle x \rangle$ und σ endlich), jedoch für allgemeine Beweise wichtig.

Beispiel: Normalverteilung

Tschebyscheff: $P(|x - \langle x \rangle| > 4\sigma) \leq 0.0625$
 \downarrow
exakt: 0.000063

Momente einer Verteilung:

Die Erwartungswerte von $g(x) = x^n$
 bzw. von $g(x) = (x - \langle x \rangle)^n$
 für $n = 0, 1, \dots$ werden Momente μ_n' bzw. zentrale
 Momente μ_n genannt:

$$\left| \begin{array}{ll} \mu_n' = E(x^n) & n\text{-tes Moment} \\ \mu_n = E((x - \langle x \rangle)^n) & n\text{-tes zentrales Moment} \end{array} \right.$$

Mit der Angabe aller Momente ist auch die Dichte $f(x)$
 bekannt (\rightarrow charakteristische Funktionen)

Spezielle Momente:

$$\mu_1' = \langle x \rangle \quad \text{Mittelwert}$$

$$\mu_2 = V(x) = \sigma^2 \quad \text{Varianz}$$

Spezielle Bezeichnungen:

- | | |
|---|---|
| <p>① $\beta_1 = \frac{\mu_3'}{\mu_2'}^2$ Schiefe</p> <p>$\gamma_1 = \sqrt{\beta_1}$ Koeffizient der Schiefe</p> | <p>$= 0$ bei zu
 $\langle x \rangle$ symmetr.
 Verteilung</p> |
| <p>② $\beta_2 = \frac{\mu_4'}{\mu_2'^2}$</p> <p>$\gamma_2 = \beta_2 - 3$ Koeffizient der
 Kurtosis</p> | <p>$\gamma_2 = 0$ für
 Normd. verteilung</p> |

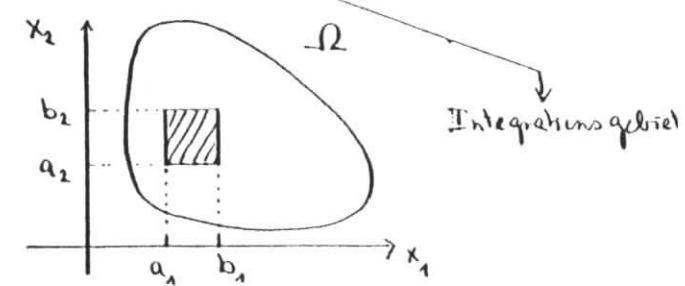
2.3 Mehrkomponentige Zufallsvariable

Zufallsvariable können mehrkomponentige Größen sein.

z.B. Streuexperiment: Winkel ϑ und Impulsbetrag p
 einzelner Schüttenläufchen werden gemessen.
 \rightarrow Vektor (ϑ, p) , gemeinsame
 Wahrscheinlichkeitsdichte $f(\vartheta, p)$

$f(x_1, x_2)$ = gemeinsame Dichtefunktion einer zweikomponen-
 tigen Zufallsvariablen (x_1, x_2)

- $$\left| \begin{array}{l} (1) \quad f(x_1, x_2) \geq 0 \quad \text{für } x_1, x_2 \in \Omega \\ (2) \quad \iint_{\Omega} f(x_1, x_2) dx_1 dx_2 = 1 \\ (3) \quad P(a_1 < x_1 \leq b_1, a_2 < x_2 \leq b_2) \\ = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2 \end{array} \right.$$



(Kumulative) Verteilungsfunktion:

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_1 dt_2$$

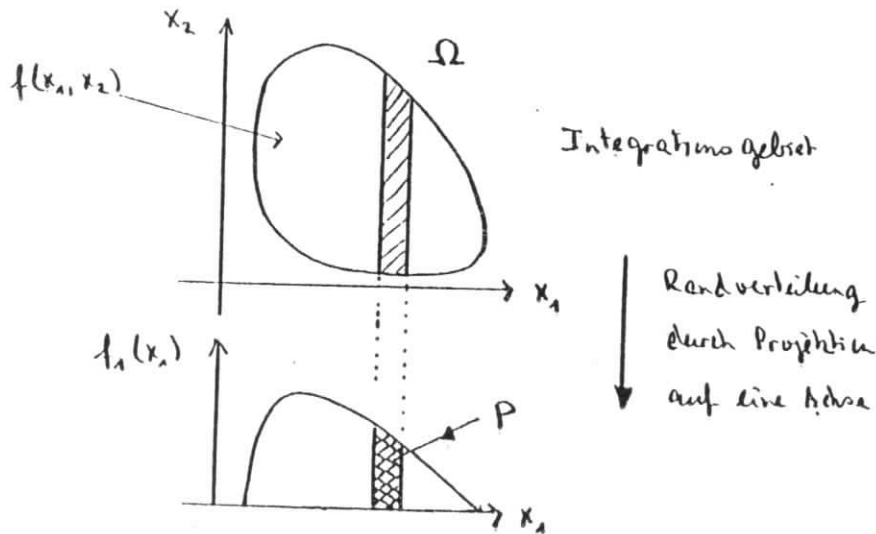
Wenn $F(x_1, x_2)$ differenzierbar: $f(x_1, x_2) = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} F(x_1, x_2)$

Randverteilungen und bedingte Verteilungen für
zweikomponentige Zufallsvariable (X_1, X_2)

Randverteilung: Die Wahrscheinlichkeit für
 $a_1 < X_1 \leq b_1$, X_2 beliebig

ist gegeben durch:

$$P(a_1 < X_1 \leq b_1) = \int_{a_1}^{b_1} \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 dx_1$$



Integrationsgebiet

Randverteilung
durch Projektion
auf eine Achse

Die "Randverteilung" von X_1 wird definiert durch

$$f_1(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2$$

dann es gilt:

$$P(a_1 < X_1 \leq b_1) = \int_{a_1}^{b_1} f_1(x_1) dx_1$$

Beispiel: Im Streuexperiment (n.v.) wird nur die Verteilung
des Winkels θ gemessen \rightarrow Randverteilung $f(\theta)$

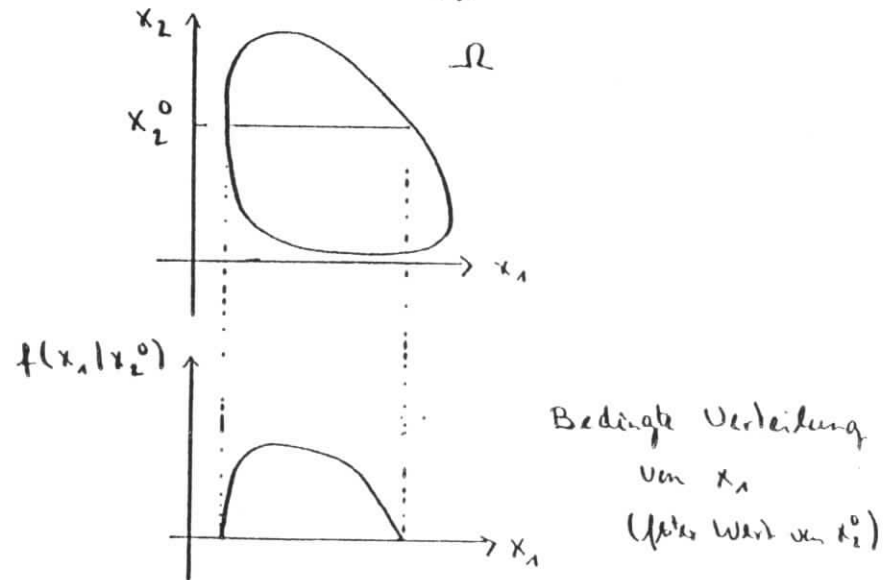
Bedingte Verteilung: Die Wahrscheinlichkeit für
 $a_1 < X_1 \leq b_1$, $X_2 = X_2^0$ (fester Wert)
ist gegeben durch

$$P(a_1 < X_1 \leq b_1 | X_2^0) = \int_{a_1}^{b_1} \underbrace{f(x_1 | X_2^0)}_{\text{Dichte der bedingten Verteilung}} dx_1$$

Dichte der bedingten Verteilung

$$f(x_1 | X_2^0) \propto f(x_1, X_2^0)$$

Normierung: $f(x_1 | X_2^0) = \frac{f(x_1, X_2^0)}{\int_{-\infty}^{+\infty} f(x_1, X_2^0) dx_1} = \frac{f(x_1, X_2^0)}{f_2(X_2^0)}$



Bedingte Verteilung
von X_1
(fester Wert von X_2^0)

Beispiel: Im Streuexperiment (n.v.) ist die Verteilung
von θ bei festem Wert des Inputbeitrags p
die bedingte Verteilung.

Aus der Beziehung

$$f(x_1, x_2) = f(x_1 | x_2) f_2(x_2) = f(x_2 | x_1) f_1(x_1)$$

folgt das Bayes Theorem für kontinuierliche Verteilungen:

$$f(x_1 | x_2) = f(x_2 | x_1) \frac{f_1(x_1)}{f_2(x_2)} \quad f_2(x_2) > 0$$

Unabhängig Variable: Die Variablen x_1 und x_2 heißen unabhängig, wenn die gemeinsame Dichtefunktion $f(x_1, x_2)$ als Produkt von zwei Funktionen $f_1(x_1) \cdot f_2(x_2)$ geschrieben werden kann:

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \quad \Rightarrow x_1, x_2 \text{ unabhängig}$$

Mittelwerte und Kovarianzen

Mittelwerte $\langle x_1 \rangle$ und $\langle x_2 \rangle$ einer zweikomponentigen Zufallsvariablen mit Dichte $f(x_1, x_2)$ definiert durch:

$$\langle x_1 \rangle = E(x_1) = \iint_{\Omega} x_1 \cdot f(x_1, x_2) dx_1 dx_2$$

$$\langle x_2 \rangle = E(x_2) = \iint_{\Omega} x_2 \cdot f(x_1, x_2) dx_1 dx_2$$

Bei einer zweikomponentigen Zufallsvariablen gibt es 3 (verschiedene) zentrale Momente 2. Ordnung:

$$\sigma_{11} = \sigma_1^2 = E[(x_1 - \langle x_1 \rangle)^2]$$

$$\sigma_{22} = \sigma_2^2 = E[(x_2 - \langle x_2 \rangle)^2]$$

$$\sigma_{12} = E[(x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle)]$$

σ_{11} und σ_{22} heißen Varianzen, σ_1 und σ_2 die zugehörigen Standardabweichungen. Die Größe σ_{12} ($= \sigma_{21}$) heißt Kovarianz zwischen x_1 und x_2 . σ_{11} , σ_{22} , σ_{12} und σ_{21} lassen sich zu einer Kovarianzmatrix zusammenfassen:

$$V \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Die Kovarianz σ_{12} läßt sich schreiben als

$$\sigma_{12} = \rho_{12} \sigma_1 \sigma_2$$

ρ_{12} = Korrelationskoeffizient; $-1 \leq \rho \leq +1$

Es gilt:

$$\begin{aligned} \sigma_{12} &= E[(X_1 - \langle X_1 \rangle)(X_2 - \langle X_2 \rangle)] \\ &= E[X_1 X_2 - \langle X_1 \rangle X_2 - X_1 \langle X_2 \rangle + \langle X_1 \rangle \langle X_2 \rangle] \\ &= E[X_1 X_2] - \langle X_1 \rangle \langle X_2 \rangle \end{aligned}$$

Die lineare Funktion

$$a X_1 + b X_2 \quad a, b \text{ beliebig}$$

wird betrachtet:

$$\begin{aligned} V(a X_1 + b X_2) &= E[(a(X_1 - \langle X_1 \rangle) + b(X_2 - \langle X_2 \rangle))^2] \\ &= a^2 \sigma_{11} + 2ab \sigma_{12} + b^2 \sigma_{22} \\ &= (a \ b) \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \\ &\quad \text{Vektor} \quad \text{Matrix} \quad \text{Vektor} \end{aligned}$$

Der Erwartungswert eines Quadrats kann nicht negativ sein

→ (lineare Algebra): Determinante ≥ 0

$$\text{Determinante } \sigma_{11} \sigma_{22} - \sigma_{12}^2 \geq 0$$

$$1 - \frac{\sigma_{12}^2}{\sigma_{11} \sigma_{22}} \geq 0$$

$$1 - \rho_{12}^2 \geq 0$$

$$\Rightarrow \boxed{-1 \leq \rho_{12} \leq +1}$$

Fall $\rho_{12} = +1$ oder -1 : Variable X_1 und X_2

linear abhängig, d.h.

Kovarianzmatrix singulär

(eine Variable überflüssig)

Unabhängige Variable: $f(X_1, X_2) = f_1(X_1) \cdot f_2(X_2)$

→ dann gilt:

$$\begin{aligned} E(X_1 X_2) &= \iint X_1 f_1(X_1) dX_1 \cdot X_2 f_2(X_2) dX_2 \\ &= \langle X_1 \rangle \langle X_2 \rangle \end{aligned}$$

Vergleich mit Ausdruck für σ_{12} zeigt: $\rho_{12} = 0$

bei unabh. Variablen

Der umgekehrte Schluss $\rho_{12} = 0 \Rightarrow X_1, X_2$ unabhängig ist nicht notwendig richtig.

Mehrkomponentige Zufallsvariable:

Die Erweiterung der behandelten Begriffe auf mehrkomponentige Zufallsvariable ist einfach.

Für eine n -komponentige Zufallsvariable (n -Vektor) gibt es $(n^2 + n) / 2$ verschiedene zentrale Momente 2. Ordnung, die zu einer symmetrischen $n \times n$ Matrix zusammengefasst werden können:

$$V(\vec{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \dots & \sigma_{nn} \end{pmatrix}$$

$$\text{mit } \vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

Erwartungswert $\langle \vec{x} \rangle$ und Varianz $V(\vec{x})$ in
Matrixschreibweise:

$$E(\vec{x}) = \langle \vec{x} \rangle = \int_{\Omega} \vec{x} \cdot f(\vec{x}) dx_1 dx_2 \dots dx_n$$

\downarrow
 $f(x_1, x_2, \dots, x_n)$

$$V(\vec{x}) = E \left[\underbrace{(\vec{x} - \langle \vec{x} \rangle) (\vec{x} - \langle \vec{x} \rangle)^T}_{n \times n \text{ Matrix}} \right]$$

T heißt transponiert

z.B. für $n=2$:

$$(\vec{x} - \langle \vec{x} \rangle) (\vec{x} - \langle \vec{x} \rangle)^T = \begin{pmatrix} (x_1 - \langle x_1 \rangle)^2 & (x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle) \\ (x_1 - \langle x_1 \rangle)(x_2 - \langle x_2 \rangle) & (x_2 - \langle x_2 \rangle)^2 \end{pmatrix}$$

symmetrische Matrix

2.4 Transformationen

Wichtig für Anwendungen (\rightarrow Frage der Fehlerfortpflanzung
 error propagation)

Lineare Transformationen:

Zufallsvariable x \rightarrow transformierte Variable $y = ax + b$
 Dichte $f(x)$ \rightarrow Dichte $g(y)$?

Intervall $(x, x + dx)$ abgebildet auf $(y, y + dy)$:

$$g(y) dy = f(x) dx$$

$$g(y) = \frac{f(x)}{\left| \frac{dy}{dx} \right|} = \frac{f(x)}{|a|}$$

$$x = \frac{1}{a}(y - b)$$

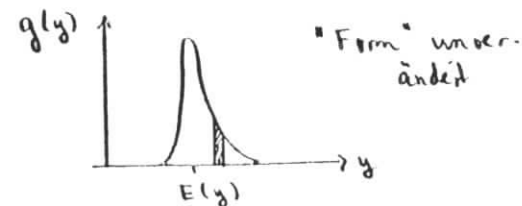
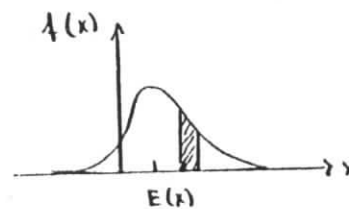
$$g(y) = \frac{1}{|a|} f\left(\frac{1}{a}(y - b)\right)$$

Mittelwert und Varianz:

$$E(y) = E(ax + b) = a E(x) + b$$

$$\begin{aligned} V(y) &= E[(ax - a E(x))^2] \\ &= a^2 E[(x - \langle x \rangle)^2] \\ &= a^2 V(x) \end{aligned}$$

$$\sigma_y = a \sigma_x$$



Nichtlineare Transformation

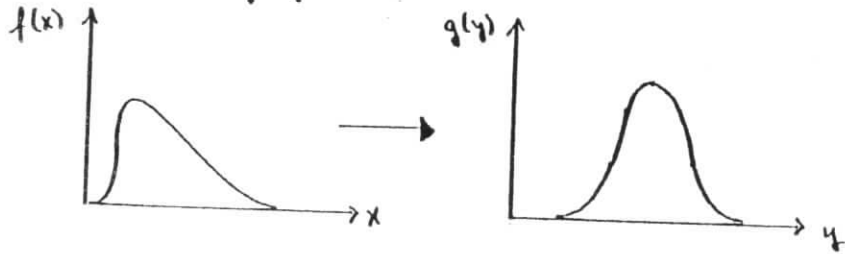
Zufallsvariable x → transformierte Variable $y = y(x)$
 Dichte $f(x)$ → Dichte $g(y)$?

① Abbildung $y \rightarrow x$ eindeutig:

$$g(y) dy = f(x) dx$$

$$g(y) = \frac{f(x)}{\left| \frac{dy}{dx} \right|}$$

Da $\frac{dy}{dx}$ nicht konstant ist, wird die "Form" der Verteilung geändert



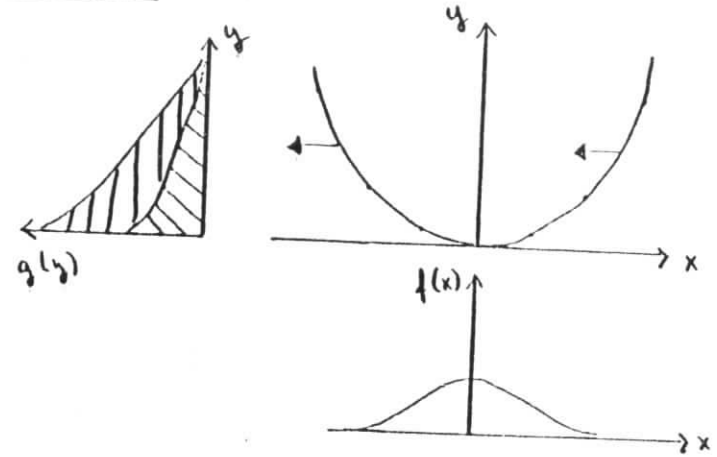
Durch geschickte Wahl der Transformation läßt sich häufig eine asymmetrisch verteilte Meßgröße x in eine symmetrisch verteilte Meßgröße transformieren, bei der viele statistische Betrachtungen einfacher sind.

② Abbildung $y \rightarrow x$ nicht eindeutig

→ summieren über alle Zweige:
 |
 (eindeutig Abb.)

$$g(y) = \sum \frac{f(x)}{\left| \frac{dy}{dx} \right|}$$

Beispiel: $y = x^2$



$$\frac{dy}{dx} = 2x$$

$$g(y) = \frac{1}{2|x|} (f(x) + f(-x))$$

$$= \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y}))$$

Mehrkomponentige Zufallsvariable

Zufallsvariable x_1, x_2 → transformierte Variable
 Dichte $f(x_1, x_2)$ →
 $y_1 = y_1(x_1, x_2)$
 $y_2 = y_2(x_1, x_2)$
 Dichte $g(y_1, y_2)$?

Es muß gelten:

$$g(y_1, y_2) dy_1 dy_2 = f(x_1, x_2) dx_1 dx_2$$

Entsprechendes Problem bei Transformation von
 Flächenintegralen (allgem. Volumenintegralen)

$$g(y_1, y_2) = f(x_1, x_2) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right|$$

$$= f(x_1, x_2) \left| \frac{\partial(y_1, y_2)}{\partial(x_1, x_2)} \right|^{-1}$$

$$J \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \text{Funktionaldeterminante (Jacobian-Det.)}$$

$$J \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Bei nichteindeutiger Abbildung $y \rightarrow x$ muß
 über alle Zweige summiert werden.

Beispiel:

$$y_1 = x_1 \qquad x_1 = y_1$$

$$y_2 = x_1 + x_2 \qquad x_2 = y_2 - y_1$$

Die Zufallsvariablen x_1, x_2 seien unabhängig, d.h.
 $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$

Jacobi-Determinante

$$J \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix} \quad |J| = 1$$

$$g(y_1, y_2) = f_1(x_1) f_2(x_2)$$

$$= f_1(y_1) f_2(y_2 - y_1)$$

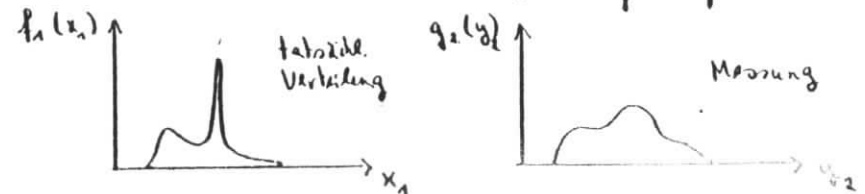
Anwendungsbeispiel: Veränderung (Störung) einer
 "idealen" Dichte $f_1(x_1) = f_1(y_1)$ durch
 begrenzte Messgenauigkeit

Tatsächliche Verteilung einer physikalischen Größe $y_1 = x_1$
 sei gegeben durch Dichte $f_1(y)$. Gemessen wird die
 Verteilung der Variablen y_2 , die durch Messfehler (x_2)
 verändert ist: $y_2 = x_1 + x_2$ nicht direkt
messbar

Rendverteilung von y_2 : Aufflämpfen

$$g_2(y_2) = \int f_1(y_1) f_2(y_2 - y_1) dy_1$$

Faltungintegral



Angenäherte Transformation:

Eine nichtlineare Transformation ist i.a. sehr kompliziert

Die wichtigsten Parameter einer Verteilung,

Mittelwert und Varianz,

lassen sich angenähert einfach transformieren

(Voraussetzung: Transf. \approx linear in Bereich, wo die Dichte groß ist)

Transformation: $x \rightarrow y(x)$

Entwicklung von $y = y(x)$ um $x = \langle x \rangle$

$$\bullet y(x) = y(\langle x \rangle) + (x - \langle x \rangle) \left. \frac{dy}{dx} \right|_{x=\langle x \rangle} + \frac{1}{2} (x - \langle x \rangle)^2 \left. \frac{d^2 y}{dx^2} \right|_{x=\langle x \rangle} + \dots$$

Mittelwert (bis 2. Ordnung):

$$E(y) = y(\langle x \rangle) + \underbrace{E(x - \langle x \rangle)}_{=0} \left. \frac{dy}{dx} \right|_{x=\langle x \rangle} + \frac{1}{2} \underbrace{E((x - \langle x \rangle)^2)}_{\sigma_x^2} \left. \frac{d^2 y}{dx^2} \right|_{x=\langle x \rangle} + \dots$$

$$\Rightarrow \langle y \rangle = y(\langle x \rangle) + \frac{1}{2} \sigma_x^2 \left. \frac{d^2 y}{dx^2} \right|_{x=\langle x \rangle} + \dots$$

i.a. weglassen bei Anwendung

Wenn ein Meßwert x in einen Wert y umgerechnet wird, hat x selbst eine S.D. von σ_x , d.h. die Umrechnung selbst hat bereits Fehler 1. Ordnung.

Varianz:

$$y(x) - \langle y \rangle = (x - \langle x \rangle) \left. \frac{dy}{dx} \right|_{x=\langle x \rangle} + \dots$$

$$V(y - \langle y \rangle) = V[(x - \langle x \rangle)] \left[\left. \frac{dy}{dx} \right|_{x=\langle x \rangle} \right]^2 + \dots$$

$$\sigma_y^2 \approx \sigma_x^2 \left(\left. \frac{dy}{dx} \right|_{x=\langle x \rangle} \right)^2$$

(einfachste Form des Gesetzes der Fehlerfortpflanzung)

Beispiel: Gemessen $q = 0.5 \pm 0.1$

\uparrow S.D.

$$\text{gesucht } p = \frac{1}{q} = \dots \pm \dots$$

$$\hat{p} = \frac{1}{\hat{q}} + \frac{1}{2} \sigma_q^2 (-2) (\hat{q})^{-3}$$
$$= 2.0 - 0.08$$

bei diesem Problem weglassen (siehe vorige Seite)

$$\sigma_p^2 \approx \sigma_q^2 \frac{1}{\hat{q}^4} = 0.16$$

$$\Rightarrow p = 2.0 \pm 0.4$$

Methode sinnvoll, wenn $\frac{q}{\sigma_q} \gg 1$

Transformation

$$x_1, x_2, \dots, x_p \longrightarrow y = y(x_1, x_2, \dots, x_p)$$

Entwicklung:

$$y = y(\langle x_1 \rangle, \langle x_2 \rangle, \dots, \langle x_p \rangle) + \sum_{i=1}^p \frac{\partial y}{\partial x_i} \Big|_{x_i = \langle x_i \rangle} (x_i - \langle x_i \rangle) + \dots$$

Mittelwert:

$$\langle y \rangle = E(y) = y(\langle x_1 \rangle, \langle x_2 \rangle, \dots, \langle x_p \rangle)$$

Varianz:

$$\begin{aligned} V(y) &= E[(y - \langle y \rangle)^2] \\ &= E\left[\left(\sum_{i=1}^p (x_i - \langle x_i \rangle) \frac{\partial y}{\partial x_i} \right)^2 \right] \\ &= \sum_{i=1}^p \sum_{j=1}^p \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} E[(x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)] \end{aligned}$$

$$V(y) = \sum_{i=1}^p \sum_{j=1}^p \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} V_{ij}(\vec{x})$$

Element ij der Matrix $V(\vec{x})$

Sonderfall: alle Zufallsvariablen x_i unabhängig (!)
 $\Rightarrow V(\vec{x}) =$ Diagonalmatrix mit $V_{ii}(\vec{x}) = \sigma_i^2$

$$V(y) = \sigma_y^2 = \sum_{i=1}^p \left(\frac{\partial y}{\partial x_i} \right)^2 \sigma_i^2$$

\equiv Quadratische Addition der Einzelfehler.
(nicht vorsichtshalber die Absolutbeträge addieren [wie es in einigen Lehrbüchern steht])

Allgemeiner Fall (ohne Beweis):

$$\begin{aligned} x_1, x_2, \dots, x_p \longrightarrow y_1 &= y_1(x_1, x_2, \dots, x_p) \\ y_2 &= y_2(x_1, x_2, \dots, x_p) \\ y_3 &= y_3(x_1, x_2, \dots, x_p) \\ &\vdots \\ y_n &= y_n(x_1, x_2, \dots, x_p) \end{aligned}$$

$$A_{ij} = \frac{\partial y_i}{\partial x_j}$$

Element ij der Matrix A

$$A = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_p} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \dots & \frac{\partial y_n}{\partial x_p} \end{pmatrix}$$

$$V_{ij}(\vec{y}) = \sum_{k=1}^p \sum_{l=1}^p A_{ik} V_{kl}(\vec{x}) A_{jl}$$

$i, j = 1 \dots n$

In Matrix-Schreibweise:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$V(\vec{y}) = A V(x) A^T$$

$A^T =$ Transp. Matrix

= allgemeines Gesetz der Fehlerfortpflanzung

3. Spezielle Verteilungen

3. Die Normalverteilung

Normalverteilung (Gauß-Verteilung) $N(\mu, \sigma^2)$

$$\text{Dichte } f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

Parameter Mittelwert = μ
 Varianz = σ^2 (σ = Standardabw.)

Durch lineare Transformation $y = \frac{x-\mu}{\sigma}$ erhält man für y die standardisierte Normalverteilung $N(0, 1)$:

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

Kumulative Verteilungsfunktion,

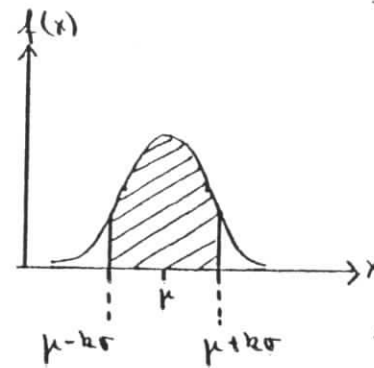
$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}y'^2} dy' = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$= \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{y}{\sqrt{2}}\right) \right) \quad \text{tabelliert}$$

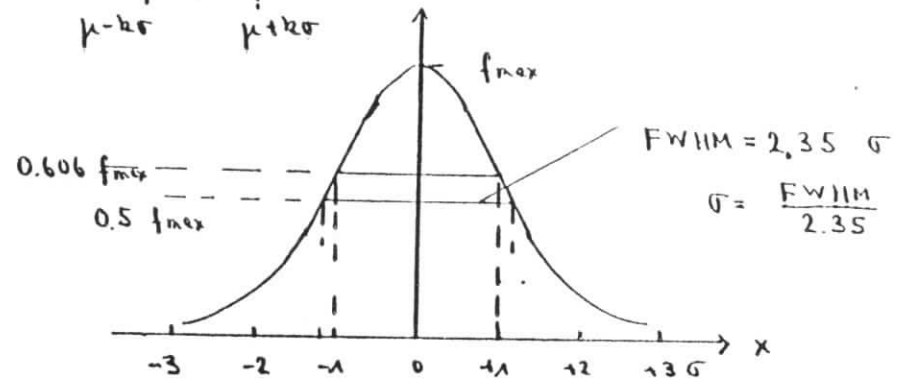
↳ FORTRAN-Funktion ERF

Berechnung des Wehrscheinlichkeitsgehalts:

$$\begin{aligned} P(\mu - k\sigma < x < \mu + k\sigma) &= \\ &= P\left(-k < \frac{x-\mu}{\sigma} < +k\right) = \Phi(k) - \Phi(-k) \end{aligned}$$



$k = 1$	$P = 68.3 \%$
$k = 2$	$P = 95.5 \%$
$k = 3$	$P = 99.8 \%$



Überragende Bedeutung der Nvt für Theorie der Beobachtungsfehler → Zentraler Grenzwertsatz

X_1, X_2, \dots, X_N sei Folge von Zufallsvariablen mit Mittelwert μ_i , Varianz σ_i^2 .

Betrachtet wird Summe $S = \sum X_i$ (= Zufallsvariable)

$$\left. \begin{aligned} E(S) &= \sum \mu_i \\ V(S) &= \sum \sigma_i^2 \end{aligned} \right\} \text{ gilt allgemein}$$

Satz: Für $N \rightarrow \infty$ ist die Zufallsvariable

$$\frac{S - \sum \mu_i}{\sqrt{\sum \sigma_i^2}}$$

normalverteilt mit Mittelwert 0, Varianz 1.

(unabhängig von den Dichten der Einzelnen X_i !)

3.2 Binomial- und Poisson-Verteilung

N Ereignisse, jedes Ereignis hat:

- entweder Eigenschaft A (mit Wahrscheinlichkeit p)
- oder \bar{A} (mit Wahrscheinlichkeit $1-p$)

Wie groß ist die Wahrscheinlichkeit, bei N Ereignissen gerade r Ereignisse mit Eigenschaft A zu haben?

Wahrscheinlichkeit, daß

die ersten r Ereignisse Eigenschaft A haben, p^r
 die nächsten $(N-r)$ " \bar{A} " $(1-p)^{N-r}$

$$\Rightarrow p^r (1-p)^{N-r}$$

Diese Reihenfolge $\underbrace{A, A, A, \dots, A}_r, \underbrace{\bar{A}, \bar{A}, \bar{A}, \dots, \bar{A}}_{N-r}$ ist eine

von $\binom{N}{r} = \frac{N!}{r!(N-r)!}$ möglich.

$$P(r) = \binom{N}{r} p^r (1-p)^{N-r} \quad \text{Binomial-Verteilung}$$

Parameter der Verteilung: N (ganze Zahl)
 p mit $0 \leq p < 1$

$$\begin{aligned} \text{Summe: } \sum_{r=0}^N P(r) &= \sum_{r=0}^N \binom{N}{r} p^r (1-p)^{N-r} \\ &= (p + (1-p))^N = 1 \quad \Rightarrow \text{normiert} \end{aligned}$$

Erwartungswert von r:

$$\begin{aligned} E(r) &= \sum r \cdot P(r) \\ &= \sum_{r=0}^N \frac{r N!}{(N-r)! r!} p^r q^{N-r} \quad q = 1-p \\ &= N \cdot p \sum_{r=1}^N \frac{(N-1)!}{(N-r)! (r-1)!} p^{r-1} q^{N-r} \quad \begin{matrix} s = r-1 \\ s = 0 \dots N-1 \end{matrix} \\ &= N \cdot p \underbrace{\sum_{s=0}^{N-1} \frac{(N-1)!}{(N-1-s)! (s)!} p^s q^{N-1-s}}_{=1} = N \cdot p \end{aligned}$$

Erwartungswert von r^2 :

$$E(r^2) = \sum r^2 P(r) = N(N-1)p^2 + Np$$

Varianz:

$$V(r) = E(r^2) - [E(r)]^2 = Np(1-p)$$

Zusammenfassung: $\langle r \rangle = N \cdot p$
 $\sigma = \sqrt{N \cdot p(1-p)}$

Beispiel für Binomialverteilung:

Wieviel Versuche sind notwendig, um mit einer Wahrscheinlichkeit $\alpha = 90\%$ mindestens ein Ereignis mit der Eigenschaft A zu finden, wenn $p(A) = 0$.

$$\Rightarrow N \text{ so bestimmen, daß } P(r \geq 1) \geq \alpha$$

$$\begin{aligned}
 P(r \geq 1) &\geq \alpha \\
 1 - P(r=0) &\geq \alpha \\
 P(r=0) &\leq 1 - \alpha \\
 (1-p)^N &\leq 1 - \alpha \\
 N &\geq \frac{\ln(1-\alpha)}{\ln(1-p)}
 \end{aligned}$$

Zahlenwerte $p=0.5$ und $\alpha=0.9$

$$N \geq \frac{\ln 0.1}{\ln 0.5} = 3.32 \Rightarrow N = 4$$

Übergang zur Poissonverteilung:

Bei der Binomialverteilung erhält man im Mittel

$$\langle r \rangle = \lambda = N \cdot p$$

Ereignisse mit der Eigenschaft A.

$P(r)$ wird betrachtet bei festem λ und anwachsendem N (d.h. p wird kleiner):

$$p = \frac{\lambda}{N}$$

$$P(r) = \frac{N!}{r!(N-r)!} \left(\frac{\lambda}{N}\right)^r \left(1 - \frac{\lambda}{N}\right)^{N-r} \quad \text{Binomialverteilung}$$

$$= \frac{\lambda^r}{r!} \left(1 - \frac{\lambda}{N}\right)^N \frac{1}{\left(1 - \frac{\lambda}{N}\right)^r} \frac{N(N-1)(N-2) \dots (N-r+1)}{N^r}$$



$$\left(1 - \frac{\lambda}{N}\right)^N \frac{1 \cdot \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{r-1}{N}\right)}{\left(1 - \frac{\lambda}{N}\right)^r}$$

$$\xrightarrow{N \rightarrow \infty} e^{-\lambda}$$

$$\xrightarrow{N \rightarrow \infty} 1$$

Für $N \rightarrow \infty$ wird also

$$P(r) = e^{-\lambda} \frac{\lambda^r}{r!}$$

Poisson-Verteilung

Man erhält also aus der Binomialverteilung die Poissonverteilung.

Normierung: $\sum_{r=0}^{\infty} P(r) = e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} = e^{-\lambda} e^{\lambda} = 1$

Die Poisson-Verteilung hat nur einen Parameter, λ

Erwartungswert: $E(r) = \langle r \rangle = \lambda$

$$V(r) = \sigma^2 = \lambda^2$$

$$\sigma = \lambda$$

Rechnungen dazu:

$$\begin{aligned}
 E(r) &= \sum_{r=0}^{\infty} r \frac{e^{-\lambda} \lambda^r}{r!} = \lambda \sum_{r=1}^{\infty} \frac{e^{-\lambda} \lambda^{r-1}}{(r-1)!} \\
 &= \lambda \sum_{s=0}^{\infty} \frac{e^{-\lambda} \lambda^s}{s!} = \lambda \quad (s=r-1)
 \end{aligned}$$

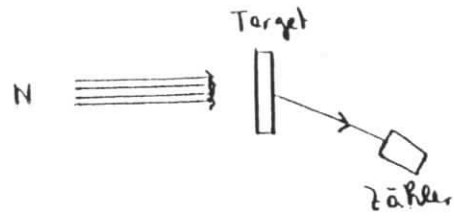
$$E(r^2) = \lambda^2 + \lambda$$

$$V(r) = E(r^2) - [E(r)]^2 = \lambda$$

Voraussetzungen für Anwendung der Poissonverteilung:

- Zufällige Ereignisse, deren Wahrscheinlichkeit, Eigenschaft A zu haben, klein ist.
- (Anzahlen!)

Beispiel: Zählrate bei Streuexperiment



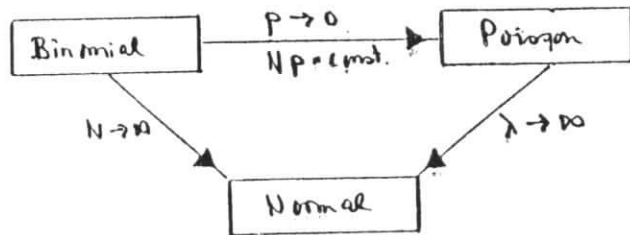
N groß
 p klein
 $\lambda = N \cdot p$ Poissonverteilt

anderes Beispiel:

Anzahl Einträge in einem Intervall eines Histogramms

Unter bestimmten Bedingungen nähern sich Binomial-
 verteilung und Poissonverteilung der Normalverteilung:

Zusammenhang:



Annäherung der Verteilungen weniger gut bei Werten, die
 weit vom Erwartungswert entfernt sind.

Regeln: Bereich guter Näherung durch Normalverteilung

Binomial	$N \geq 8$	p nicht nahe bei 0, 1
Poisson	$\lambda \geq 5$	

Beispiel: Binomialverteilung mit $N = 300$, $p = \frac{1}{3}$
 d.h. $E(r) = 100$

Gesucht:

$$P(85 \leq r \leq 115) = \sum_{r=85}^{115} \binom{N}{r} p^r (1-p)^{N-r}$$

aufwendig zu berechnen!

Näherung durch Normalverteilung:

Transformation $x = \frac{r - \langle r \rangle}{\sigma}$ $\langle r \rangle = N \cdot p = 100$
 $\sigma = \sqrt{N \cdot p \cdot (1-p)}$
 $= 10 \cdot \sqrt{\frac{2}{3}}$

Grenzen: $x_1 = \frac{85 - 100}{10 \sqrt{\frac{2}{3}}} = -1.84$
 $x_2 = +1.84$

$$P(85 \leq r \leq 115) \approx \Phi(1.84) - \Phi(-1.84) = 0.934$$

Beispiel: Transformation von Poisson verteilten Daten

Die Beurteilung von Histogrammen wird dadurch erschwert
 daß die einzelnen Intervallinhalte (Poissonverteilt)
 ungleiche Varianzen haben (Inhalt $y_i \rightarrow$ Varianz σ^2
 und $\sigma = \sqrt{y_i}$). Die folgenden Transformationen stabilisieren die Varianzen:

$\sqrt{y_i}$ Varianz 0.25, $\sigma = 0.5$

$\sqrt{y_i} + \sqrt{y_i + 1}$ Varianz 1.00, $\sigma = 1.00$
 \uparrow
 $\pm 6\%$ für $E(y_i) \geq 1$

3.3 χ^2 Verteilung

Wichtig für statistische Tests.

n Zufallsvariable X_i seien unabhängig verteilt nach $N(0,1)$.

$N(0,1)$ = Normalverteilung, Mittelwert 0, Varianz 1.

Dann ist die Zufallsvariable

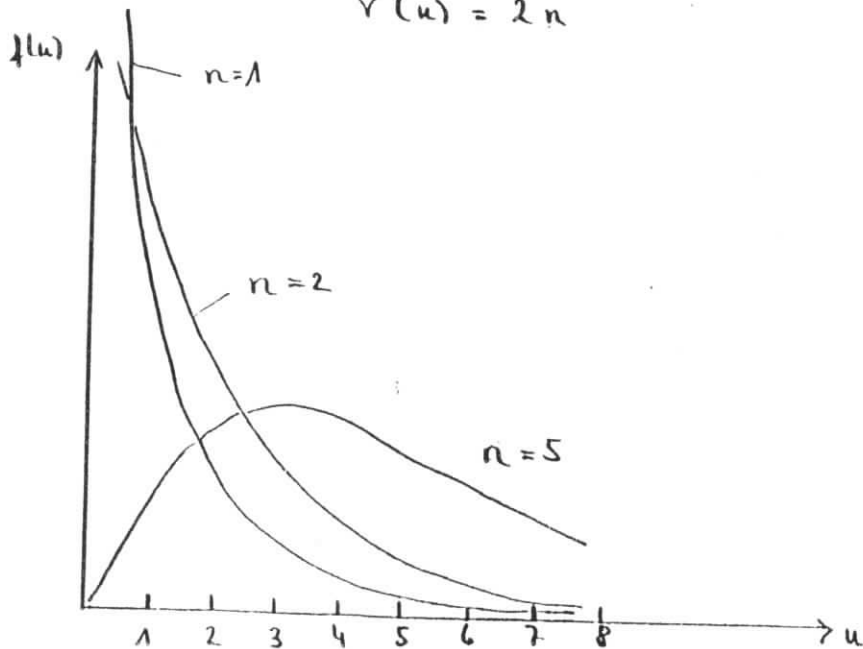
$$u = \sum_{i=1}^n X_i^2$$

χ^2 verteilt mit n Freiheitsgraden.

Dichtefunktion: $f(u) = \frac{1}{2} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} e^{-u/2} \frac{1}{\Gamma(\frac{n}{2})}$

Erwartungswert: $E(u) = n$

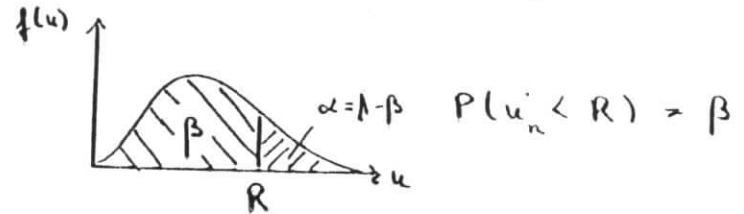
$$V(u) = 2n$$



Für großes n nähert sich die χ^2 -Verteilung der Normalverteilung mit Mittelwert n , und Varianz $2n$; (gute Näherung für $n > 30$).

$$\text{d.h. } z = \frac{u_n - n}{\sqrt{2n}} \rightarrow N(0,1)$$

In Anwendungen ist häufig die Grenze R zu vorgegebener Wahrscheinlichkeit β zu betrachten:



R wird als der β -Wert der χ^2 -Verteilung mit n Freiheitsgraden bezeichnet. $R = \chi^2_{\beta}(n)$

Diese Größe R ist in Statistikbüchern tabelliert für einige Werte β und $n = 1, 2, 3, \dots$

Eine andere Konvention ist.

$$R = \chi^2_{1-\beta}(n) = \chi^2_{\alpha}(n)$$

Diese Konvention wird in der Funktion

$$P = \text{PROB}(R, N) \quad \text{bzw.} \quad \text{PROCH}(R, N)$$

benutzt, um R in P umzurechnen.

(je größer R , desto kleiner P)

4. Bestimmung von Parametern

4.1 Eigenschaften von Methoden zur Parameterbestimmung

Problem: Es liegt eine Menge von Beobachtungen (Stichprobe)

vor: x_1, x_2, \dots, x_N (Daten)

Aus diesen Daten soll ein Parameter a bestimmt werden:

$\hat{a} = \hat{a}(x_1, x_2, \dots, x_N)$ Stichprobenfunktion

Beispiel: Winkelverteilung beim Zerfall eines Elementarteilchens soll gemessen werden, sie ist theoretisch von der Form: $f(x) = \frac{1 + ax}{2}$; $x = \cos \vartheta$

$N=4$ Zerfälle werden beobachtet: $x_i = 0.8; 0.5; 0.0; -0.5$
Gesucht: $\hat{a} = \hat{a}(x_1, x_2, x_3, x_4)$

Der Parameter soll so bestimmt werden, daß die in den Daten stehende Information möglichst vollständig ausgenutzt wird.

Forderungen:

① Konsistenz

\hat{a} konsistent, wenn $\lim_{N \rightarrow \infty} \hat{a}_N = a^0$ (wahr. Wert)

(asymptotische Eigenschaft)

② Erwartungstreue

\hat{a} basiert auf N Einzelwerten (N endlich)

Systematische Abweichung = Bias =
 $= b_N(\hat{a}) = E(\hat{a}) - a^0 = E(\hat{a} - a^0)$

Ein Schätzwert \hat{a} heißt erwartungstreu, wenn
 $b_N(\hat{a}) = 0$ für alle N

(3) Effektivität

Effektivität $\text{eff}(\hat{a}) = \frac{I^{-1}}{V(\hat{a})} \leq 1$

I = Information, berechenbar unter gewissen Voraus.
(u.a. Dichte $f(x, a)$ bekannt für erwartungstreue Schätzwerte)

Ein Schätzwert heißt effektiv, wenn minimale Varianz erreicht wird: $V_{\min}(\hat{a}) = I^{-1}$; $\text{eff} =$

(4) Robustheit

Unempfindlichkeit des Schätzwertes \hat{a} gegen kleine Abweichungen von der zugrundeliegenden Verteilung (gegen "falsche" Daten)

Beispiel zu (1) und (2):

Bestimmung von Mittelwert μ und Varianz σ^2
aus Daten $x_i, i = 1 \dots N$

Methode:

$$\hat{\mu} = \frac{1}{N} \sum x_i = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum (x_i - \bar{x})^2$$

Ist $\hat{\mu}$ erwartungstreu und konsistent? ja

$$E(\hat{\mu}) = E\left(\frac{1}{N} \sum x_i\right) = \frac{1}{N} E\left(\sum x_i\right) = \frac{1}{N} \sum \underbrace{E(x_i)}_{\mu}$$

= μ
 \Rightarrow erwartungstreu

$$V(\hat{\mu}) = V\left(\frac{1}{N} \sum x_i\right) = \frac{1}{N^2} V\left(\sum x_i\right) = \frac{N \sigma^2}{N^2} = \frac{\sigma^2}{N}$$

d.h. $V(\hat{\mu}) \rightarrow 0$ für $N \rightarrow \infty \Rightarrow$ konsistent

Ist $\hat{\sigma}^2$ erwartungstreu? nein

$$\hat{\sigma}^2 = \frac{1}{N} \sum \left[(x_i - \mu) - (\bar{x} - \mu) \right]^2$$

$$= \frac{1}{N} \sum \left[(x_i - \mu)^2 + (\bar{x} - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu) \right]$$

$$= \frac{1}{N} \left[\sum (x_i - \mu)^2 - \sum (\bar{x} - \mu)^2 \right]$$

$$E(\hat{\sigma}^2) = \frac{1}{N} \left\{ \underbrace{E\left[\sum (x_i - \mu)^2\right]}_{N \cdot \sigma^2} - \underbrace{E\left[\sum (\bar{x} - \mu)^2\right]}_{N \cdot \frac{\sigma^2}{N}} \right\}$$

$$= \frac{N-1}{N} \sigma^2 = \sigma^2 - \frac{1}{N} \sigma^2$$

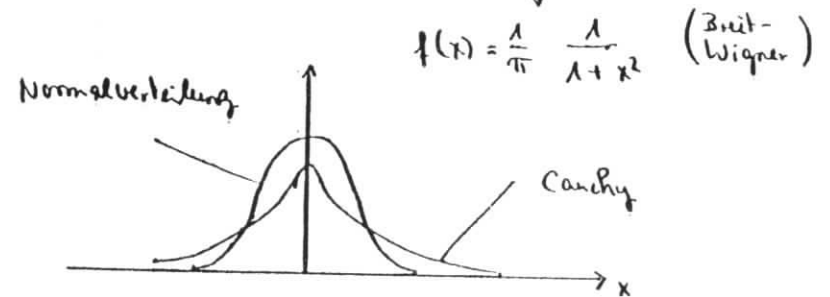
= Bias

Durch Neuformulierung der Formeln kann der Bias behoben werden:

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Beispiel zu (4): Robustheit

Das Zentrum einer asymmetrischen Verteilung soll bestimmt werden, Experimentelle Verteilung = Normalverteilung + kleine Beimischung einer Cauchy-Verteilung.



einige Werte bei großen $|x|$ können den Mittelwert \bar{x} stark beeinflussen; Median (= 50% Wert der exp. Vert.) wird kaum gestört.

Methode	Asymptotische Effektivität	
	Normalvert.	Cauchy-Vert.
Mittelwert der Stichprobe	1.0	0.0
Median der Stichprobe	0.64	0.82

Effektivität kann für Mittelwert der Stichprobe zwischen 0 und 1 liegen \Rightarrow wenig robust
 Effektivität des Medians $\geq 64\%$...

Getrimmter Mittelwert =

Mittelwert bei Weglassen von 23% der kleinsten und 23% der größten Werte

$$\tilde{x} = \frac{1}{0.54 N} \sum_{i=0.23 N}^{0.77 N} x_i$$

\tilde{x} hat Effektivität $\geq 82\%$ für Normalverteilung, Cauchy-Verteilung und Doppel-Exponentialverteilung

Robustheit = wichtige Forderung, jedoch bei komplizierten Problemen schwer zu verwirklichen.

Allgemeine Methoden zur Bestimmung von Parametern

- Momentenmethode
- Maximum-Likelihood-Methode
- Methode der kleinsten Quadrate

4.2 Momentenmethode

Älteste allgemeine Methode, wegen einfacher Anwendung wichtig, obwohl keine optimalen Eigenschaften

Prinzip: Gleichsetzen von experimentellen und theoretischen Momenten

Beispiel:

Dichte = Winkelverteilung $f(x) = \frac{1+ax}{2}$; $x \in (-1, 1)$
 N Werte $x_i = \cos \alpha_i$; gemessen
 Parameter a ist zu bestimmen

theoretischer Mittelwert: $E(x) = \int_{-1}^1 x f(x) dx$
 $= \int_{-1}^1 x \frac{1+ax}{2} dx = \frac{1}{3} a$

experimenteller Mittelwert: $\bar{x} = \frac{1}{N} \sum x_i$ ← gleichsetzen

durch Gleichsetzen: $\hat{a} = 3 \bar{x}$

Zahlenwerte: $x_i = 0.8 ; 0.5 ; 0.0 ; -0.5$

$\bar{x} = \frac{1}{4} \cdot 0.8 = 0.2$

$\Rightarrow \hat{a} = 0.6$

AllgemeinErwartungswert einer Funktion $h(x)$,

$$E[h(x)] = \int h(x) f(x) dx \Rightarrow g(a)$$

Stichprobe

$$\bar{h}(x) = \frac{1}{N} \sum_{i=1}^N h(x_i)$$

Gleichsetzen: $g(\hat{a}) = \bar{h}(x)$

$$\hat{a} = g^{-1}(\bar{h}(x))$$

Nach dem zentralen Grenzwertsatz ist \bar{h} asymptotisch normalverteilt mit Varianz $\frac{1}{N} V(h(x))$

$$\Rightarrow V(\hat{a}) = \frac{1}{N} \left(\frac{\partial g^{-1}}{\partial a} \right)^2 V(h(x))$$

Fortsetzung des Beispiels:

$$h \equiv x$$

$$V(h) = V(x) = E\left[\left(x - \frac{1}{3}a\right)^2\right] = \frac{1}{3} - \frac{1}{9}a^2$$

$$V(\hat{a}) = \frac{1}{N} (3 - a^2)$$

Zahlenwerte:

$$V(\hat{a}) = \frac{1}{4} (3 - 0.6^2) = 0.66$$

$$\sigma = 0.81 = \sqrt{0.66}$$

$$\longrightarrow a = 0.60 \pm 0.81$$

Eigenschaften der Momentenmethode.

Für $h(x)$ wird Normalität asymptotisch erreicht, jedoch durch die Umkehrfunktion $[g^{-1}(\bar{h}(x))]$ wieder zerstört.

Möglicher Bias $b_N(\hat{a}) \propto \frac{1}{N}$

Asymptotisch nicht effektiv (d.h. nicht kleinstmögliche Var.)

... aber einfache Methode für Näherungswert.

4.3 Maximum-Likelihood-Methode

N Werte (Messwerte) x_1, x_2, \dots, x_N gegeben
Dichte $f(x, a)$ bekannt bei unbekanntem Wert
des Parameters a

Gemeinsame Dichtefunktion der Messwerte x_i :

$$L(x|a) = L(x_1, \dots, x_N|a) = \prod_{i=1}^N f(x_i, a)$$

Funktion $L(x|a)$, als Funktion des Parameters a
betrachtet bei gegebenen Messdaten, heißt
Likelihood-Funktion

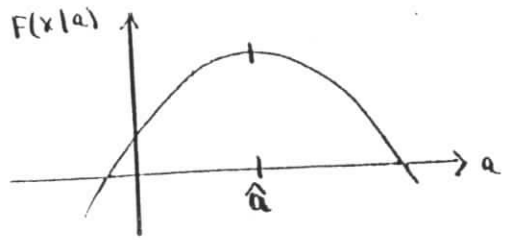
Schätzwert \hat{a} für Parameter a bei der MLM =
der Wert, für den $L(x|a)$ maximal wird.

Praktische Anwendung:

$$F(x|a) = \ln L(x|a) = \sum_{i=1}^N \ln f(x_i, a)$$

notwendige Bedingung für Maximum ist Verschwinden
der Ableitung nach a :

$$\frac{\partial}{\partial a} F(x|a) = 0 \quad \text{Likelihood-Gleichung}$$



Beispiel: $f(x) = \frac{1+ax}{2}$

$$F(a) = \sum_{i=1}^N \ln \frac{1+ax_i}{2}$$

$$\frac{\partial F}{\partial a} = \sum_{i=1}^N \frac{x_i}{1+ax_i} = 0 \quad \mapsto \hat{a}$$

Praktische Durchführung mit Newton-Methode:

Startwert $a^{(1)}$
$$a^{(i+1)} = a^{(i)} - \frac{\left(\frac{\partial F}{\partial a}\right)}{\left(\frac{\partial^2 F}{\partial a^2}\right)} \quad \begin{matrix} \text{iteratives} \\ \text{Verfahren} \end{matrix}$$

$$\frac{\partial^2 F}{\partial a^2} = - \sum_{i=1}^N \left(\frac{x_i}{1+ax_i}\right)^2$$

Allgemeine Eigenschaften der MLM: asymp. optimal

- (1) Konsistent, d.h. $\hat{a} \xrightarrow{N \rightarrow \infty} a^0$
- (2) "Invariant": Schätzwert \hat{b} einer Funktion $b(\cdot)$ ist $b(\hat{a})$. Allerdings kann nur Funktion $b(a)$ einen Bias von 0 zu

- (3) Asymptotische Normalität:
 \hat{a} ist asymptotisch normalverteilt mit minimaler

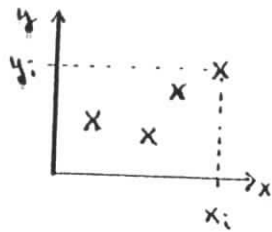
Varianz
$$v(\hat{a}) \rightarrow \left(E \left[\left(\frac{\partial \ln L}{\partial a} \right)^2 \right] \right)^{-1}$$

Praxis:
$$v(\hat{a}) = \left[\left(- \frac{\partial^2 F}{\partial a^2} \right)_{a=\hat{a}} \right]^{-1}$$

4.4 Methode der kleinsten Quadrate

Geradenanpassung als Beispiel

Gegeben: n Wertepaare (y_i, x_i)
durch Messung, alle y_i mit gleicher Genauigkeit.



Modell: $f(x) = a_1 + a_2 x$

a_1 und a_2 unbekannte Parameter, die durch die Messung bestimmt werden sollen.

Annahme: $E(y_i) = a_1 + a_2 x_i$

d.h. abgesehen von den Messfehlern beschreibt das Modell die Daten richtig.

Definition: Residuen $r_i = y_i - (a_1 + a_2 x_i)$
ist Funktion der Parameter a_1, a_2

Mögliche Kriterien für die Bestimmung optimaler Werte von a_1, a_2 :

- $Q(a_1, a_2) = \sum_i |r_i|$ minimal
(numerische Durchführung schwierig)
- $Q(a_1, a_2) = \max_i |r_i|$ minimal
Tschebyscheff - Ausgleichung
- $Q(a_1, a_2) = \sum_i r_i^2$ minimal
Methode der kleinsten Quadrate

Tschebyscheff - Ausgleichung (\rightarrow lineare Optimierung):

sinnvolles Prinzip, wenn z.B. für Interpolationszwecke eine tabellierte Abhängigkeit durch eine Funktion ausgedrückt werden soll; bei Messfehlern weniger günstig.

Methode der kleinsten Quadrate:

"optimale" Methode zur Bestimmung von Parametern bei mit Messfehlern behafteten Daten
 \rightarrow liefert Schätzwerte \hat{a}_1, \hat{a}_2 mit den Eigenschaften:

- ① Erwartungstreue: $E(\hat{a}) = a^0$
 \hookrightarrow wahren Wert
- ② effektiv, d.h. kleinste Varianz (unter gewissen allgemeinen Voraussetzungen - Gauß-Markoff-Theorem)

Außerdem ist die numerische Durchführung einfach.

Mathem. Durchführung der Geradenanpassung nach der Methode der kleinsten Quadrate

$$Q(a_1, a_2) = \sum_i r_i^2 = \sum_i (y_i - (a_1 + a_2 x_i))^2$$

$$= \sum_i (y_i^2 + a_1^2 + a_2^2 x_i^2 + 2a_1 a_2 x_i - 2a_1 y_i - 2a_2 x_i y_i)$$

Aufsuchen des Minimums von $Q(a_1, a_2)$ bzgl. a_1, a_2 durch Nullsetzen der Ableitungen von Q nach a_1, a_2 :

$$\frac{\partial Q}{\partial a_1} = \sum_i (2a_1 + 2a_2 x_i - 2y_i)$$

$$\frac{\partial Q}{\partial a_2} = \sum_i (2a_1 x_i + 2a_2 x_i^2 - 2x_i y_i)$$

Nullsetzen der Ableitungen

$$a_1 \sum 1 + a_2 \sum x_i = \sum y_i$$

$$a_1 \sum x_i + a_2 \sum x_i^2 = \sum x_i y_i$$

führt auf ein lineares Gleichungssystem für a_1, a_2

Praktische Durchführung:

① Bildung der Summen: $S_1 = \sum 1 = n$

$$S_x = \sum x_i$$

$$S_{xx} = \sum x_i^2$$

$$S_y = \sum y_i$$

$$S_{xy} = \sum x_i y_i$$

$$S_{yy} = \sum y_i^2$$

(wird später gebraucht)

② Lösung des lin. Gleichungssystems

$$a_1 S_1 + a_2 S_x = S_y$$

$$a_1 S_x + a_2 S_{xx} = S_{xy}$$

durch:

$$D = S_1 S_{xx} - S_x^2$$

$$\hat{a}_1 = \frac{1}{D} (S_{xx} S_y - S_x S_{xy})$$

$$\hat{a}_2 = \frac{1}{D} (S_1 S_{xy} - S_x S_y)$$

↳ einfache, direkte Methode (kein iteratives Verfahren).
zu Bestimmung der Schätzwerte \hat{a}_1, \hat{a}_2 .

Numerisches Beispiel:

① $S_1 = \sum 1 = 9.0$

$$S_x = \sum x_i = 45.0$$

$$S_{xx} = \sum x_i^2 = 285.0$$

$$S_y = \sum y_i = 40.9$$

$$S_{xy} = \sum x_i y_i = 230.0$$

$$S_{yy} = \sum y_i^2 = 203.75$$

② $D = S_1 S_{xx} - S_x^2 = 9 \cdot 285.0 - (45.0)^2 = 540.0$

$$\hat{a}_1 = \frac{1}{D} (S_{xx} S_y - S_x S_{xy}) = \frac{1}{540.0} (285.0 \cdot 40.9 - 45.0 \cdot 230.0) = 2.4194$$

$$\hat{a}_2 = \frac{1}{D} (S_1 S_{xy} - S_x S_y) = \frac{1}{540.0} (9 \cdot 230.0 - 45.0 \cdot 40.9) = 0.4250$$

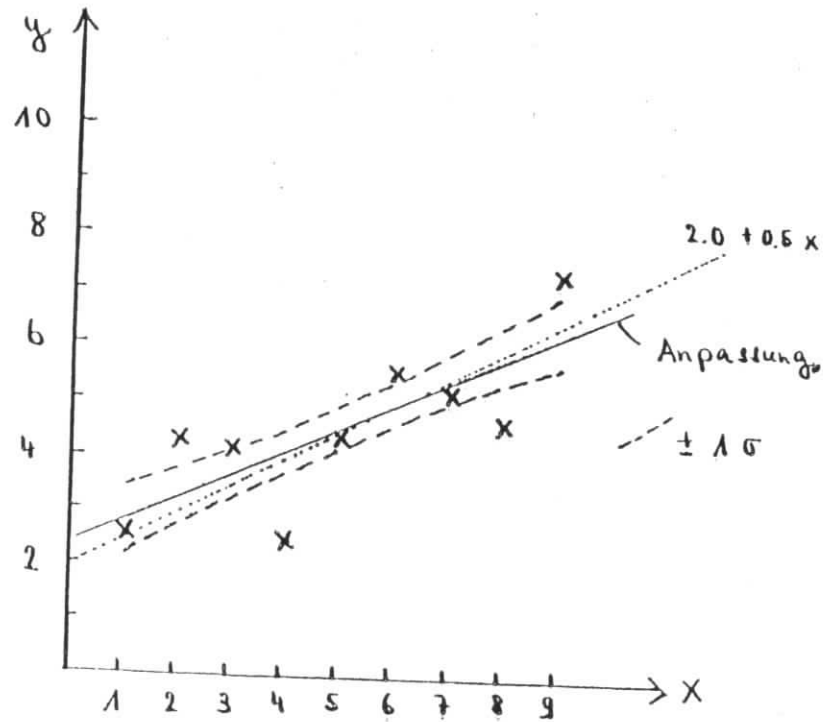
$$\Rightarrow \hat{y} = 2.4194 + 0.4250 \cdot x$$

$$(\text{"wahre" Gerade} = 2.0 + 0.5 \cdot x)$$

"Wahre" Gerade $y = 2.0 + 0.5x$

Meßpunkte mit Fehler (σ) = 1.0

i	x_i	y_i
1	1.0	2.6
2	2.0	4.3
3	3.0	4.2
4	4.0	2.5
5	5.0	4.4
6	6.0	5.6
7	7.0	5.2
8	8.0	4.7
9	9.0	7.4



Matrix - Schreibweise:

für weitergehende Betrachtungen ist die Matrix-Schreibweise geeigneter

- Meßwerte = Vektoren der Länge n

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- Vektor der Parameter (Länge 2): $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$

- Vektor der Erwartungswerte $E(y_i)$:

$$E(\vec{y}) = \underbrace{\begin{pmatrix} a_1 + a_2 x_1 \\ a_1 + a_2 x_2 \\ \vdots \\ a_1 + a_2 x_n \end{pmatrix}}_{A \vec{a}} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{n \times 2 \text{ - Matrix } A} \cdot \underbrace{\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}_{\vec{a}}$$

- Vektor der Residuen r_i :

$$\vec{r} = \vec{y} - A \vec{a}$$

- Skalar $Q = \sum r_i^2$ in Matrix-Schreibweise:

$$\begin{aligned} Q &= \vec{r}^T \vec{r} && T = \text{transponiert} \\ &= (\vec{y} - A \vec{a})^T (\vec{y} - A \vec{a}) \\ &= \vec{y}^T \vec{y} + \vec{a}^T A^T A \vec{a} - 2 \vec{a}^T A^T \vec{y} \end{aligned}$$

Was für eine Matrix ist $A^T A$?

- sie ist symmetrisch, denn ihre Transponierte ist identisch mit der ursprünglichen Matrix

$$(A^T A)^T = (A^T A)$$

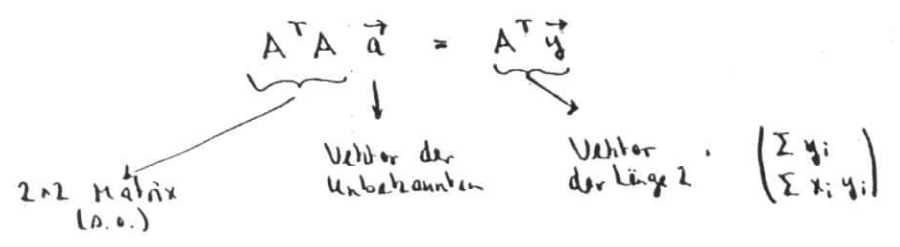
(Transponierte eines Matrixprodukts, Umkehren der Reihenfolge und transponieren aller Matrizen)

$$A^T A = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

Aufsuchen des Minimums von $Q(\vec{a})$ bezgl. des Vektors \vec{a} .
 Gradienten = Vektor der Ableitungen Null setzen.

$$\vec{\text{grad}} Q = \left(\frac{\partial Q}{\partial \vec{a}} \right) = 2 (A^T A \vec{a} - A^T \vec{y})$$

Nullsetzen führt auf lineares Gleichungssystem:



Formale Lösung:

$$\vec{\hat{a}} = \underbrace{(A^T A)^{-1}}_{\text{zu } (A^T A) \text{ inverse Matrix}} (A^T \vec{y})$$

Fehler der Schätzwerte \hat{a}_1 und \hat{a}_2 :

Die Fehler der Schätzwerte \hat{a}_1 und \hat{a}_2 ergeben sich durch Anwendung elementarer Formeln der Statistik auf den Fehler der Originaldaten (y_i). Es gibt keine geheimnisvollen "fit"-Fehler!

$\vec{\hat{a}}$ entsteht durch lineare Abbildung der y_i (bzw. \vec{y}):

$$\vec{\hat{a}} = (A^T A)^{-1} A^T \vec{y} = B \vec{y}$$

mit $B = (A^T A)^{-1} A^T = 2 \times n$ Matrix

Die Fehler der y_i seien bekannt und gleich σ .
↑
Standardabweichung

Kovarianzmatrix von \vec{y} :

$$V(\vec{y}) = \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{1}$$

Einheitsmatrix

Anwendung der Formel der Fehlerfortpflanzung (error propagate)

$$\vec{\hat{a}} = B \vec{y}$$

$$V(\vec{\hat{a}}) = B V(\vec{y}) B^T$$

(denn B ist eine feste Matrix, und unabhängig von den y_i)

einsetzen von $(A^T A)^{-1} A^T$ für B:

$$V(\vec{\hat{a}}) = \sigma^2 \underbrace{(A^T A)^{-1} A^T \mathbf{1} A (A^T A)^{-1}}_{(A^T A)^{-1}}$$

$$V(\vec{\hat{a}}) = \sigma^2 (A^T A)^{-1}$$

Kovarianzmatrix der \hat{a}_1, \hat{a}_2

Die Kovarianzmatrix (2x2 Matrix) der Schätzwerte $\vec{\hat{a}}$ kann in folgender Form geschrieben werden:

$$V(\vec{\hat{a}}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \quad \sigma_{12} = \sigma_{21}$$

Darin ist:

$$\sigma_{11} = \sigma_1^2 \quad \sigma_1 = \text{Standardabw. (Fehler) um } \hat{a}_1$$

$$\sigma_{22} = \sigma_2^2 \quad \sigma_2 = \text{Standardabw. (Fehler) um } \hat{a}_2$$

$$\sigma_{21} = \sigma_{12} = \rho_{12} \sigma_1 \sigma_2 \quad \sigma_{12} = \text{Kovarianz zwischen } \hat{a}_1 \text{ und } \hat{a}_2$$

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11} \sigma_{22}}} = \text{Korrelationsparameter zwischen } \hat{a}_1 \text{ und } \hat{a}_2, \quad -1 \leq \rho_{12} \leq +1$$

Das Ergebnis einer Geradenanpassung kann in folgender Form angegeben werden:

$$\begin{matrix} \hat{a}_1 \pm \sigma_1 \\ \hat{a}_2 \pm \sigma_2 \end{matrix}$$

Bei dieser Angabe fehlt der Wert von ρ_{12} bzw. σ_{12} , dieser ist jedoch wichtig, wenn mit den Parametern \hat{a}_1 und \hat{a}_2 weitere Rechnungen angestellt werden (siehe weiter unten); daher stets die volle Kovarianzmatrix für weitere Rechnungen aufbewahren.

Die bisher betrachteten Fehler (von \hat{a}_1, \hat{a}_2) werden "externe" Fehler genannt, weil sie auf Angaben von außen her beruhen ("die y_i haben Fehler σ ").

Es ist möglich, sogenannte "interne" Fehler aus den Daten selbst zu bestimmen; dies ist notwendig, wenn keine Angaben über die Genauigkeit der y_i vorliegen.

"Interne" Fehler aus Residuen:

$$r_i^2 \text{ im Mittel} = \sigma^2$$

σ = Standardabw. der Einzeldaten y_i

$$Q_{\min} = \sum r_i^2 = (n-2) \sigma^2$$

↓
minimaler Wert bzgl. a_1, a_2

↳ -2, weil 2 Parameter aus den Daten bestimmt werden

(diese Betrachtung läßt sich quantitativ beenden; insbes. wenn die y_i normalverteilt sind).

↳ Schätzwert $\hat{\sigma}^2 = s^2$ für σ^2 aus:

$$s^2 = \frac{1}{n-2} Q_{\min} \quad \text{genau für } n \gg 1$$

Dieser (interne, d.h. aus den Daten selbstbestimmte) Wert kann an Stelle von σ^2 in die vorher betrachtete Formel eingesetzt werden.

Berechnung von Q_{\min} :

$$Q_{\min} = \vec{y}^T \vec{y} + \vec{a}^T (A^T A) \vec{a} - 2 \vec{a}^T A^T \vec{y}$$

↑
 $\vec{a} = (A^T A)^{-1} A^T \vec{y}$ einsetzen

$$Q_{\min} = \vec{y}^T \vec{y} + \vec{a}^T \underbrace{(A^T A) (A^T A)^{-1} A^T \vec{y}}_1 - 2 \vec{a}^T A \vec{y}$$

$$= \vec{y}^T \vec{y} - \underbrace{\vec{a}^T A^T \vec{y}}_{\text{dies war bereits berechnet}}$$

Praktische Durchführung:

$$Q_{\min} = S_{yy} - (\hat{a}_1 S_y + \hat{a}_2 S_{xy})$$

→ Berechnung der Residuenquadrat-Summe ohne Berechnung der einzelnen r_i möglich.

$$V(\hat{\vec{a}}) = \frac{Q_{\min}}{n-2} (A^T A)^{-1} \quad \text{"interne" Fehler}$$

Fortsetzung des numerischen Beispiels:

① externe Fehler: alle y_i haben S.D. $\sigma = 1$

$$V(\hat{\vec{a}}) = \sigma^2 (A^T A)^{-1}$$

$$A^T A = \begin{pmatrix} S_x & S_x \\ S_x & S_{xx} \end{pmatrix} \quad (A^T A)^{-1} = \frac{1}{D} \begin{pmatrix} S_{xx} + S_x & -S_x \\ -S_x & S_x \end{pmatrix}$$

$$(A^T A)^{-1} = \frac{1}{540.0} \begin{pmatrix} 285.0 & -45.0 \\ -45.0 & 9.0 \end{pmatrix} = \begin{pmatrix} 0.5278 & -0.0833 \\ -0.0833 & 0.0167 \end{pmatrix}$$

$$\sigma_1 = \sqrt{0.5278} = 0.7265$$

$$\sigma_2 = \sqrt{0.0167} = 0.1291$$

$$S_{12} = -0.8885$$

$$\Rightarrow a_1 = 2.4194 \pm 0.7265 \quad \rightarrow 2.42 \pm 0.73$$

② interne Fehler - Abschätzung von σ^2 aus den Daten

$$Q_{\min} = S_{yy} - \hat{a}_1 S_y - \hat{a}_2 S_{xy}$$

$$= 203.75 - 2.4194 \cdot 40.9 - 0.4250 \cdot 230.0$$

$$= 7.0465$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-2} Q_{\min}$$

$$= \frac{1}{7} 7.0465 = 1.0066 \quad (\text{vergleichen mit})$$

hier kommt zufällig praktisch $\sigma = 1$ heraus

Berechnung ausgeglichener Werte

Wenn Schätzwerte \hat{a}_1, \hat{a}_2 vorliegen, kann für beliebigen ein "ausgeglichener" Wert gemäß

$$y = \hat{a}_1 + \hat{a}_2 x$$

berechnet werden.

Wie groß ist der Fehler des ausgeglichenen Wertes?

Hier wird die Bedeutung der Kovarianz σ_{12} klar werden.

Naive (falsche) Betrachtung:

\hat{a}_1 und \hat{a}_2 haben Fehler σ_1 bzw. σ_2 . Der Fehler von $y(x)$ ist daher

$$\Delta y = (\sigma_1^2 + \sigma_2^2 x^2)^{1/2}$$

(quadratische Addition der Einzelfehler)

Zahlenbeispiel: für $x = 9$ ist $y = 6.2444$

$$\Delta y = (0.5278 + 0.0167 \cdot 81)^{1/2}$$

$$= (1.8805)^{1/2} = 1.3713 > 1$$

das Ergebnis ist natürlich falsch!

Korrektes Verfahren:

$$y = \hat{a}_1 + \hat{a}_2 x$$

$$= (1 \quad x) \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \end{pmatrix} = \vec{b}^T \vec{\hat{a}} \quad \text{mit } \vec{b} = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

Lineare Abbildung von $\vec{\hat{a}}$

Formel für Fehlerfortpflanzung:

$$V(y) = \vec{b}^T V(\vec{\hat{a}}) \vec{b}$$

$$\begin{matrix} \uparrow \\ \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \end{matrix}$$

$$V(y) = \Delta y^2 = \underbrace{\sigma_{11}} + 2x \sigma_{12} + x^2 \sigma_{22}$$

zusätzlicher Term

Zahlenbeispiel: für $x = 9$

$$\Delta y = (0.5278 - 2 \cdot 9 \cdot 0.0833 + 81 \cdot 0.0167)^{1/2}$$

$$= (0.3811)^{1/2} = 0.6173 < 1.0$$

Die Formel für $V(y) = \Delta y^2$ kann in die folgende Form gebracht werden:

$$V(y) = \Delta y^2 = \sigma^2 \left[\frac{1}{S_1} + \frac{S_1}{D} \left(x - \frac{S_x}{S_1} \right)^2 \right]$$

↳ σ = S.D. der Einzelmessung

Für den zentralen Wert $x = \frac{S_x}{S_1} = \frac{\sum x_i}{n} = \bar{x}$ ist der Fehler des ausgefallenen Werts y am kleinsten; mit wachsendem Abstand von \bar{x} nimmt der Fehler Δy rasch zu
(vgl. Abb. 1; gestrichelte Kurve = $y \pm \Delta y$ (1 S.D.))

Anderer Ansatz für Geradenanpassung

Statt des Ansatzes $y = a_1 + a_2 x$

könnte auch der Ansatz $y = a_1' + a_2' (x - c)$

mit beliebiger Konstante c gewählt werden.

Umrechnung:

$$a_1 = a_1' - a_2' \cdot c$$

$$a_2 = a_2'$$

bzw.

$$a_1' = a_1 + a_2 \cdot c$$

$$a_2' = a_2$$

$$\begin{pmatrix} a_1' \\ a_2' \end{pmatrix} = T \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \text{mit } T = \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$$

Die Umrechnung der $\vec{\hat{a}}$ in $\vec{\hat{a}}'$ ist trivial; für die Kovarianzmatrix gilt:

$$V(\vec{\hat{a}}') = T V(\vec{\hat{a}}) T^T$$

$$\begin{matrix} \uparrow \\ \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \end{matrix}$$

$$V(\vec{\hat{a}}') = \left(\begin{array}{c|c} \sigma_{11} + 2c\sigma_{12} + c^2\sigma_{22} & \sigma_{12} + c\sigma_{22} \\ \hline \sigma_{21} + c\sigma_{22} & \sigma_{22} \end{array} \right)$$

Mit $c = -\frac{\sigma_{12}}{\sigma_{22}}$ verschwindet die Kovarianz $\sigma_{12} + c\sigma_{22}$,

die Parameter a_1' und a_2' werden unkorreliert. Dann

wird:

$$\sigma_{11} + 2c\sigma_{12} + c^2\sigma_{22} = \sigma_{11} - 2 \frac{\sigma_{12}^2}{\sigma_{22}} + \frac{\sigma_{12}^2}{\sigma_{22}^2}$$

$$= \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} < \sigma_{11}$$

→ durch eine bestimmte Transformation läßt sich die Kovarianz zum Verschwinden bringen; gleichzeitig werden die Varianzen σ'_{11} und σ'_{22} kleiner oder bleiben gleich.

Umgekehrt ausgedrückt und verallgemeinert: bei Wahl von stark miteinander korrelierten Parametern werden die Varianzen vergrößert.

Was bedeutet $c = -\frac{\sigma_{12}}{\sigma_{22}}$?

$$c = -\frac{s_x}{s_y} = -\bar{r}$$

d.h. Ansatz $y = a_1' + a_2' (x - \bar{x})$
 $\rightarrow a_1'$ und a_2' unkorreliert

auch numerische Vorteile:

$$(A^T A) \rightarrow \begin{pmatrix} \sum 1 & \sum (x_i - \bar{x}) = 0 \\ 0 & \sum (x_i - \bar{x})^2 \end{pmatrix}$$

= Diagonalmatrix, läßt zu invertieren

Folgerungen: Die Kovarianzmatrix ist von Bezugspunkten abhängig. Numerisch (und auch in anderer Beziehung) vorteilhaft ist ein Ansatz, bei dem die Kovarianzmatrix eine Diagonalmatrix ist, bzw die Korrelationskoeffizienten zumindest klein (gegen 1) sind.

5. Konfidenzintervalle

Konfidenzintervall = Vertrauensintervall

Methoden zur Schätzung von Parametern:

→ führt auf "beste" Werte \hat{a} von Parametern a .

Die Angabe von \hat{a} als Ergebnis eines Experiment alleine ist unvollständig.

• Ziel: Intervall $a_e \leq a \leq a_h$ (Konfidenzintervall)

angeben, das den wahren Wert a^0 mit vorgegebener Wahrscheinlichkeit β enthält.

(β = Confidence level, C.L.)

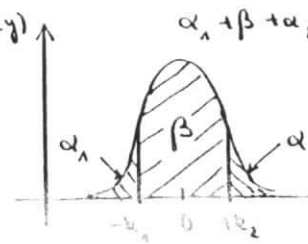
5.1 Normalverteilte Größen

Betrachtet wird eine Zufallsvariable x , die normal verteilt ist mit (unbekanntem) Mittelwert $\langle x \rangle = \mu$ bei bekannter Standardabweichung σ .

Gemessen ist ein \hat{x} (Schätzwert für $\langle x \rangle = \mu$), gesucht ist eine Aussage über $\langle x \rangle = \mu$

Transformation $y = \frac{x - \mu}{\sigma}$ ergibt Normalverteilung $N(0,1)$ mit Dichte

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$



$$\beta = \int_{-k_1}^{k_2} f(y) dy = P(-k_1 \leq y \leq +k_2) = \Phi(k_2) - \Phi(k_1)$$

$$\beta = P(\mu - k_1 \sigma \leq \hat{x} \leq \mu + k_2 \sigma)$$

Dies ist eine Aussage über ein Intervall, in das \hat{x} mit Wahrscheinlichkeit β fällt.

Gesucht ist jedoch eine Aussage über den unbekanntem Mittelwert μ . Dazu werden die Ungleichungen

$$\hat{x} \leq \mu + k_2 \sigma \quad \mu - k_1 \sigma \leq \hat{x}$$

umgewandelt in die Ungleichungen

$$\hat{x} - k_2 \sigma \leq \mu \quad \mu \leq \hat{x} + k_1 \sigma$$

man erhält die (inverse) Aussage

$$\beta = P(\hat{x} - k_2 \sigma \leq \mu \leq \hat{x} + k_1 \sigma)$$

d.h. mit Wahrscheinlichkeit β liegt der wahre Mittelwert $\langle x \rangle = \mu$ in den Grenzen

$$\hat{x} - k_2 \sigma \leq \langle x \rangle \leq \hat{x} + k_1 \sigma$$

Diese Inversion der Aussage ist im Falle der Normalverteilung einfach, weil die Dichte $f(y)$ in \hat{x} und μ symmetrisch ist.

(Inversion führt zur Vertauschung $k_1 \leftrightarrow k_2$)

Wahl von β :

im Prinzip kann β frei gewählt werden, je nach Fragestellung kann auch α_1 und α_2 (mit $\alpha_1 + \alpha_2 = 1 - \beta$) verschieden gewählt werden.

Zentrales Konfidenzintervall: $\alpha_1 = \alpha_2 = \frac{1 - \beta}{2}$

untere Grenzen: $\alpha_1 = 1 - \beta$; $\alpha_2 = 0$

obere Grenzen: $\alpha_2 = 1 - \beta$; $\alpha_1 = 0$

• Folgende Werte von β sind üblich:

① $\beta = 68.3\%$ entspricht gerade ± 1 Standardabw. bei der Normalverteilung

Wird ein Messergebnis in der Form

$$a = \hat{a} \pm sa$$

angegeben (ohne Angabe des Konfidenzniveaus β), sollte stets 68.3% , d.h. $\pm 1\sigma$ gemeint sein.

② $\beta = 95\%$ entspricht bei einer Normalverteilung ± 2 Standardabweichungen (genau $\pm 1.96\sigma$).

Angabe z.B.

$$2.7 \pm 2.9 \quad (95\% \text{ C.I.})$$

Bei Normalvet. ist i.a. die Angabe nur einer Standardabweichung sinnvoll, weil das andere Intervall leicht daraus berechnet werden können.

$\beta = 95\%$ entspricht bei dreiwertigen "geraden" $\pm 1.5\sigma$

Beispiele für Grenzen bei normalverteilten Größen

1. Beispiel: Varianz einer Normalverteilung

Bei einer Stichprobe vom Umfang n aus einer Normalverteilung (d.h. Daten x_1, x_2, \dots, x_n) hat die

Größe
$$u = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$$

eine χ^2 -Verteilung mit $(n-1)$ Freiheitsgraden, (σ unbekannt)

Ein Konfidenzintervall ^{für σ^2} kann durch zwei Werte u_1, u_2 gefunden werden, sodass

$$P(u_1 \leq u \leq u_2) = \int_{u_1}^{u_2} f(u) du = \beta \text{ ist.}$$

Invertieren der Aussage liefert mit

$$P\left(\frac{\sum(x_i - \bar{x})^2}{u_2} \leq \sigma^2 \leq \frac{\sum(x_i - \bar{x})^2}{u_1}\right) = \beta$$

ein Konfidenzintervall für σ^2 . Die Länge des Konfidenzintervalls ist

$$\left(\frac{1}{u_1} - \frac{1}{u_2}\right) \sum (x_i - \bar{x})^2$$

Da der Aufwand, das kleinste Intervall zu finden, sehr groß ist, wird i. a. ein zentrales Konfidenzintervall mit $u_1 = u_2 = \frac{1}{2}(1-\beta)$ benutzt.

Zahlenbeispiel: $S = \sum (x_i - \bar{x})^2 = 10$ bei $n=3$

$$\hat{\sigma}^2 = \frac{S}{n-1} = 5 \quad \hat{\sigma} = 2.24$$

Konfidenzintervall zu $\beta = 95\%$ $\rightarrow u_1 = u_2 = 2.5\%$

aus χ^2 -Tabelle: $u_1 = 0.0506$
 $u_2 = 7.38$

$$\frac{10}{7.38} \leq \sigma^2 \leq \frac{10}{0.0506}$$

$$1.35 < \sigma^2 < 197.6 \quad 95\% \text{ C.L.}$$

$$1.16 < \sigma < 14.5$$

2. Beispiel: Mittelwert einer Normalverteilung bei unbekannter Varianz, $\beta = 95\%$

Die Varianz des Mittelwerts \bar{x} ist $\frac{\sigma^2}{n}$.

$$P\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = \beta = 95$$

Einschön des Schätzwerts $\hat{\sigma}$ für σ kann insbesondere bei kleinen Stichproben zu falschen Aussagen führen, da die Verteilung von $\hat{\sigma}$ nicht berücksichtigt wird.

Exakte Berechnung: Benutzung der t -Verteilung.

$$t = \frac{\bar{x} - \mu}{\sqrt{\sum(x_i - \bar{x})^2 / (n(n-1))}} = \frac{\bar{x} - \mu}{\sqrt{\hat{\sigma}^2 / n}} \text{ hat } t\text{-Verteilung mit } n_d = n-1 \text{ Freiheits}$$

Bei $\beta = 95\%$: $n_d = 1 \quad 2 \quad \dots \quad 10$ (tabelliert in Büchern)

5.2 Der allgemeine Fall

Gegeben Schätzwert $\hat{a} = \hat{a}(x_1, x_2, \dots, x_n)$ für Parameter a ,
aus Daten x_1, x_2, \dots bestimmt.

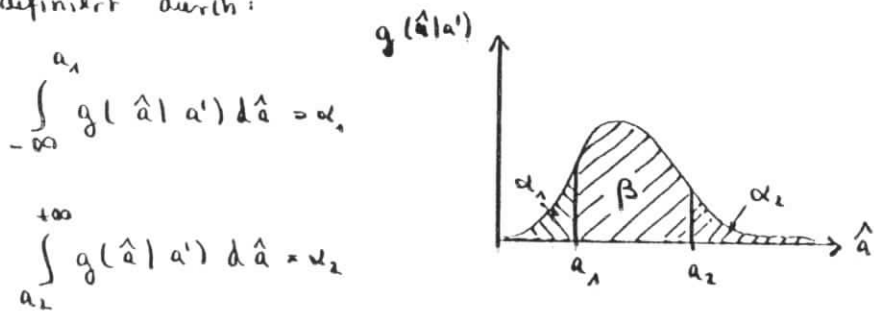
Gesucht: Konfidenzintervall zu vorgegebener Wahrscheinlichkeit β

Sei bei wahren Wert a die Dichtefunktion des Schätzwerts \hat{a}
 $g(\hat{a} | a)$

Zu jedem vorgegebenen Wert a' des Parameters lassen
sich zwei Grenzen a_1 und a_2 berechnen mit:

$$\beta = P(a_1 \leq \hat{a} \leq a_2) = \int_{a_1}^{a_2} g(\hat{a} | a') d\hat{a}$$

Diese Grenzen a_1 und a_2 sind mit $\alpha_1 + \beta + \alpha_2 = 1$
definiert durch:



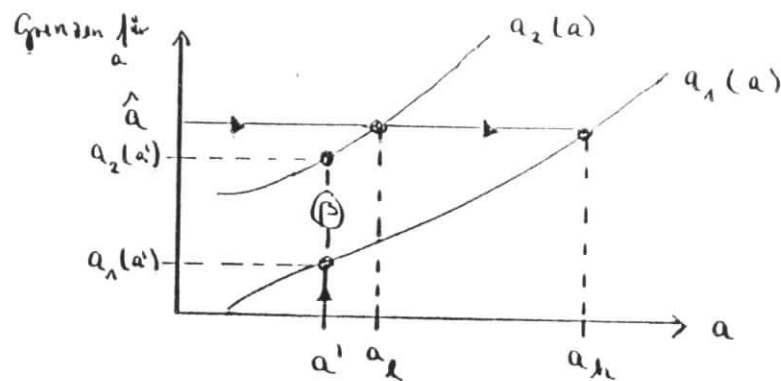
$$\int_{-\infty}^{a_1} g(\hat{a} | a') d\hat{a} = \alpha_1$$

$$\int_{a_2}^{+\infty} g(\hat{a} | a') d\hat{a} = \alpha_2$$

Die Grenzen a_1, a_2 hängen vom unbekanntem wahren
Wert von a ab, sie lassen sich aber (bei
bekannter Dichtefunktion) für jeden beliebigen Wert a'
berechnen

→ auftragen als Funktion von a

Konstruktion der Kurven: vertikal ↓
 a' vorgeben, Grenzen berechnen



Erläuterung: Da a' der wahre Wert, so liegt \hat{a} zwischen $a_1(a')$
und $a_2(a')$ in einem Anteil β von vielen, wiederholten Experimenten

Benutzung der Kurven: horizontal →

durch Messung bestimmt: Schätzwert \hat{a}
eine horizontale Linie durch \hat{a} schneidet die Kurven
in zwei Punkten, diese ergeben, auf die a -Achse
projiziert, die unteren und oberen Grenzen, mit
 $\beta = P(a_2 \leq a \leq a_1)$

Beweis: Der wahre Wert sei a' . Die Wahrscheinlichkeit,
daß ein Schätzwert \hat{a}' zwischen $a_1(a')$ und $a_2(a')$
fällt, ist β (nach Konstruktion). Für jeden Schätzwert \hat{a} ,
der in dieses Intervall fällt, ergibt die vertikale Grenzen,
die den wahren Wert enthalten. Also ist die Wahrscheinlichkeit
100% zu enthalten.

Die Grenzen a_L und a_H ergeben sich aber bei gegebenem Schätzwert \hat{a}' aus den Gleichungen:

$$\int_{-\infty}^{\hat{a}'} g(\hat{a} | a_H) d\hat{a} = \alpha_1$$

$$\alpha_1 + \alpha_2 + \beta = 1$$

$$\int_{\hat{a}'}^{+\infty} g(\hat{a} | a_L) d\hat{a} = \alpha_2$$

Beispiel: Eine physikalische Größe a wird gemessen, das Ergebnis ist

$$\hat{a} = 5 \pm 2$$

Von direktem physikalischen Interesse sei jedoch die reziproke Größe $b = \frac{1}{a}$. Gesucht sind 95% Grenzen für b .
Annahme: Messgröße a normalverteilt. Wichtig zu wissen ist, wie die Standardabw. von \hat{a} bei anderem Ergebnis für \hat{a} wäre.

Annahme (a) σ unabhängig von \hat{a} , z.B. 10 ± 2

Annahme (b) σ prop. zu \hat{a} , z.B. 10 ± 4

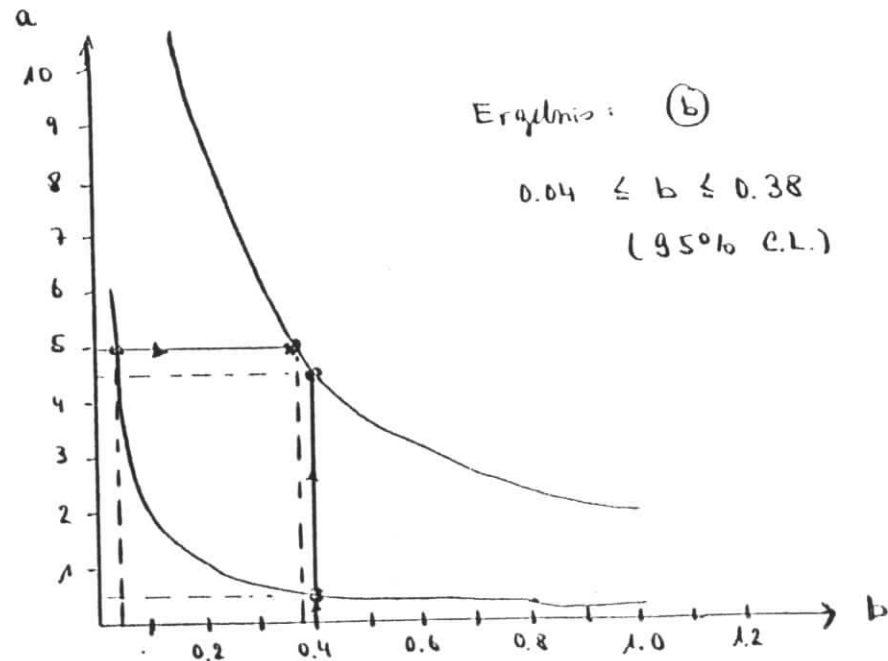
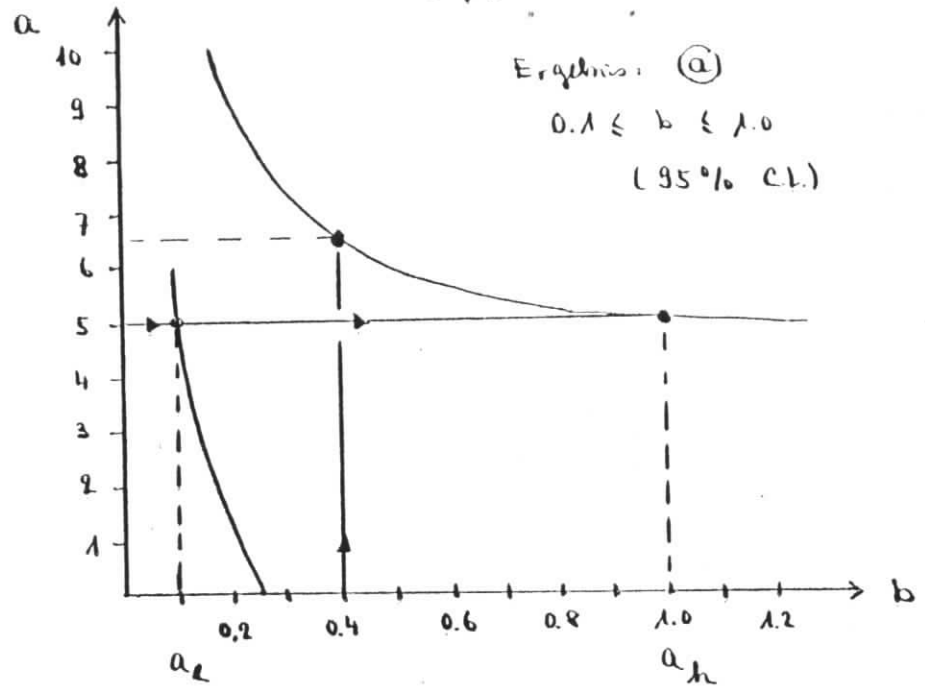
Für beide Fälle (a) und (b) werden Kurven entspr. den unten und oben Grenzen konstruiert.

z.B. $b = 0.4$ vorgeben $\Rightarrow a = 2.5$

(a) $\sigma = 2$ $a_1 = 2.5 - 2 \cdot 2$ $a_2 = 2.5 + 2 \cdot 2$

(b) $\sigma = 1$ $a_1 = 2.5 - 2 \cdot 1$ $a_2 = 2.5 + 2 \cdot 1$

Annahmen (a) und (b) führen auf sehr verschiedene Grenzen für b .



Konfidenzintervall bei Poisson-Verteilung:

Von einem bestimmten Typ von Ereignissen werden in einem Experiment n Ereignisse beobachtet. Gesucht sind Grenzen für den Parameter λ der Poisson-Verteilung zu vorgegebener Wahrscheinlichkeit (C.L.)

Dichtefunktion für n bei gegebenem Parameter λ der Poisson-Verteilung: $P(r=n | \lambda) = \frac{e^{-\lambda} \lambda^r}{r!}$

Da die Verteilung diskret ist, werden die Integrale durch Summen ersetzt, d.h. die unteren und oberen Grenzen λ_e und λ_h sind definiert durch:

$$\sum_{r=0}^n \frac{e^{-\lambda_h} \lambda_h^r}{r!} = \alpha_1$$

$$\sum_{r=n}^{\infty} \frac{e^{-\lambda_e} \lambda_e^r}{r!} = \alpha_2$$

Es gibt folgende (exakte) Beziehungen zwischen den Summen und der χ^2 -Verteilung:

$$\sum_{r=0}^n \frac{e^{-\lambda} \lambda^r}{r!} = 1 - P(u_{2n+2} \leq 2\lambda)$$

$$\sum_{r=n}^{\infty} \frac{e^{-\lambda} \lambda^r}{r!} = P(u_{2n} \leq 2\lambda)$$

Die Gleichungen für die Grenzen sind also erfüllt, wenn

$$1 - P(u_{2n+2} < 2\lambda_h) = \alpha_1$$

$$P(u_{2n+2} < 2\lambda_h) = 1 - \alpha_1$$

$$\Rightarrow \lambda_h = \frac{1}{2} \chi_{1-\alpha_1}^2(2n+2)$$

$$\text{analog } \lambda_e = \frac{1}{2} \chi_{\alpha_2}^2(2n)$$

Zahlenbeispiel:

$n = 9$ Ereignisse beobachtet, Grenzen für λ zu 95% C.L. gesucht.

Näherung durch Normalverteilung ergibt mit $\sigma = \sqrt{9} = 3$
 $3 \leq \lambda \leq 15$

Exakte Grenzen:

$$\beta = 95\% \quad \alpha_1 = \alpha_2 = 2.5\%$$

$$\chi_{97.5}^2(2n+2=20) = 34.2 \quad \lambda_h = 17.1$$

$$\chi_{2.5}^2(2n=18) = 8.2 \quad \lambda_e = 4.1$$

$$\Rightarrow 4.1 \leq \lambda \leq 17.1$$

Für die Interpretation von exper. Ergebnissen ist es manchmal sinnvoll, einseitige Grenzen anzugeben, d.h. obere bzw. untere Grenzen.

Untere Grenze:

$$\lambda_e = \frac{1}{2} \chi^2_{1-\beta} (2n)$$

Oberer Grenze:

$$\lambda_{h_1} = \frac{1}{2} \chi^2_{\beta} (2n+2)$$

Zahlenbeispiel: $n=0$ Ereignisse beobachtet, obere Grenze für λ zu $\beta = 95\%$ C.L. gesucht.

$$\lambda_{h_1} = \frac{1}{2} \chi^2_{95} (2) = \frac{1}{2} 5.99 = 3$$

d.h. $\lambda < 3$ Ereignisse (95% C.L.)

Aussage: Wenn λ tatsächlich > 3 ist, ist die Wahrscheinlichkeit, $n=0$ zu beobachten, kleiner als 5%

Inverse Aussage: Mit 95% Wahrscheinlichkeit ist $\lambda < 3$

5.3 Mehrdimensionale Zufallsvariable

Mehrdimensionale Zufallsvariable $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$

Praktisch bedeutsam vor allem mehrdimensionale Normalverteilung (übliche Schätzmethoden liefern asympt. normalverteilte Zufallsvariable). \rightarrow nur Not betrachte

Ist die Konstruktion mehrdimensionaler Konfidenzintervalle möglich?

Der Vektor der Mittelwerte $\langle \vec{a} \rangle = \vec{a}^0$ und die Kovarianzmatrix V bestimmen die Dichte der mehrdimensionalen Nvt. eindeutig

Quadratische Form: $Q(\vec{a}, \vec{a}^0) = (\vec{a} - \vec{a}^0)^T V^{-1} (\vec{a} - \vec{a}^0)$

Dichte:

$$f(\vec{a} | \vec{a}^0) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{1/2}} \exp\left(-\frac{1}{2} Q(\vec{a}, \vec{a}^0)\right)$$

Die Quadratische Form ist selbst Zufallsvariable, folgt der χ^2 -Verteilung mit n Freiheitsgrade

$Q(\vec{a}, \vec{a}^0) = \text{const.}$ entspricht Hyperebene in n -dim \vec{a} -Raum

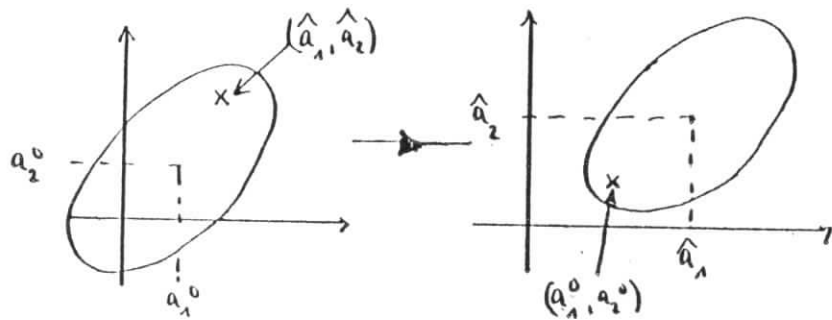
Wahrscheinlichkeitsaussage:

$$P(Q(\vec{a}, \vec{a}^0) \leq R_\beta) = \beta$$

mit $R_\beta = \chi^2_\beta(n)$ β -Wert der χ^2 -Vert.
mit n Freiheitsgraden:

Inversion der Wahrscheinlichkeitsaussage einfach
wegen Symmetrie zwischen \vec{a} und \vec{a}^0

Anschaulich für $n=2$:

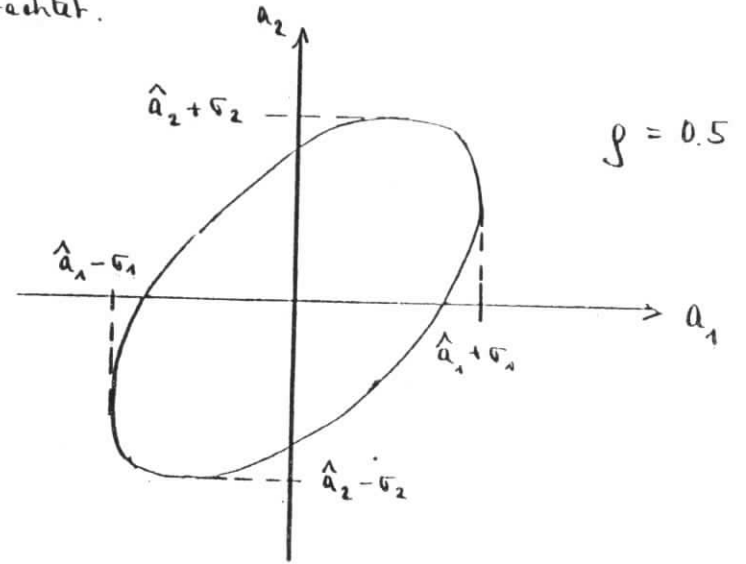


Formel ist die Konstruktion von Hyperellipsoiden
mit gegebenem Wahrscheinlichkeitsgehalt β einfach:

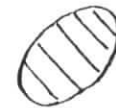
$$R_\beta = \chi^2_\beta(n)$$

$$Q(\vec{a}, \vec{a}^0) = R_\beta \quad \text{Hyperellipsoid}$$

Um die praktischen Probleme, die auftreten, zu
illustrieren, wird die Ellipse ($n=2$) zu
 $R_\beta = 1$ (χ^2 -Tafel: $\beta = 39.3\%$)
betrachtet.



Wahrscheinlichkeitsinhalt verschiedener Konfidenzbereiche



$$\beta_1 = 39.3\%$$

definiert durch
 $Q \leq 1$



$$\beta_2 = 49.8\%$$

(hängt von β ab)

$$\hat{a}_1 \pm \sigma_1$$

$$\hat{a}_2 \pm \sigma_2$$



$$\beta_3 = 68.3\%$$

$$\hat{a}_1 \pm \sigma_1$$

Folgerungen:

- ① Nur in dem Falle, daß nur 1 Parameter (von vielen) betrachtet wird, hat das Intervall ± 1 Standardabweichung den üblichen Wahrscheinlichkeitsinhalt ($\beta_3 = 68.3\%$)
- ② Wenn gleichzeitig zwei (oder mehr) Parameter betrachtet werden, z.B.

$$\hat{a}_1 \pm \sigma_1$$

$$\hat{a}_2 \pm \sigma_2$$
 so entspricht dem dadurch definierten (rechteckigen) Konfidenzbereich eine kleinere Wahrscheinlichkeit, die übrigens noch von der Korrelation abhängt.
- ③ Weitergehende Rechnungen sind nur mit voller Kovarianzmatrix möglich (z.B. Transformationen, Test von Hypothesen).

6. Test von Hypothesen

6.1 Einführung

Beispiel für Test von statistischen Hypothesen:

Gemessen sind zwei Werte

$$\left. \begin{aligned} x_1 \pm \sigma_1 &= 2.05 \pm 0.01 \\ x_2 \pm \sigma_2 &= 2.09 \pm 0.02 \end{aligned} \right\} \text{normal-verteilt}$$

Zu prüfen ist die statistische

Hypothese H: Die beiden Werte sind miteinander verträglich.

Ergebnis des Test ist

- entweder: Annahme der Hypothese
- oder: Zurückweisung der Hypothese

Zu definieren sind:

- eine Testgröße t
- ein Signifikanzniveau α

1. Möglichkeit:

Differenz bilden: $\Delta x \pm \sigma = (x_2 \pm \sigma_2) - (x_1 \pm \sigma_1)$
 $= 0.04 \pm 0.0224$

Hypothese H besagt:

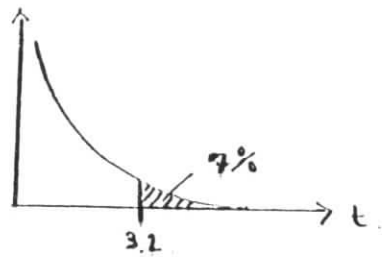
$\frac{\Delta x}{\sigma}$ ist normalverteilt $N(0, 1)$

$t = \left(\frac{\Delta x}{\sigma}\right)^2$ ist χ^2 verteilt mit 1 Fr.

Wenn H wahr ist, ist die Dichte von t etwa

Zahlenwert: $t = 3.2$

$P(t > 3.2) = 7\%$



Vorgehen in der Praxis:

Ein bestimmtes Signifikanzniveau α wird vorgegeben, dies definiert durch

$P(t > t_c) = \alpha$

eine Grenze t_c für die Testgröße t

übliche Wahl: $\alpha = 5\%$

$\Rightarrow t_c = 3.84$ (aus χ^2 Vert. mit 1 Frei.)

Für $t \leq t_c$ wird die Hypothese angenommen,

--- $t > t_c$ wird die Hypothese verworfen

Was bedeutet dies:

Sei H tatsächlich wahr. Dann wird die Hypothese mit Wahrscheinlichkeit α (fälschlicherweise) verworfen.
= Fehler 1. Art

2. Möglichkeit (nur formal anders):

Testgröße $t = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{\sigma_i^2}$

Wenn die Hypothese richtig ist, und \bar{x} bekannt ist, dann ist $t \sim \chi^2$ verteilt mit n Freiheitsgraden.

Hier ist aber \bar{x} nicht bekannt, muß aus den Daten bestimmt werden.

\bar{x} so wählen, daß t möglichst klein:

$\Rightarrow \bar{x}$ bestimmt durch

$\frac{dt}{d\bar{x}} = -2 \sum_{i=1}^n \frac{(x_i - \bar{x})}{\sigma_i^2} = 0$

$\Rightarrow \sum_{i=1}^n \frac{x_i}{\sigma_i^2} = \bar{x} \sum_{i=1}^n \frac{1}{\sigma_i^2}$

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

mit $w_i = \frac{1}{\sigma_i^2}$
"Gewicht"

Da 1 Parameter aus den Daten bestimmt wurde, ist nun $t \sim \chi^2$ verteilt mit $(n-1)$ Freiheitsgraden.

Zahlenwerte

$\bar{x} = 2.056$

$t = 3.20$ wie vorher

• Allgemein: zu vorgegebenem α kann für die χ^2 Verteilung der Abschneidewert der Testgröße t aus Tabellen abgelesen werden

Beispiele zu $\alpha = 5\%$

$t_c = 3.84$ bei 1 Freiheitsgrad

$t_c = 18.3$ bei 10 " " "

Allgemeine Definition eines statistischen Tests:

Zu testen ist eine Hypothese H_0 .

Statt der Hypothese H_0 können auch andere Hypothesen H_1, H_2, \dots zutreffen.

- Null-Hypothese H_0
- Alternativ-Hypothesen H_1, H_2, \dots

Ein Signifikanzniveau α wird vorgegeben, und eine Testgröße t wird definiert. Zu gegebenem α läßt sich ein kritischer Bereich W für t definieren durch

$$P(t \in W | H_0) = \alpha$$

- Signifikanzniveau α
- Testgröße t = Fkt. der Beobachtungen
- Kritischer Bereich W

Ist $t \in W$, wird H_0 verworfen. Wenn H_0 wahr ist, wird H_0 mit Wahrscheinlichkeit α verworfen.

$$= \text{Fehler 1. Art}$$

Der Grad der Nützlichkeit eines Tests hängt von der Fähigkeit ab, gegebene Alternativhypothesen zu diskriminieren.

$$P(t \notin W | H_1) = \beta$$

$$P(t \in W | H_1) = 1 - \beta$$

Wenn H_1 wahr ist, wird die Hypothese H_0 mit Wahrscheinlichkeit β akzeptiert.

$$= \text{Fehler 2. Art}$$

Der kritische Bereich W sollte so gewählt werden bei vorgegebenem α , daß die Wahrscheinlichkeit β möglichst klein wird.
(optimaler Test)

Begriffe:

einfache Hypothese: Wahrscheinlichkeit vollst. bestimmt
(kein freier Parameter)

- Parameter-Tests: Test des Zahlenwerts eines oder mehrerer Parameter.
- Anpassungstests: Test von Verteilungsfunktionen
(Alternative zu H_0 , alle andere Hypothesen)

6.2 Parametertest

Als Beispiel wird der Test einer einfachen Hypothese gegen eine einfache Alternative betrachtet.

$$\left. \begin{aligned} H_0 : a = a_0 = 2 \\ H_1 : a = a_1 = 0 \end{aligned} \right\} \begin{array}{l} \text{Normalverteilung,} \\ \text{mit Varianz 1} \end{array}$$

für Stichprobe $x_1, x_2, x_3, \dots, x_n$

d.h. zu entscheiden ist zwischen der Hypothese, daß die Werte x_i aus einer NvH mit Mittelwert 0, Varianz 1 stammen gegen die Hypothese, daß die Werte x_i aus einer NvH mit Mittelwert 2, Varianz 1 stammen.

Testgröße:

$$t = \frac{\prod_{i=1}^n f(x_i; a_0)}{\prod_{i=1}^n f(x_i; a_1)} = \frac{L(a_0)}{L(a_1)} \quad \text{Likelihood-Verhältnis}$$

$t > k \quad \rightarrow H_0$ akzeptieren

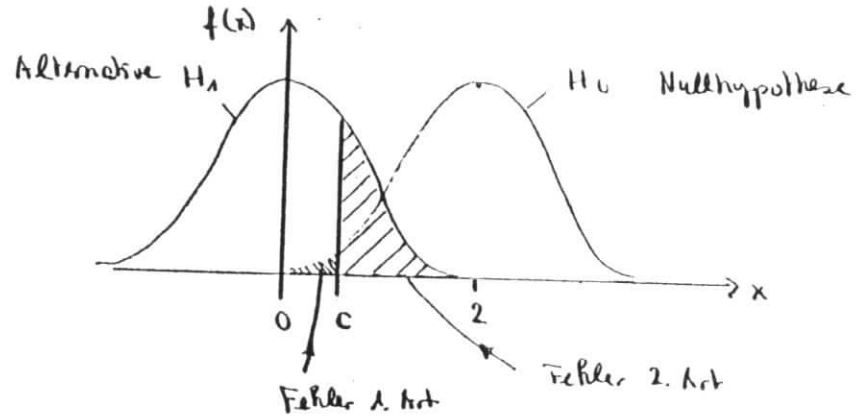
$t < k \quad \rightarrow H_1$ akzeptieren.

Einssetzen von $f(x, a) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-a)^2\right]$

Liefert: $t = (a_0 - a_1) n \bar{x} - \frac{1}{2}(a_0^2 - a_1^2) n$
 $= 2n\bar{x} - 2n$

Die Grenze $t = k$ entspricht

$$\bar{x} = c = \frac{\ln k}{2} + 1$$



Fehler 1. Art: $P(\bar{x} < c | a=2) = \alpha$

Zahlenwerte: $\alpha = 5\%$, $n=4 \quad \rightarrow \quad c = 1.1775$

Fehler 2. Art: $P(\bar{x} > 1.1775 | a=0) = 0.9\%$

Wie α gewählt wird, und welche Werte von β zulässig sind, hängt von der Art des speziellen Problems ab.

Beispiel: Das Schauerzählersignal, daß durch den Durchgang eines Teilchens durch einen Schauerzähler hervorgerufen wird, soll zur Unterscheidung zwischen Elektronen und Hadronen dienen:

Elektronen: Signal groß
 Hadronen: Signal klein

Zur Bestimmung der Grenze Fehler 1. Art + 2. Art, und Häufigkeit von Hadronen, Elektronen bestimmen.

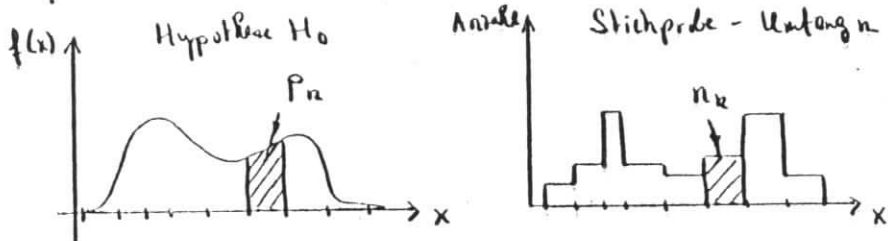
6.3 Anpassungstest

Anpassungstest \equiv goodness-of-fit test

hier χ^2 -Test

Verfahri der Verteilung einer Groe x mit der Annahme eines bestimmten Hypothese

$f(x)$ = Dichte zur Hypothese H_0



Einteilung in Intervalle

$\rightarrow p_1, p_2, \dots, p_r$

in Intervall k :

$n \cdot p_k$ erwartet

n_k beobachtet

$$\text{Testgroe } \chi^2 = \sum \frac{(n_k - n \cdot p_k)^2}{n \cdot p_k}$$

χ^2 = Zufallsvariable, die einer χ^2 -Verteilung mit $(r-1)$ Freiheitsgraden folgt.

(bei groen n)

(1 Parameter, namlich n , ist den Daten entnommen)

Die Anzahlen n_k konnen eine Poisson-Verteilung (bei festem n Multinomialverteilung), diese kann durch Normalverteilung mit Mittelwert $n \cdot p_k$

Varianz $n \cdot p_k$

angenohert werden. $\rightarrow \chi^2$ -Verteilung, tabelliert.

in der Praxis: $n \cdot p_k > 5$, sonst Intervalle zusammenfassen.

Beispiel: Ein Tahler misst jeweils fur 1 Minute eine Fahrrate $n_k, k=1, \dots, r$

$$H_0: \text{Fahrrate konstant}$$

$$n = \sum_{k=1}^r n_k$$

$$\rightarrow p_k = \frac{1}{r} \quad ; \quad n \cdot p_k = \frac{1}{r} n$$

$$\text{Testgroe } \chi^2 = \sum_{k=1}^r \frac{(n_k - n \cdot \frac{1}{r})^2}{n \cdot \frac{1}{r}} = \frac{r}{n} \sum_{k=1}^r (n_k - n \cdot \frac{1}{r})^2$$

Fastbedeutetes Signifikanzniveau: $\alpha = 5\%$

$$\rightarrow n_k = 100 \quad 90 \quad 80 \quad 130 \quad r = 4$$

$$n = 400 \quad \rightarrow \chi^2 = 14.0$$

Aus χ^2 -Tabelle bei $r-1 = 3$ Freiheitsgraden: $\chi^2_c = 7.81$

$\chi^2 > \chi^2_c$ \leftarrow Hypothese verwerfen

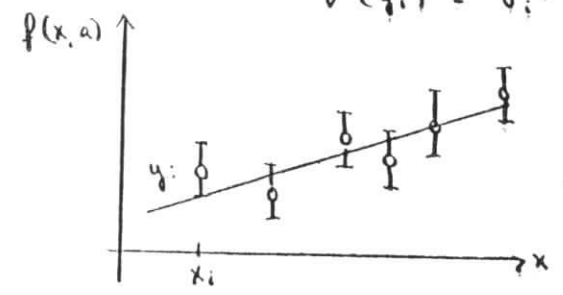
$$14.0 > 7.81$$

$\bullet \rightarrow$ die Annahme, da die Fahrrate konstant ist, ist (vermutlich) falsch.

χ^2 - Test bei Methode der kleinsten Quadrate:
vgl. 4.4; eine Größe Q wird minimiert.

$$Q = \sum_{i=1}^n \frac{(y_i - f(x_i, a))^2}{\sigma_i^2} = \sum_{i=1}^n \left(\frac{r_i}{\sigma_i} \right)^2$$

Hypothese H_0 : $E(y_i) = f(x_i, a)$
 $V(y_i) = \sigma_i^2$, normal. } Modell



$\Rightarrow \frac{r_i}{\sigma_i}$ normalverteilt $N(0, 1)$

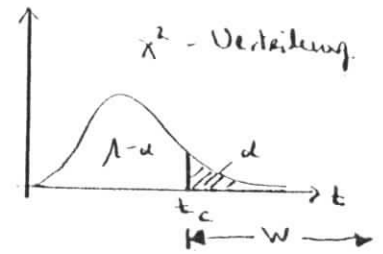
d.h. Q Zufallsvariable mit χ^2 -Verteilung, mit n Freiheitsgraden, wenn kein Parameter aus den Daten bestimmt wird;

$(n-k)$ Freiheitsgrade, wenn k Parameter a_1, a_2, \dots, a_k aus den Daten bestimmt werden (und $f(x, a)$ linear von den Parametern a_1, \dots, a_k abhängt)

Testgröße $t = Q$

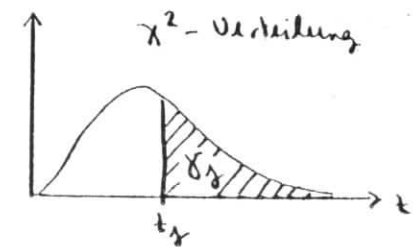
Signifikanzniveau α festlegen, und dazu den $(1-\alpha)$ Wert der χ^2_{n-k} -Verteilung aus Tabelle ablesen: $\Rightarrow t_c$

Kritischer Bereich W :
 $t > t_c$



Andere Möglichkeit für Hypothesenprüfung, bei sehr vielen gleichartigen Fällen:

jedes t_j umrechnen in Wahrscheinlichkeit γ_j

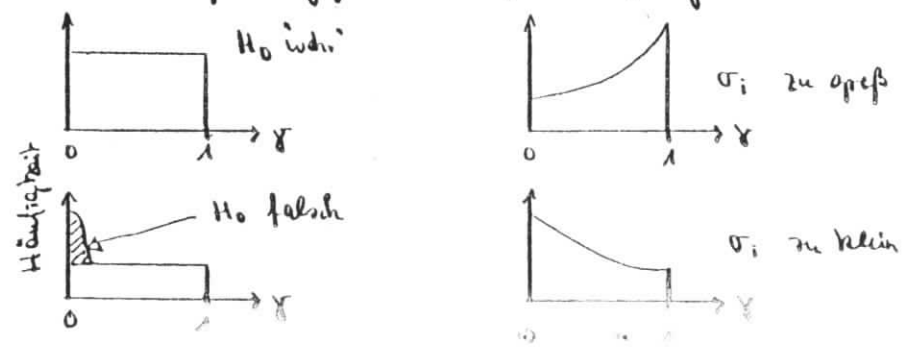


$$\gamma_j = \text{PROB}(t_j, n-k)$$

$$\gamma_j = \text{PROCH}(t_j, n-k)$$

\uparrow Zahl der Freiheitsgrade

Verteilung von γ_j als Histogramm auftragen:



B. Anwendung der Maximum-Likelihood-Methode (MLM)

1. Problemstellung

Die MLM ist eine allgemeine Methode zur Bestimmung (Schätzung) von Parametern aus Daten.

In diesem Abschnitt werden vier verschiedene Probleme behandelt, und der Ansatz für die MLM beschrieben.

→ (a) Gegeben n Messwerte x_1, x_2, \dots, x_n

Annahme: die x_i folgen einer Verteilung mit der Dichte $f(x_i, a)$

mit unbekanntem Parameterwert a_1, a_2, \dots, a_p .

Voraussetzung: die Dichte $f(x_i, a)$ sei normiert.

Gemeinsame Dichtefunktion der x_i

$$L(x_i | a) = \prod_{i=1}^n f(x_i, a)$$

Die Funktion $L(x_i | a)$, als Funktion von a (bei gegebenen Messwerten x_i) betrachtet, heißt Likelihood-Funktion.

Prinzip der MLM: Der Schätzwert \hat{a} ist gegeben durch den Wert der Parameter a , bei dem $L(x | a)$ maximal wird.

Praktische Durchführung:

Logarithmus von $L(x | a)$ betrachten

$$F(a) = \ln L(x | a)$$

$$F(a) = \sum_{i=1}^n \ln f(x_i, a)$$

Maximum von $F(a)$ bestimmen

Beispiel: Ein Winkelverteilung wird gemessen, $x_i = \cos$
Erwartet wird eine Verteilung der Form

$$1 + a \cos^2 \theta = 1 + a x^2$$

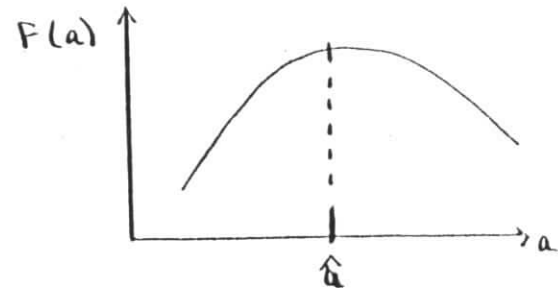
Diese Verteilung ist nicht normiert.

$$\int_{-1}^{+1} (1 + a x^2) dx = 2 \left(1 + \frac{a}{3} \right)$$

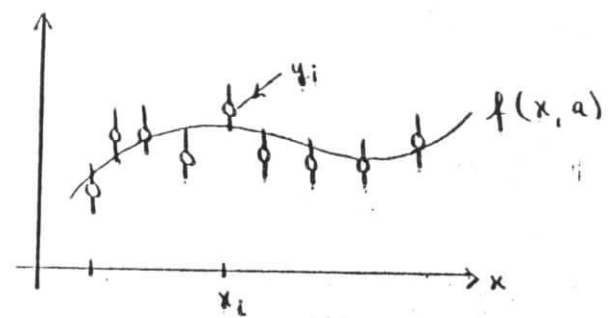
Durch Division durch das Integral erhält man eine normierte Verteilung:

$$f(x, a) = \frac{1 + a x^2}{2 \left(1 + \frac{a}{3} \right)}$$

Für gegebene Werte x_1, x_2, \dots, x_n $F(a)$ berechnen für verschiedene Werte von a .



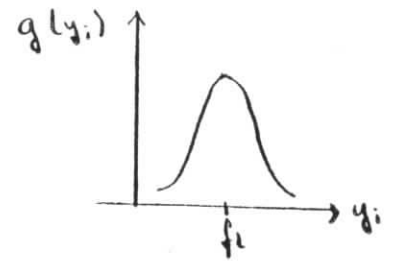
→ (b) Gegeben n Meßwerte y_1, y_2, \dots, y_n als Funktion einer unabh. Variablen x .



Annahme: die Abhängigkeit der Werte $y_i = y_i(x)$ kann durch eine Funktion $f(x_i, a)$ mit unbekanntem Parameterwert a beschrieben werden.

d.h. $E(y_i) = f(x_i, a) = f_i$

Voraussetzung: y_i normalverteilt um f_i
 $V(y_i) = \sigma_i^2$ (bekannt)



Für ein einzelnes y_i gilt:
 $g(y_i) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{(y_i - f_i)^2}{2\sigma_i^2}\right)$
 $\int_{-\infty}^{+\infty} g(y_i) dy_i = 1$ normiert

Gemeinsame Dichtefunktion:
 $L(y | a) = \prod_{i=1}^n g(y_i, a)$

$$F(a) = \ln L(y | a) = -\frac{1}{2} \sum \frac{(y_i - f_i)^2}{\sigma_i^2} - \sum \ln \sigma_i - \frac{1}{2} n \ln(2\pi)$$

Die letzten beiden Terme hängen nicht von a ab, sie können weggelassen werden:

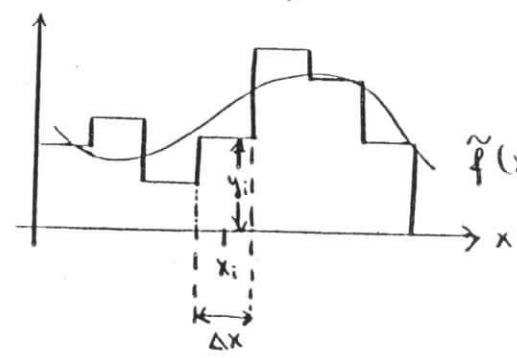
$$F(a) = -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - f_i)^2}{\sigma_i^2}$$

Vergleich mit der Methode der kleinsten Quadrate: dort würde man das Minimum der Funktion

$$Q(a) = \sum_{i=1}^n \frac{(y_i - f_i)^2}{\sigma_i^2}$$

bestimmen, das Ergebnis ist das gleiche: die Methode der kleinsten Quadrate setzt jedoch nicht die Normalverteilung voraus, sie hat optimale Eigenschaften auch bei anderen Verteilungen.

→ (c) Gegeben als Meßwerte n Brechen y_1, y_2, \dots, y_n , z.B. Einträge in einem Histogramm.

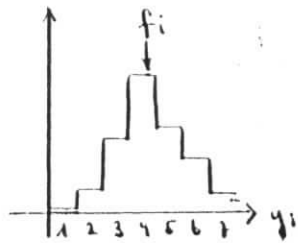


Annahme: y_i durch Fkt. $\tilde{f}(x, a)$ beschrieben.

Erwartungswert für y_i im Intervall $x_i \pm \frac{1}{2} \Delta x$.

$$= f(x_i, a) = \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} \tilde{f}(x, a) dx \approx \tilde{f}(x_i, a) \cdot \Delta x$$

Die y_i sind Poisson verteilt, mit Dichte



$$g(y_i) = \frac{f_i^{y_i}}{y_i!} \exp(-f_i)$$

Gemeinsame Dichtefunktion:

$$L(y|a) = \prod_{i=1}^n g(y_i)$$

$$F(a) = \sum_{i=1}^n y_i \ln f_i - \underbrace{\sum_{i=1}^n \ln y_i!}_{\text{const.}} - \sum_{i=1}^n f_i$$

Der konstante Term kann weglassen werden:

$$F(a) = \sum_{i=1}^n y_i \ln f(x_i, a) - \sum_{i=1}^n f(x_i, a)$$

Fall (c) tritt bei der Auswertung von Experimenten der Hochenergiephysik häufig auf.

In der Praxis wird meist die Methode der kleinsten Quadrate benutzt, mit der Näherung:

$$\sigma_i = \sqrt{f_i} \quad (\text{oder auch } \sigma_i = \sqrt{y_i})$$

→ Wenn alle $y_i \gg 1$ ist, ist dies eine gute Näherung, die Resultate von Methoden (b) und (c) sind dann praktisch gleich.

Im Prinzip ist $\sigma_i = \sqrt{f_i}$ konst., ist jedoch bei der Maximalisierung instabiler (die σ_i hängen dann von den Werten der Parameter ab).

Die Näherung $\sigma_i = \sqrt{y_i}$ ist stabiler.

Gute Methode: erst mit $\sigma_i = \sqrt{y_i}$ rechnen, nach Konvergenz Rechnung mit $\sigma_i = \sqrt{f_i}$ wiederholen.

→ Wenn einige y_i klein (oder gar 0) sind, ist die Näherung durch Methode der kleinsten Quadrate nicht anwendbar.

→ (d) Extended Maximum-Likelihood-Methode

Gegeben n Messwerte x_1, \dots, x_n

Annahme: die x_i folgen einer Verteilung mit der "Dichte"
 $f(x_i, a)$

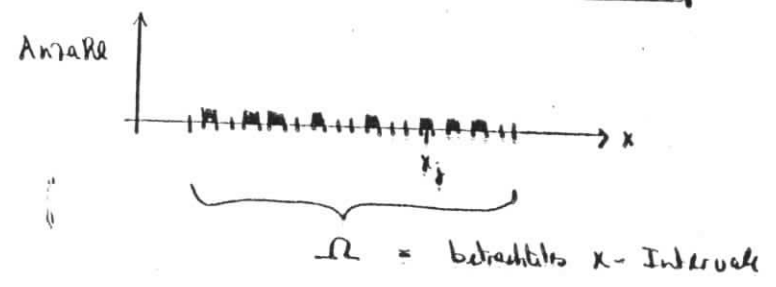
mit unbekanntem Parameterwerten a_1, \dots, a_p

soweit wie Fall (a), jedoch,

keine Voraussetzung, dass $f(x_i, a)$ normiert ist,

z.B. $f(x, a) = 1 + a x^2 \quad -1 \leq x \leq 1$

Einteilen des x -Bereichs in n kleine Intervalle Δx ,
dass in jedem Intervall höchstens 1 Eintrag ist.



Erwartungswert in einem x -Intervall

$$= f(x_j, a) \cdot \Delta x = f_j \cdot \Delta x$$

Insgesamt n Intervalle $x_j, j=1, \dots, k$

Bei zugrundeliegender Poisson-Verteilung in jedem Intervall
(vgl. (a)) hat man nur $y_j = 0$ oder 1 zu betrachten

$$g(0) = \frac{(f_j \cdot \Delta x)^0}{0!} \exp(-f_j \cdot \Delta x) = \exp(-f_j \cdot \Delta x)$$

$$g(1) = \frac{(f_j \cdot \Delta x)^1}{1!} \exp(-f_j \cdot \Delta x)$$

Durch Einsetzen in

$$L(y|a) = \prod_{j=1}^k g(y_j)$$

ergibt sich

$$F(a) = \ln L(y|a) = \sum_{i=1}^n \ln f_i + \underbrace{k \ln \Delta x}_{\text{const.}} - \sum_{j=1}^k f_j \cdot \Delta x$$

Durch Weglassen des konstanten Terms und Grenzübergang
 $\Delta x \rightarrow 0$ erhält man

$$F(a) = \sum_{i=1}^n \ln f_i - \int_{\Omega} f(x, a) dx$$

Der Vergleich mit (a) zeigt, dass der Integralterm
hinzugekommen ist.

Was ist $\int_{\Omega} f(x, a) dx$?

$\int_{\Omega} f(x, a) dx =$ Erwartungswert von n
(Anzahl Ereignisse)

Extended MLH geht von unnormierten Funktionen $f(x, a)$
aus, und ergibt zusätzlich zu besten Parameterwerten
noch einen Wert für die Zahl der erwarteten Ereignisse,
der "observer" sein sollte als der beobachtete Wert n .

Beispiel: Winkelverteilung $(1 + a_1 x^2) a_2 \quad x = \cos \theta$
 n ist proportional zum Wirkungsquerschnitt

Extended MLH gibt Parameter der Winkelverteilung und
besten Wert für den Wirkungsquerschnitt.

2. Eigenschaften der Maximum-Likelihood-Methode

Gemeinsame Dichtefunktion

$$L(y|a) = \prod_{i=1}^n g(y_i, a)$$

Likelihood-Funktion = Fkt von a

Folgende Formeln für einen Parameter a geschrieben,
am Ende des Abschnitts Formeln mit Verallgemeinerung
auf p Parameter a_1, a_2, \dots, a_p (Vektor \vec{a})

Logarithmus von $L(y|a)$:

$$F(a) = \ln L(y|a) = \sum_{i=1}^n \ln g(y_i|a)$$

Notwendige Bedingung für Maximum

$$\frac{\partial}{\partial a} F(a) = 0 \quad \text{Likelihood-Gleichung}$$

Schätzwert \hat{a} der MLM durch Lösen dieser Gleichung

Der Schätzwert \hat{a} hat folgende Eigenschaften

a.) Konsistenz: Der Schätzwert \hat{a} der MLM ist
asymptotisch ($n \rightarrow \infty$) konsistent, d.h.
$$\hat{a} \xrightarrow[n \rightarrow \infty]{} a^0 \quad (a^0 = \text{wahrer Wert des Parameters})$$

b.) Normalität: Unter sehr allgemeinen Voraus.
ist \hat{a} asymptotisch normalverteilt mit minimalem
Varianz $V(\hat{a})$:

$$V(\hat{a}) \xrightarrow[n \rightarrow \infty]{} \left\{ E \left(\frac{\partial \ln g}{\partial a} \right)^2 \right\}^{-1}$$

↳ Likelihood-Fkt. asymptotisch eine lineare Funktion
von a.

Wichtige Eigenschaft: Invarianz

d.h. Schätzwert \hat{b} einer Funktion $b = b(a)$
gegeben durch $\hat{b} = b(\hat{a})$

Jedoch, für endliche n hat nur eine Funktion b(
keine system. Verschiebung (Bias),
d.h. i.a. ist ein Bias ($\propto \frac{1}{n}$) vorhanden.

Diese Aussagen werden im folgenden begründet:

a.) Konsistenz:

Schätzwert \hat{a} ergibt sich durch Lösen der Gleichung

$$\frac{\partial F}{\partial a} = \sum_{i=1}^n \frac{\partial \ln g(y_i, a)}{\partial a} = 0$$

bzw. durch Lösen der Gleichung

$$Q(a) = 0 \quad \text{mit} \quad Q(a) = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln g(y_i, a)}{\partial a}$$

hängt von exp. Werten y_i ab.

Gesetz der großen Zahl:

Für jede Funktion $z(y)$ einer Zufallsvariablen y mit Dichtefunktion $h(y)$ gilt:

$$\frac{1}{n} \sum_{i=1}^n z(y_i) \xrightarrow{n \rightarrow \infty} E[z(y)] = \int z(y) h(y) dy$$

Daher gilt:

$$Q(a^0) = \frac{1}{n} \sum \frac{\partial \ln g(y_i, a^0)}{\partial a} \xrightarrow{n \rightarrow \infty} E \left[\frac{\partial \ln g}{\partial a} \right] = 0$$

Der Erwartungswert von $\frac{\partial \ln g}{\partial a} = \frac{1}{g} \frac{\partial g}{\partial a}$ verschwindet:

$$E \left[\frac{\partial \ln g}{\partial a} \right] = \int_{\Omega} \frac{1}{g} \frac{\partial g}{\partial a} g dy = \frac{\partial}{\partial a} \int_{\Omega} g dy = 0$$

$\underbrace{\int_{\Omega} g dy}_{=1}$

Daher gilt: $Q(a^0) \xrightarrow{n \rightarrow \infty} 0$

Da \hat{a} durch Lösen von $Q(a) = 0$ bestimmt wird, ist

$$\hat{a} \xrightarrow{n \rightarrow \infty} a^0 \quad (\text{Konsistenz})$$

b.) Asymptotische Normalität:

Entwicklung der Funktion $Q(a)$ um a^0 :

$$Q(a) = Q(a^0) + (a - a^0) \left(\frac{\partial Q}{\partial a} \right)_{a=a^0} + \dots$$

\Rightarrow Asymptotisch gilt (bei Abbruch der Entwicklung nach dem linearen Term):

$$\hat{a} - a^0 = \frac{Q(a^0)}{\left(\frac{\partial Q}{\partial a} \right)_{a=a^0}} \quad (1)$$

In diesem Ausdruck nähert sich der Nenner der rechten Seite einer (negativen) Konstanten. Die Varianz um \hat{a} ergibt sich daher aus der Varianz von $Q(a^0)$.

Betrachtung des Nenners:

$$E \left[\left(\frac{\partial Q}{\partial a} \right)_{a=a^0} \right] = \text{asymptotische Steigung um } a$$

$$\frac{\partial Q}{\partial a} = \frac{1}{n} \frac{\partial}{\partial a} \sum_{i=1}^n \frac{\partial \ln g(y_i, a)}{\partial a} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln g(y_i, a)}{\partial a^2}$$

Anwendung des Gesetzes der großen Zahl:

$$\left(\frac{\partial Q}{\partial a} \right)_{a=a^0} \xrightarrow{n \rightarrow \infty} E \left[\frac{\partial^2 \ln g}{\partial a^2} \right]$$

Ausrechnung:

$$\begin{aligned} \frac{\partial^2 \ln g}{\partial a^2} &= \frac{\partial}{\partial a} \left(\frac{\partial \ln g}{\partial a} \right) = \frac{\partial}{\partial a} \left(\frac{1}{g} \frac{\partial g}{\partial a} \right) \\ &= \frac{1}{g} \frac{\partial^2 g}{\partial a^2} - \left(\frac{\partial \ln g}{\partial a} \right)^2 \end{aligned}$$

$$E \left[\frac{\partial^2 \ln g}{\partial a^2} \right] = \underbrace{E \left[\frac{1}{g} \frac{\partial^2 g}{\partial a^2} \right]}_{=0} - \underbrace{E \left[\left(\frac{\partial \ln g}{\partial a} \right)^2 \right]}_{\neq 0}$$

1. Term der rechten Seite:

$$E \left[\frac{1}{g} \frac{\partial^2 g}{\partial a^2} \right] = \int_{-n}^n \frac{1}{g} \frac{\partial^2 g}{\partial a^2} g dy = \frac{\partial^2}{\partial a^2} \underbrace{\int_{-n}^n g dy}_{=1} = 0$$

$$\begin{aligned} \Rightarrow E \left[\left(\frac{\partial Q}{\partial a} \right)_{a=a^0} \right] &= E \left[\frac{\partial^2 \ln g}{\partial a^2} \right] = - E \left[\left(\frac{\partial \ln g}{\partial a} \right)^2 \right] \quad (2) \\ &= \text{negative Konstante} \end{aligned}$$

Betrachtung des Zählers:

$Q(a^0)$ asymptotischer Erwartungswert = 0

Varianz von $Q(a^0)$ aus zentralem Grenzwertsatz.

Der zentrale Grenzwertsatz:

Für jede Funktion $z(y)$ einer Zufallsvariablen y mit Dichte $h(y)$ ist die Größe

$$\frac{1}{n} \sum_{i=1}^n z(y_i)$$

asymptotisch normalverteilt mit Mittelwert

$$E(z(y)) = \int z(y) h(y) dy = \langle z \rangle$$

und mit Varianz

$$\frac{1}{n} V(z(y)) = \frac{1}{n} \int (z(y) - \langle z \rangle)^2 h(y) dy$$

Daher gilt für die Varianz von $Q(a^0)$:

$$V(Q(a^0)) \xrightarrow{n \rightarrow \infty} \frac{1}{n} V \left[\frac{\partial \ln g}{\partial a} \right] = \frac{1}{n} E \left[\left(\frac{\partial \ln g}{\partial a} \right)^2 \right]$$

Die Varianz von \hat{a} ergibt sich nun gemäß Gl. (1) durch Transformation:

$$V(\hat{a}) = \left\{ E \left[\left(\frac{\partial Q}{\partial a} \right)_{a=a^0} \right] \right\}^{-1} \cdot V[Q(a^0)] \cdot \left\{ E \left[\left(\frac{\partial Q}{\partial a} \right)_{a=a^0} \right] \right\}$$

Durch Einsetzen von Gl. (2) und (3):

$$\begin{aligned} V(\hat{a}) &= \frac{1}{n} \left\{ E \left[\left(\frac{\partial \ln g}{\partial a} \right)^2 \right] \right\}^{-1} \\ &= \left\{ E \left[n \left(\frac{\partial \ln g}{\partial a} \right)^2 \right] \right\}^{-1} = - \left\{ E \left[n \frac{\partial^2 \ln g}{\partial a^2} \right] \right\}^{-1} \end{aligned}$$

Die eingeraumte Formel gibt die asymptotische Varianz der MLM-Schätzwerte \hat{a} , die Normalität folgt aus dem zentralen Grenzwertsatz.

→ $\left\{ \begin{array}{l} \text{Berechnung der} \\ \text{Varianz des Schätzwerts} \\ \text{vor Ausführung des Experiments möglich} \end{array} \right.$

Verallgemeinerung auf p Parameter a_1, a_2, \dots, a_p :

→ Kovarianzmatrix $V(\hat{a})$

Asymptotisch

$$\begin{aligned} [V(\hat{a})^{-1}]_{jk} &= E \left[n \frac{\partial \ln g}{\partial a_j} \frac{\partial \ln g}{\partial a_k} \right] \\ &= -E \left[n \frac{\partial^2 \ln g}{\partial a_j \partial a_k} \right] \end{aligned}$$

↙ Element jk der Inversen der Kovarianzmatrix.

Praktische Durchführung (bei gegebenen Messwerten):

$V(\hat{a})$ bestimmt durch 2. Ableitung der Funktion $F(\vec{a}) = \sum_{i=1}^n \ln g(y_i; \vec{a})$ für $\vec{a} = \hat{a}$

$$[V(\hat{a})^{-1}]_{jk} = - \left(\frac{\partial^2 F}{\partial a_j \partial a_k} \right)_{\vec{a} = \hat{a}}$$

Beachte: Wegen Normalität sind alle höhere Ableitungen

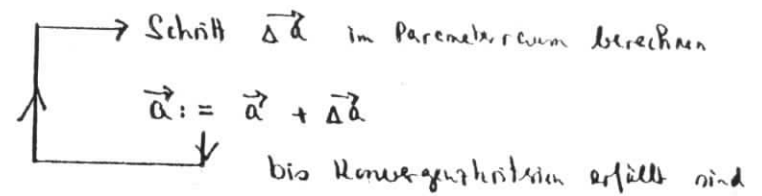
asymptotisch Null $\rightarrow \frac{\partial^2 F}{\partial a_j \partial a_k} = \text{Konstante}$

3. Numerische Durchführung

Bestimmung des Maximums einer log Likelihood-Fkt. $F(\vec{a})$ muß i.a. iterativ erfolgen.

Iterative Methode:

Näherungswert \vec{a} vorgeben



Taylor-Entwicklung um Näherungswert \vec{a}

$$F(\vec{a} + \Delta \vec{a}) = F(\vec{a}) + \vec{g}^T \cdot \Delta \vec{a} - \frac{1}{2} \Delta \vec{a}^T G \Delta \vec{a} + \dots$$

mit $g_j = \frac{\partial F}{\partial a_j}$

$$G_{jk} = - \frac{\partial^2 F}{\partial a_j \partial a_k} \quad \left[G^{-1} = \text{Kovarianzmatrix} \right]$$

Ableitung von zwei Methoden aus der nach dem quadrat. Glied abgebrochen Entwicklung:

→ a.) Newton-Verfahren

Entwicklung nach \vec{a} differenzieren, Ableitung gleich Null setzen:

$$\vec{g} - G \cdot \Delta \vec{a} = 0$$

$$\Rightarrow \boxed{\Delta \vec{a}_N = G^{-1} \cdot \vec{g}}$$

(durch Lösen eines linearen Gleichungssystems)

Wenn $F(\vec{a})$ in \vec{a} quadratisch ist, d.h. alle höheren Ableitungen verschwinden, führt bereits ein Schritt zum Maximum.

Da Likelihood-Funktionen asymptotisch normal sind, sind die log. Likelihood-Funktionen $F(a)$ asymptotisch quadratisch in \vec{a} .

Für endliche n gilt diese Eigenschaft angenähert zumindest in der Nähe des Maximums

b.) Gradienten-Methode

Der Schritt wird in Richtung des Gradienten \vec{g} gewählt (Richtung des stärksten Anstiegs der Funktion $F(\vec{a})$), also mit skalarem Größe t :

$$\Delta \vec{a} = t \cdot \vec{g}$$

Einsetzen in quadratische Näherung,

Funktion von t :

$$F(\vec{a} + t \cdot \vec{g}) = F(\vec{a}) + t \vec{g}^T \vec{g} - \frac{1}{2} t^2 \vec{g}^T G \vec{g} +$$

Nullsetzen der Ableitung nach t :

$$\vec{g}^T \vec{g} - t \vec{g}^T G \vec{g} = 0$$

$$\text{Lösung: } t = \frac{\vec{g}^T \vec{g}}{\vec{g}^T G \vec{g}}$$

Damit ergibt sich bei der Gradientenmethode ein Schritt:

$$\Delta \vec{a}_G = \frac{\vec{g}^T \vec{g}}{\vec{g}^T G \vec{g}} \vec{g}$$

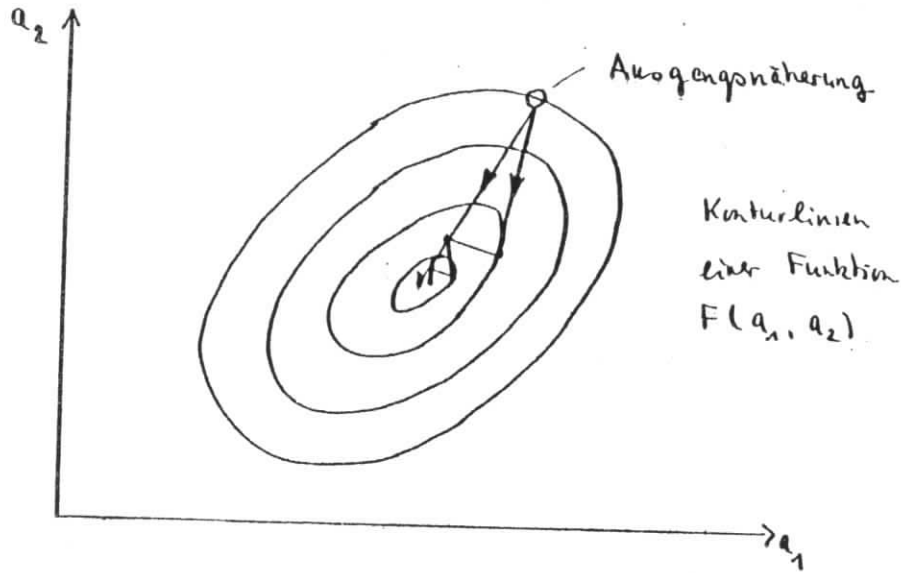
Eigenschaften der beiden Methoden:

→ Gradientenmethode konvergiert langsam, aber sicher

→ Newton-Methode ist normalerweise vorzuziehen, denn (bei guter Ausgangsnäherung) führt sie schnell zur Konvergenz; bei ungünstiger Ausgangsnäherung führt sie jedoch zur Divergenz \Rightarrow auf langsam aber sichere Methode umschalten.

Beispiel für die Schritte bei

- Newton - Methode
- Gradienten - Methode



Verfahren bei Programm MLFIT A:

- normalerweise Newton - Schritte
- bei Divergenz: Interpolation zwischen Newton - und Gradientenmethode mit steuerbarer Gewichtung

$$\vec{\Delta d} = K^{-1} \vec{g}$$

$$\text{mit } K = G + \gamma \frac{\vec{g}^T G \vec{g}}{\vec{g}^T \vec{g}} I$$

$$\gamma = 0.1 (2^k - 1) \quad k=0,1,2$$

$l = 0$ $\gamma = 0$ Newton - Schritt
 $l > 0$ $\gamma > 0$ Interpolation

Fälle:

Divergenz l um 2 erhöhen
 langsame Konvergenz l um 1 erhöhen
 schnelle Konvergenz l um 1 erniedrigen (bis $l=0$)

Konvergenzerkennung:

Vergleich von

tatsächlicher Funktionsänderung $\Delta F_t = F(\vec{d} + \Delta \vec{d}) - F(\vec{d})$
 mit
 erwarteter Funktionsänderung $\Delta F_e = \vec{g}^T \Delta \vec{d} - \frac{1}{2} \Delta \vec{d}^T G \Delta \vec{d}$

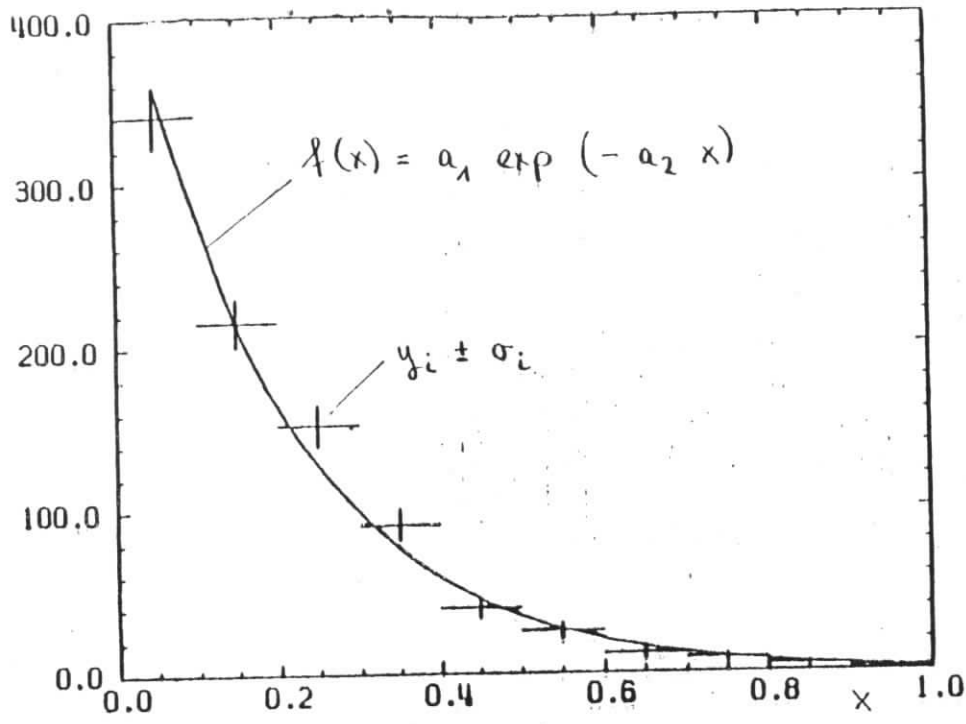
Iterationsverfahren konvergierend, wenn $\Delta F_t \approx \Delta F_e$
 divergierend, wenn $\Delta F_t < 0$
 langsam konvergierend, wenn $\Delta F_t \ll \Delta F_e$

Bei Newton - Schritt ist

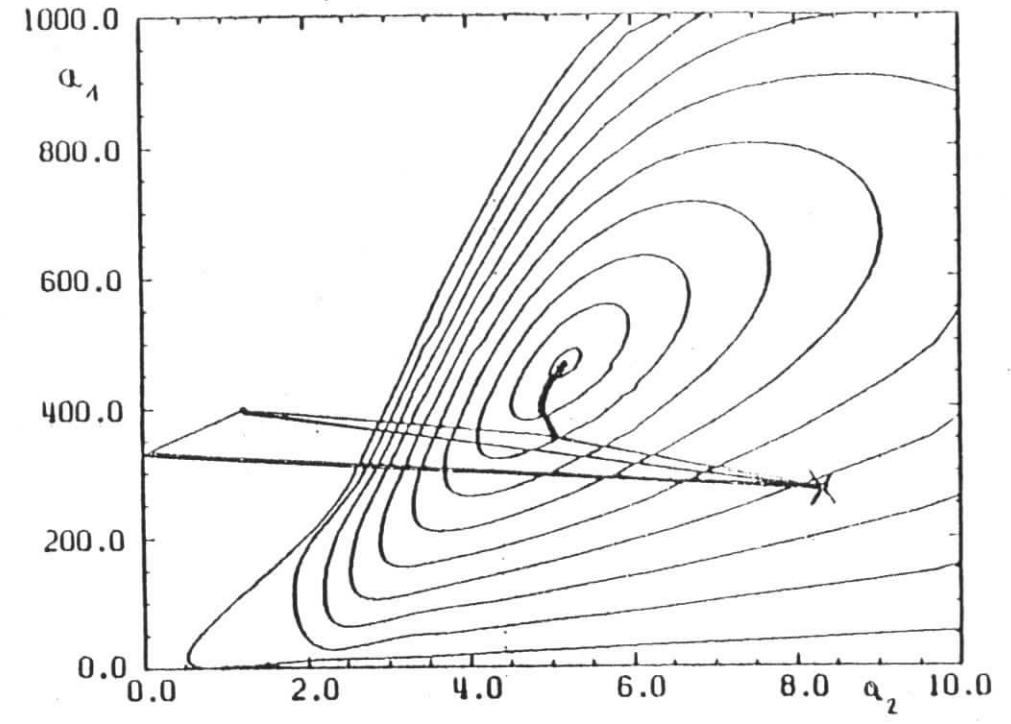
$$\Delta F_e = \vec{g}^T G^{-1} \vec{g} - \frac{1}{2} \vec{g}^T G^{-1} G G^{-1} \vec{g} - \frac{1}{2} \vec{g}^T G \vec{g}$$

Auf dem Fehlerellipsoid (Fehlerellipse bei 2 Parametern) zu einer Standardabweichung ist die Abweichung um maximalen Wert gerade $1/2$.

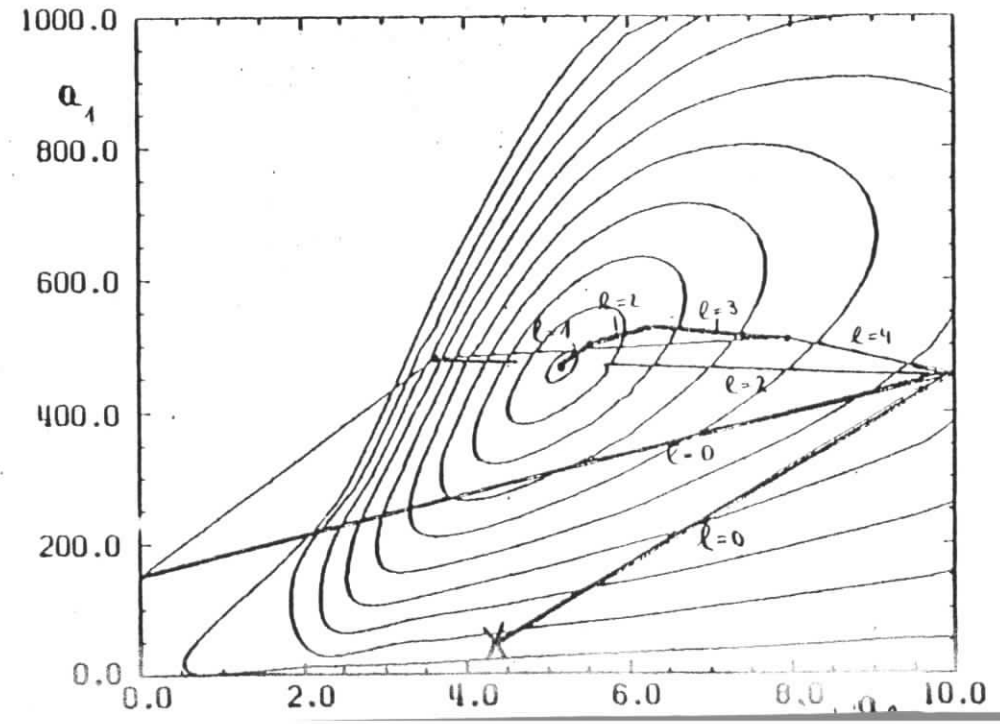
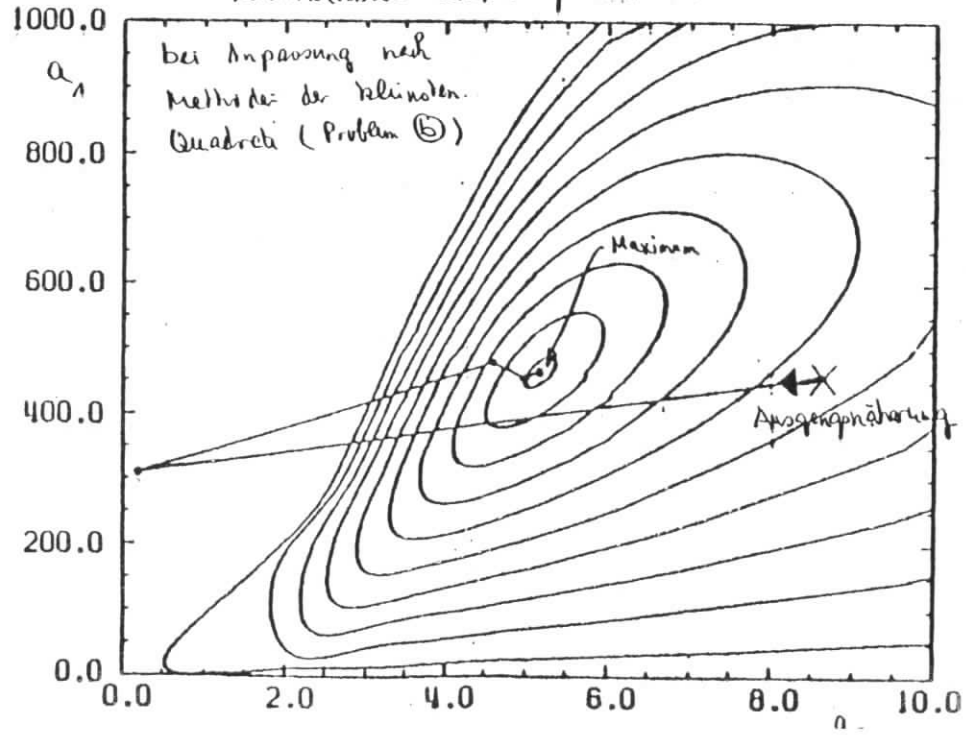
Konvergenz erreicht, wenn $\Delta F_t \approx \Delta F_e < \frac{1}{2}$
 (in der Nähe des Max. ist die quadratische Näherung gut)



Beispiel für Schritte \vec{a} zum Maximum



Konturlinien = Linien gleichen Werts von $F(a)$



Für die Durchführung der Rechnungen müssen die 1. und 2. Ableitungen der Funktion $F(\vec{a})$ nach den Parametern a_1, a_2, \dots, a_p berechnet werden.

Diese Ableitungen ergeben sich bei Problemen (b) und (c) bereits aus den 1. Ableitungen von $f(x_i, \vec{a})$ nach den Parametern a_1, a_2, \dots, a_p .

Problem (b): normalverteilte Daten

$$F(\vec{a}) = -\frac{1}{2} \sum_{i=1}^n w_i (f(x_i, \vec{a}) - y_i)^2 \quad w_i = 1/\sigma_i^2$$

$$g_j = \frac{\partial F}{\partial a_j} = \sum_{i=1}^n w_i \frac{\partial f}{\partial a_j} (y_i - f(x_i, \vec{a}))$$

$$G_{jk} = \frac{\partial^2 F}{\partial a_j \partial a_k} = \sum w_i \left(\frac{\partial f}{\partial a_j} \frac{\partial f}{\partial a_k} - \frac{\partial^2 f}{\partial a_j \partial a_k} (y_i - f(x_i, \vec{a})) \right)$$

vernachlässigen

Problem (c): Poisson verteilte Daten

$$F(\vec{a}) = \sum_{i=1}^n y_i \ln f(x_i, \vec{a}) - \sum_{i=1}^n f(x_i, \vec{a})$$

$$g_j = \sum_{i=1}^n y_i \frac{\frac{\partial f}{\partial a_j}}{f(x_i, \vec{a})} - \sum_{i=1}^n \frac{\partial f}{\partial a_j}$$

$$G_{jk} = \sum_{i=1}^n y_i \frac{\frac{\partial f}{\partial a_j} \frac{\partial f}{\partial a_k} - \frac{\partial^2 f}{\partial a_j \partial a_k} f(x_i, \vec{a})}{f^2(x_i, \vec{a})} - \sum_{i=1}^n \frac{\partial^2 f}{\partial a_j \partial a_k}$$

vernachlässigen

- 116 -

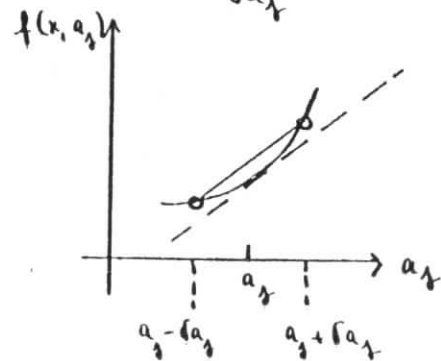
Die in den Matrixelementen G_{jk} auftretenden 2. Ableitungen von $f(x_i, \vec{a})$ können weggelassen werden, ohne dass das Ergebnis merklich beeinflusst wird. (Position des Maximums wird nicht verändert).

→ Arbeiterspenis G_{jk} positiv definit

• Analytische Berechnung von $\frac{\partial f}{\partial a_j}$ häufig kompliziert

→ numerische Berechnung

$$\frac{\partial f}{\partial a_j} \approx \frac{f(x_i, a_j + \delta a_j) - f(x_i, a_j - \delta a_j)}{2 \delta a_j}$$



→ wenig empfindlich gegen Größe der Schrittweite δa_j

(bei zu kleinem δa_j Rundungsfehler!)

Diese Möglichkeit kann in MLFITA benutzt werden, die δa_j werden im Laufe der Rechnung dem Problem angepasst.

Wenn analytische Berechnung der Ableitungen nicht möglich und $F(a)$ sehr kompliziert ist (Probleme (c) und besonders (d))

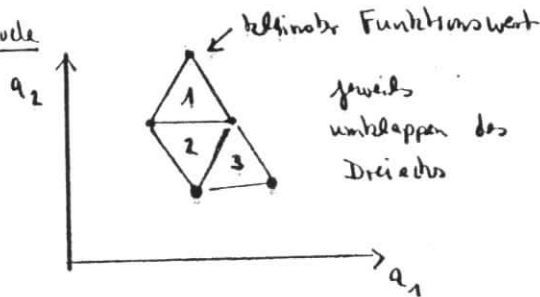
→ nur Berechnung von $F(a)$ programmieren und allgemeines Maximierungsverfahren verwenden

z.B. Programm MINUIT (CERN)
(Beschreibung in DEST-Bibliothek)

Enthält 2 wählbare Verfahren:

① Simplex-Methode

anschaulich
(bei 2 Parametern)



robustes Verfahren, daß sich dem Maximum nähert, jedoch in der Nähe des Maximums langsam konvergiert

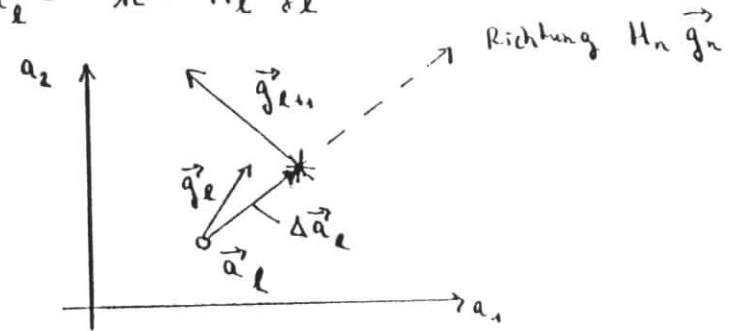
② Newton-Verfahren mit iterativer Berechnung der ^{Inverse} der Matrix der 2. Ableitungen aus den numerisch berechneten 1. Ableitungen von $F(\vec{a})$ nach den a_j (MIGRAD)

Iterative Methode (nach Davidon):

H_e = Näherung für Matrix G^{-1}
(zu Beginn z.B. $H_0 = I$)

g_e = Vektor der 1. Ableitungen (numerisch)

$$\Delta \vec{a}_e = h \cdot H_e \vec{g}_e$$



Suche in Richtung $H_e \vec{g}_e$ bis zum Maximum (→ definiert h)

$$\vec{a}_{e+1} = \vec{a}_e + \Delta \vec{a}_e$$

\vec{g}_{e+1} berechnen (bei \vec{a}_{e+1})

Neue Näherung für G^{-1} :

$$H_{e+1} = H_e + A_e + B_e$$

$$A_e = h \frac{(H_e \vec{g}_e)(H_e \vec{g}_e)^T}{(H_e \vec{g}_e)^T \vec{d}}$$

$$B_e = \frac{H_e \vec{d} \vec{d}^T H_e}{\vec{d}^T H_e \vec{d}}$$

$$\vec{d} = \vec{g}_{e+1} - \vec{g}_e$$

$H \rightarrow G^{-1}$ nach p Schritten
(bei p Parametern a_1, a_2, \dots, a_p)
bei in \vec{a} quadratischer Funktion

Praxis:

durch Rundungsfehler stark beeinflusst (bei IBM Rechnen)
 H nicht stets positiv definit

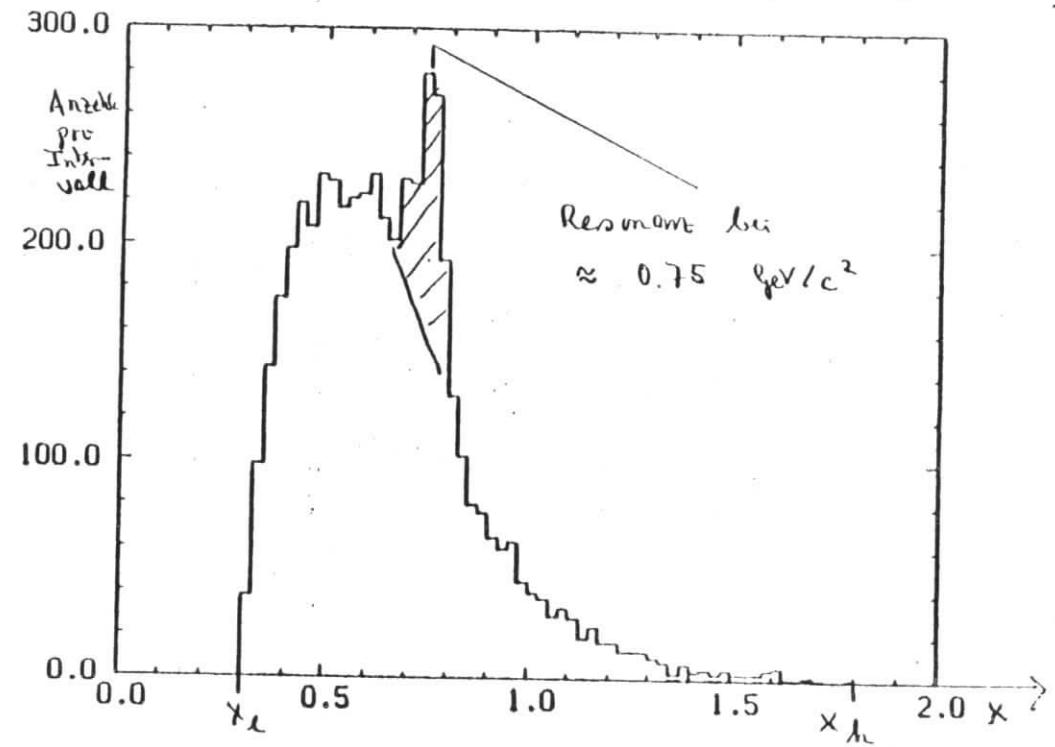
↳ also Vorsicht: MINUIT bricht denn ab mit
Fehlerausdruck (not positive definite); Kovarianzmatrix
strukturell falsch, Maximum nicht erreicht).

Abhilfe:

mit Kommando HESSIAN
numerische Berechnung der Matrix G und
evtl. MINRAD wiederholen

4. Beispiel

Verteilung einer Zwei-Teilchen-Masse gemessen ($x \equiv \text{Masse}$)



Problem: Resonanzanteil über offtem Untergrund bestimmen

Ansatz: $f(x) = f_p(x) [a_1 + a_2 f_R(x)]$

Untergrundverlauf Breit-Wigner

Breit-Wigner: $f_R(x) = \frac{\frac{1}{4} a_4^2}{(x - a_3)^2 + \frac{1}{4} a_4^2}$

$a_3 = \text{Resonanzmasse}$

$a_4 = \text{FWHM (Breite der Resonanz)}$

Funktion für Untergrundverlauf:

Möglicher Ansatz: Polynom in x

Nachteil: viele Parameter notwendig, um Verlauf zu beschreiben

→ möglichst wenige freie Parameter

Besserer Ansatz

① Nullstellen x_e und x_h fest einbauen:

$$f(x) = (x - x_e) \cdot (x_h - x)$$

② Abfall bei großen x etwa exponentiell

$$f(x) = (x - x_e) \cdot (x_h - x) e^{-a_5 x}$$

③ Umschreiben, damit Parameter a_1, a_2 leicht abzulesen sind für Ausgangsnäherung

$$f_p(x) = \frac{4(x - x_e)(x_h - x)}{(x_h - x_e)^2} e^{-a_5(x - \frac{1}{2}(x_e + x_h))}$$

$$f_p(x) = 1 \text{ bei } x = \frac{1}{2}(x_e + x_h)$$

→ a_1 aus Abb. ablesen

Durchführung mit MLFITA:

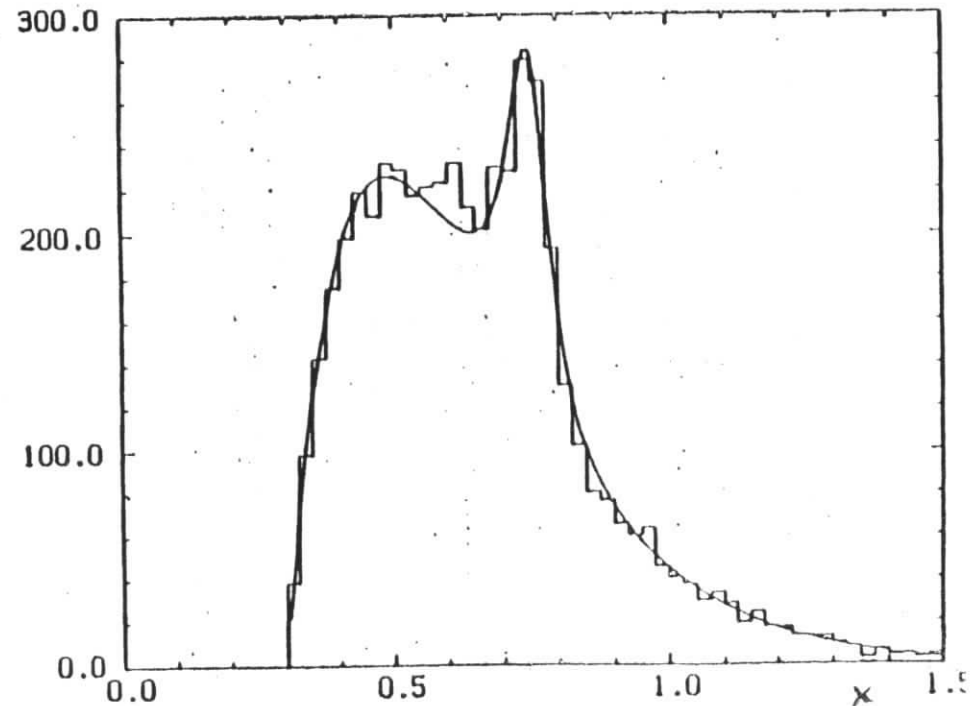
```
COMMON/MLCOM/EV,FV,WT,PA(50),ST(50),DA(50),AR(300)
REAL Y(68),X(68)
PHSP(X)=4.0*(X-0.3)*(2.0-X)*EXP(-PA(5)*(X-1.15))/2.89
BWIG(X)=0.25*PA(4)**2/((X-PA(3))**2+0.25*PA(4)**2)
CALL MLFITA(5,10,2)
PA(1)=17.0
PA(2)=41.0
PA(3)=.750
PA(4)=0.1
PA(5)=5.0
ST(1)=0.1*PA(1)
ST(2)=0.1*PA(2)
ST(3)=0.01
ST(4)=0.01
ST(5)=0.1
10 DO 20 I=1,68
EV=Y(I)
WT=1.0
IF (EV.GT.0.0) WT=1.0/EV
X=XX(I)
15 FV=PHSP(X(I))*(PA(1)+PA(2)*BWIG(X(I)))
CALL MLDER(I,15)
20 CALL MLLESQ
CALL MLITER(1,610)
```

Definition von Anfangswerten PA(I) und Schrittweiten für num. Diff. ST(I)

Iterationen
 10 DO 20 I=1,68
 EV=Y(I)
 WT=1.0
 IF (EV.GT.0.0) WT=1.0/EV
 X=XX(I)
 15 FV=PHSP(X(I))*(PA(1)+PA(2)*BWIG(X(I)))
 CALL MLDER(I,15)
 20 CALL MLLESQ
 CALL MLITER(1,610)

Siehe
 üb
 68 D
 pu

MLLESQ durch MLPOIS ersetzen → Poissonverteilter Daten



C. Methode der kleinsten Quadrate

1.) Problemstellung der vermittelnden Ausgleichung

Gegeben n Werte y_1, y_2, \dots, y_n (Messwerte)
 jedem y_i zugeordnet: σ_i = Standardabweichung
 und unabhängige Koordinate x_i

Modell:

Abhängigkeit der Variablen y von x beschrieben durch
 Funktion $f(x; a_1, a_2, \dots, a_p)$,
 die von p Parametern $a_1, a_2, a_3, \dots, a_p$ abhängt

Voraus.:

- (1) $E(y_i - f(x_i, a_1, \dots, a_p)) = 0$
- (2) $V(y_i - f(x_i, a_1, \dots, a_p)) = \sigma_i^2$

Keine Voraussetzungen über Verteilung!

Bei diesem Problem sind folgende Fragestellungen möglich:

- ① Bestimmung der Parameter a_1, \dots, a_p
- ② Test des Modells (Anpassungstest)
- ③ Ausgleich der Messwerte

Residuen: $r_i = y_i - f(x_i, a_1, \dots, a_p)$

- Voraus.: $\begin{cases} (1) E(r_i) = 0 \\ (2) V(r_i) = \sigma_i^2 \end{cases}$

Prinzip der Methode der kleinsten Quadrate:

Parameterwerte bestimmen, für die

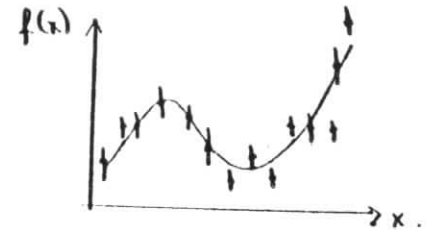
$$Q = \sum_{i=1}^n \frac{r_i^2}{\sigma_i^2}$$

ein Minimum annimmt. ($\rightarrow \hat{a}$).

1. Beispiel:

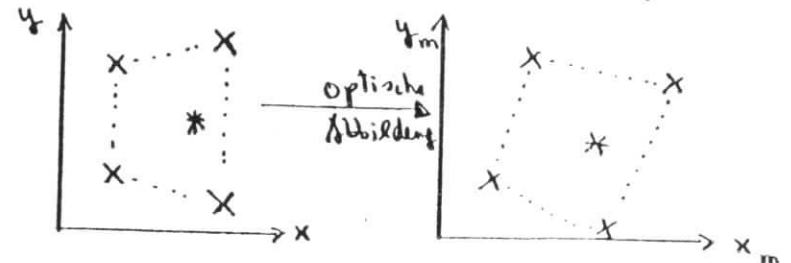
Polynomfit:

$$f(x) = a_1 + a_2 x + a_3 x^2 + a_4 x^3$$



2. Beispiel:

Bestimmung von Transformationskoeffizienten für lin. Abb.



Wahre Koordinaten von Bezugspunkten X

Messebene

Annahme: optische Abbildung linear

$$a_1 x + a_2 y + a_3 \rightarrow x_m$$

$$a_4 x + a_5 y + a_6 \rightarrow y_m$$

Lösung des Problems, die Koeffizienten der Transformation zu bestimmen, mit Methode der kleinsten Quadrate möglich

$$\text{Residuen} \quad r_x = x_m - (a_1 x + a_2 y + a_3)$$

$$r_y = y_m - (a_4 x + a_5 y + a_6)$$

$a_1 \dots a_6$ so bestimmen, daß
 $\longrightarrow Q = \sum r_x^2 + \sum r_y^2$ minimal wird

\Rightarrow je 2 Gleichungen pro (x, y) Paar
 6 Unbekannte

\Rightarrow mindestens 3 Punkte notwendig
 bei > 3 Punkten überbestimmtes Problem

Dieses Beispiel zur Demonstration, daß mit der Methode der kleinsten Quadrate nicht nur Kurven angepaßt werden können (allerdings hier etwas andere Struktur der Gleichungen, andere Bezeichnungen)

2. Lineare Probleme

$f(x, \vec{a})$ soll nur linear von den a_j abhängen

z.B. $f(x, \vec{a}) = a_1 \cdot p_1(x) + a_2 \cdot p_2(x) + a_3 \cdot p_3(x)$

$p_2(x)$ unabh. von den a_j

$$f(x, \vec{a}) = \sum_{j=1}^p p_j(x) \cdot a_j$$

Matrix-Schreibweise:

$$\text{Messwerte } \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{Funktionswerte } \vec{f} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}$$

$$\vec{f} = \vec{A} \vec{a} \quad \text{mit } \vec{A} = \begin{pmatrix} p_1(x_1) & p_2(x_1) & \dots & p_p(x_1) \\ p_1(x_2) & p_2(x_2) & \dots & p_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(x_n) & p_2(x_n) & \dots & p_p(x_n) \end{pmatrix}$$

d.h. $A_{ij} = p_j(x_i)$

$$\text{Parametervektor } \vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

Residuenvektor:

$$\vec{r} = \vec{y} - \vec{f} = \vec{y} - \vec{A} \vec{a}$$

Kovarianzmatrix $V(\vec{y}) = n \times n$ Matrix

Gewichtsmatrix: $W = V^{-1}$

Bei unabhängigen Messwerten sind beide Matrizen Diagonalmatrizen, d.h.

$$V(\vec{y}) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \\ & & & \sigma_n^2 \end{pmatrix} \quad W = \begin{pmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \frac{1}{\sigma_2^2} & \\ 0 & & \ddots \\ & & & \frac{1}{\sigma_n^2} \end{pmatrix}$$

Im folgenden werden in den Formeln allgemeine Matrizen V bzw. W benutzt.

Summe der Quadrate der gewichteten Residuen:

$$Q = \vec{r}^T W \vec{r}$$

(das ist bei unabhängigen Messwerten)

$$Q = \sum_{i=1}^n \frac{r_i^2}{\sigma_i^2}$$

Einschreiben der Residuenvektoren, ausmultiplizieren:

$$Q(\vec{a}) = (\vec{y} - A\vec{a})^T W (\vec{y} - A\vec{a}) \\ = \vec{y}^T W \vec{y} - 2 \vec{a}^T A^T W \vec{y} + \vec{a}^T A^T W A \vec{a}$$

Halbe Ableitung von Q nach \vec{a} :

$$\frac{1}{2} \frac{\partial Q}{\partial \vec{a}} = -A^T W \vec{y} + A^T W A \vec{a}$$

\vec{a} durch $\frac{\partial Q}{\partial \vec{a}} = 0$ bestimmt.

$$-A^T W \vec{y} + A^T W A \vec{a} = 0$$

$$\longrightarrow \vec{\hat{a}} = (A^T W A)^{-1} A^T W \vec{y}$$

(Lösen eines linearen Gleichungssystems)

$\vec{\hat{a}}$ = lineare Abbildung $B \vec{y}$ von \vec{y} mit
 $B = (A^T W A)^{-1} A^T W$

• (Quadrat der Fehlerfortpflanzung), Kovarianzmatrix von $\vec{\hat{a}}$

$$V(\vec{\hat{a}}) = V(B \vec{y}) = B V(\vec{y}) B^T \\ = (A^T W A)^{-1} A^T W W^{-1} W A (A^T W A)^{-1} \\ = (A^T W A)^{-1}$$

\hookrightarrow war bereits oben zur Bestimmung von $\vec{\hat{a}}$ benutzt.

• Wert von $Q(\vec{\hat{a}})$ beim Minimum durch Einsetzen:

$$Q_{\min} = \vec{y}^T W \vec{y} - 2 \vec{\hat{a}}^T A^T W \vec{y} + \vec{\hat{a}}^T A^T W A \vec{\hat{a}} \\ = \vec{y}^T W \vec{y} - 2 \vec{\hat{a}}^T A^T W \vec{y} + \vec{\hat{a}}^T A^T W A (A^T W A)^{-1} A^T W \vec{y} \\ = \vec{y}^T W \vec{y} - \vec{\hat{a}}^T A^T W \vec{y}$$

war bereits oben zur Berechnung von $\vec{\hat{a}}$ notwendig.

- Ausgewählte Werte von den y_i ("vermittelnde Ausgleichung")
durch Einsetzen von $\vec{\hat{a}}$

$$\vec{\hat{y}} = A \vec{\hat{a}} = \underbrace{A (A^T W A)^{-1} A^T W}_{\text{lineare Abbildung von } \vec{y} \text{ auf } \vec{\hat{y}}}$$

Kovarianzmatrix der $\vec{\hat{y}}$,

$$V(\vec{\hat{y}}) = A (A^T W A)^{-1} A^T$$

Rechenoperationen zur Lösung eines linearen Problems

Definition der Vektoren \vec{y}
und Matrizen A und W

$$C = A^T W A$$

$$\vec{b} = A^T W \vec{y}$$

$$C^{-1} \text{ durch Invertieren} \quad \mapsto \quad V(\vec{\hat{a}}) = C^{-1}$$

$$\vec{\hat{a}} = C^{-1} \vec{b}$$

$$Q_{\min} = \vec{y}^T W \vec{y} - \vec{\hat{a}}^T \vec{b}$$

$$\vec{\hat{y}} = A \vec{\hat{a}}$$

$$V(\vec{\hat{y}}) = A C^{-1} A^T$$

$$f(x) = a_1 + a_2 x^2$$

Beispiel: Bei der Gradanpassung (und unabhängigen Messwerten) ergeben sich folgende Matrizen:

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

$$C = A^T W A = \begin{pmatrix} \sum w_i & \sum w_i x_i \\ \sum w_i x_i & \sum w_i x_i^2 \end{pmatrix}$$

$$\vec{b} = A^T W \vec{y} = \begin{pmatrix} \sum w_i y_i \\ \sum w_i y_i x_i \end{pmatrix}$$

3. Lösungseigenschaften und Folgerungen

- (1) $\vec{\hat{a}}$ ist ein erwartungstreuer Schätzwert von \vec{a}^0
d.h. $E(\vec{\hat{a}}) = \vec{a}^0$

Beweis:

$$\vec{\hat{a}} = (A^T W A)^{-1} A^T W \vec{y}$$

$$E(\vec{\hat{a}}) = (A^T W A)^{-1} A^T W \underbrace{E(\vec{y})}_{A \vec{a}^0} = \vec{a}^0$$

- (2) (Gauß-Markoff-Theorem) Unter den Schätzern die erwartungstreu und Linearkombinationen der Werte y_i sind, liefert die Methode der kleinsten Quadrate die kleinste Varianz.

Beweis: $V(\vec{\hat{a}}) = (A^T W A)^{-1} = C^{-1}$

Fall: Daten y_i nicht unabhängig, d.h. $V(\vec{y})$ und $W(\vec{y})$ nicht diagonal

Durch lineare Transformation von \vec{y} mit einer quadratischen Matrix P ergibt sich ein Vektor $\vec{z} = P \vec{y}$

mit Kovarianzmatrix

$$V(\vec{z}) = P V(\vec{y}) P^T$$

bzw. $V(\vec{y}) = P^{-1} V(\vec{z}) (P^{-1})^T$

P kann so gewählt werden, daß $V_z = I$ (Einheitsmatrix) ist. Dann ist

$$Q = (\vec{z} - PA\vec{a})^T (\vec{z} - PA\vec{a}) \\ = (\vec{y} - A\vec{a})^T P^T P (\vec{y} - A\vec{a})$$

Es gilt: $(P^T P)^{-1} = P^{-1} (P^{-1})^T = V(\vec{y})$

↳ die beiden Probleme (in \vec{z} und in \vec{y}) sind äquivalent.

Fall: normalverteilte Daten y_i

Dann sind (wegen Linearität der Abbildung) auch die \hat{a}_f normalverteilt, und

Q_{\min} ist χ^2 verteilt mit $(n-p)$ Freiheitsgr.

↳ Anpassungstest möglich

Ein anderer erwartungstreu Schätzwert sei $\vec{a}^* = U W \vec{y}$

Aus $E(\vec{a}^*) = U W E(\vec{y}) = U W A \vec{a}^0 = \vec{a}^0$

folgt: $U W A = I$

$$V(\vec{a}^*) = U W W^{-1} W U = U W U^T$$

↳ vergleiche $U W U^T$ mit $(A^T W A)^{-1} = C^{-1}$

$$U W U^T = C^{-1} + (U - C^{-1} A^T) W (U - C^{-1} A^T)^T \\ = C^{-1} + U W U^T + C^{-1} A^T W A C^{-1} - C^{-1} A^T W U - U W A C^{-1} \\ = C^{-1} + U W U^T + C^{-1} - C^{-1} - C^{-1} = U W U^T \\ \Rightarrow V(\vec{a}^*) = V(\vec{a}) + \underbrace{(U - C^{-1} A^T) W (U - C^{-1} A^T)^T}$$

hat Diagonalelemente $>, 0$ ($= 0$ nur für $U = C^{-1} A$)

$$\longrightarrow V(\vec{a}^*)_{ff} \gg V(\vec{a})_{ff} \quad \text{q.e.d.}$$

• (3) Der Erwartungswert von Q_{\min} ist $(n-p)$:

$$E(Q_{\min}) = n - p$$

↳ erlaubt Bestimmung eines Feldes, mit dem $V(\vec{y})$ und damit auch $V(\vec{a})$ multipliziert werden muß

$$s^2 = \frac{Q_{\min}}{n-p} \quad \Rightarrow V(\vec{a}) = s^2 (A^T W A)^{-1}$$

4. Nichtlineare Probleme

Iterative Lösungsmethoden notwendig, wenn die Funktion $f(x_i, a_1, \dots, a_p)$ nicht linear von den a_j abhängt.

Ausgangsnäherung $\vec{a}^{(0)}$

Linearisierung des Problems für jede Näherung $\vec{a}^{(v)}$

Berechnung von Korrekturen $\Delta \vec{a}$

$$\vec{a}^{(v+1)} := \vec{a}^{(v)} + \Delta \vec{a}$$

Linearisieren:

$$f(x_i, \vec{a}^{(v)} + \Delta \vec{a}) \approx f(x_i, \vec{a}^{(v)}) + \underbrace{\sum_{j=1}^p \Delta a_j \frac{\partial f(x_i, \vec{a})}{\partial a_j}}_{A \Delta \vec{a}} \Big|_{\vec{a} = \vec{a}^{(v)}}$$

$$A = \begin{pmatrix} \frac{\partial f(x_1)}{\partial a_1} & \frac{\partial f(x_1)}{\partial a_2} & \dots & \frac{\partial f(x_1)}{\partial a_p} \\ \frac{\partial f(x_2)}{\partial a_1} & \dots & & \\ \vdots & & & \\ \frac{\partial f(x_n)}{\partial a_1} & & & \frac{\partial f(x_n)}{\partial a_p} \end{pmatrix}$$

$$A_{ik} = \frac{\partial f(x_i)}{\partial a_k}$$

Einsetzen der Näherung in die gewichtete Quadratsumme Q und Nullsetzen der Ableitungen ergibt analog zu Kap. (2) mit

$$\vec{r} = \vec{y} - \vec{f}(x, \vec{a}^{(v)})$$

für die Korrekturen:

$$\longrightarrow \Delta \vec{a} = (A^T W A)^{-1} A^T W \vec{r}$$

Diese Korrektur $\Delta \vec{a}$ ergibt in der linearen Näherung eine Reduktion der gewichteten Quadratsumme um

$$\Delta Q = \Delta \vec{a}^T A^T W \vec{r}$$

Wenn $\Delta Q < 1$ (vgl. $E(Q)_{\min} = n - p$)

kann angenommen werden, daß lineare Näherung ausreichend genau ist, und weitere Iterationen nicht erforderlich sind.

Eigenschaften: In der Praxis werden im nichtlinearen Fall die gleichen Formeln für $V(\vec{a})$ wie im linearen Fall (Kap. 2) benutzt;

aber: $\left\{ \begin{array}{l} \text{kein Beweis optimaler Eigenschaften,} \\ \text{Verteilung von } Q \text{ bei normalverteilten Daten} \\ \text{nur noch annähernd } \chi^2 \end{array} \right.$

Rechneschema bei nichtlinearen Problemen

Definition von \vec{y} und W

Definition von Anfangswerten $\vec{a}^{(0)}$

$v = 0$

Iteration:

$$\vec{r} = \vec{y} - f(x, \vec{a}^{(v)})$$

$$A \text{ mit } A_{ik} = \left. \frac{\partial f(x_i, a)}{\partial a_k} \right|_{\vec{a} = \vec{a}^{(v)}}$$

$$C = A^T W A$$

$$\vec{b} = A^T W \vec{r}$$

$$\Delta \vec{a} = C^{-1} \vec{b}$$

$$v = v + 1$$

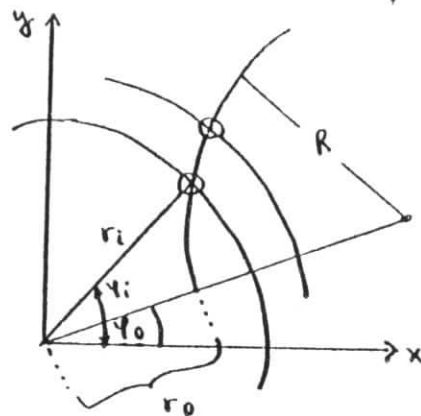
$$\vec{a}^{(v)} = \vec{a}^{(v-1)} + \Delta \vec{a}$$

$$\Delta Q = \Delta \vec{a}^T \vec{b}$$

für $\Delta Q > 1$ weitere Iteration notwendig

Beispiel für nichtlineares Problem:

Bestimmung von Spurparametern durch Anpassung eines Kreises an Spurpunkte in Ebene $\perp \vec{B}$



Messung mit zylindr. Können

φ_i zu Radius r_i

Gesucht: Spurparameter, z.B. R ($\propto p$ Impuls)

$$\text{Kreisgleichung: } r_i^2 - 2 r_i (R + r_0) \cos(\varphi - \varphi_0) + (R + r_0)^2 = R^2$$

umformen in ' (mit $\alpha = 1/R$) ungünstig bei $R \gg r_0, r_i$

$$\alpha (r_i^2 + r_0^2) - 2 r_i (\alpha + \alpha r_0) \cos(\varphi - \varphi_0) + 2 r_0 = 0$$

$$\varphi_i = \varphi_0 + \arccos \left(\frac{\alpha (r_i^2 + r_0^2) + 2 r_0}{2 r_i (\alpha + \alpha r_0)} \right)$$

$$= f(\varphi_0, \alpha, r_0 | r_i)$$

Spurparameter

Zur Anpassungsrechnung muß $f(\varphi_0, \alpha, r_0)$ nach φ_0, α, r_0 differenziert werden.

Probleme: Messfehler, Vieldeutigkeit.

5. Bedingte Ausgleichung

Gleiche Problemstellung wie in Kap. 1, d.h.

gegeben \vec{y} , $V(\vec{y})$; $W(\vec{y}) = V^{-1}(\vec{y})$

lineares Modell

$$E(\vec{y}) = A \vec{a} \quad \text{mit (1) } E(\vec{y} - A \vec{a}) = 0$$

$$(2) V(\vec{y} - A \vec{a}) = W^{-1}$$

zusätzlich:

m lineare Bedingungsbedingungen

$$h_k + \sum_{j=1}^p H_{kj} a_j = 0 \quad k=1, \dots, m$$

bzw. $\vec{h} + H \vec{a} = 0$ in Matrixschreibweise

Methode der kleinsten Quadrate:

$$\text{Minimum von } Q = (\vec{y} - A \vec{a})^T W (\vec{y} - A \vec{a})$$

$$\text{unter Bedingung } \vec{h} + H \vec{a} = 0$$

→ Methode der Lagrange'schen Multiplikatoren
zusätzlich m Parameter λ_k einführen ($\vec{\lambda}$)

$$Q(\vec{a}, \vec{\lambda}) = (\vec{y} - A \vec{a})^T W (\vec{y} - A \vec{a}) + 2 \vec{\lambda}^T (\vec{h} + H \vec{a})$$

Notwendige Bedingung für Minimum von $Q(\vec{a}, \vec{\lambda})$ ist
Verschwinden der Ableitungen von Q nach \vec{a} und $\vec{\lambda}$:

$$\frac{1}{2} \frac{\partial Q}{\partial \vec{a}} = A^T W A \vec{a} - A^T W \vec{y} + H^T \vec{\lambda}$$

$$\frac{1}{2} \frac{\partial Q}{\partial \vec{\lambda}} = \vec{h} + H^T \vec{a}$$

→ 2 gekoppelte lineare Gleichungssysteme für die
Vektoren \vec{a} und $\vec{\lambda}$

Matrixaufschreibung in Blockform:

$$\left(\begin{array}{c|c} A^T W A & H^T \\ \hline H^T & \emptyset \end{array} \right) \begin{pmatrix} \vec{a} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} A^T W \vec{y} \\ -\vec{h} \end{pmatrix}$$

oder mit den Abkürzungen $C = A^T W A$
 $B = A^T W \vec{y}$

$$\left(\begin{array}{c|c} C & H^T \\ \hline H & \emptyset \end{array} \right) \begin{pmatrix} \vec{a} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} \vec{b} \\ -\vec{h} \end{pmatrix}$$

→ Inverse der links stehenden Matrix wird benötigt

$$\left(\begin{array}{cc} C & H^T \\ H^T & \emptyset \end{array} \right)^{-1} = \begin{pmatrix} P & F \\ F & R \end{pmatrix}$$

C und P: $p \times p$ Matrizen

\emptyset und R: $m \times m$ — —

H und F: $m \times p$ — —

Multiplikation der beiden Matrizen \rightarrow Einheitsmatrix $\mathbb{1}$
bzw.:

$$C P + H^T F = \mathbb{1}$$

$$C F^T + H^T R = \emptyset$$

$$H P = 0$$

$$H F^T = \mathbb{1}$$

Aus diesen Gleichungen erhält man die Matrizen P , F und R wie folgt:

$$F = (H C^{-1} H^T)^{-1} H C^{-1}$$

$$P = C^{-1} - C^{-1} H^T (H C^{-1} H^T)^{-1} H C^{-1}$$

$$R = - (H C^{-1} H^T)^{-1}$$

$$\begin{aligned} \vec{a} &= P \vec{b} - F^T \vec{h} = P A^T W \vec{y} - F^T \vec{h} \\ \vec{\lambda} &= F \vec{b} - R \vec{h} = F A^T W \vec{y} - R \vec{h} \end{aligned}$$

\vec{a} und $\vec{\lambda}$ entstehen durch lineare Abbildungen von \vec{y}

Eigenschaften:

$$(1) \quad E(\vec{a}) = \vec{a}^0$$

$$E(\vec{\lambda}) = \vec{0}$$

Beweis:

$$\begin{aligned} E(\vec{a}) &= E(P A^T W A \vec{a} + F^T H \vec{a}) \\ &= P A^T W A E(\vec{a}) + F^T H E(\vec{a}) \\ &= [P C + F^T H] E(\vec{a}) = \vec{a}^0 \end{aligned}$$

$$\begin{aligned} E(\vec{\lambda}) &= E(F A^T W A \vec{a} + R H \vec{a}) \\ &= [F C + R H] E(\vec{a}) = \vec{0} \end{aligned}$$

$$(2) \quad V \begin{pmatrix} \vec{a} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} P & \emptyset \\ \emptyset & -R \end{pmatrix}$$

Beweis:

$$V \begin{pmatrix} \vec{y} \\ \vec{h} \end{pmatrix} = \begin{pmatrix} W^{-1} & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \quad \begin{pmatrix} \vec{a} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} P & F^T \\ F & R \end{pmatrix} \begin{pmatrix} A^T W \vec{y} \\ -\vec{h} \end{pmatrix}$$

$$\begin{aligned} V \begin{pmatrix} \vec{a} \\ \vec{\lambda} \end{pmatrix} &= \begin{pmatrix} P & F^T \\ F & R \end{pmatrix} \begin{pmatrix} A^T W W^{-1} W A & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \begin{pmatrix} P & F^T \\ F & R \end{pmatrix}^T \\ &= \begin{pmatrix} P A^T W A P & P A^T W A F^T \\ F A^T W A P & F A^T W A F^T \end{pmatrix} \end{aligned}$$

Betrachtung der einzelnen Blockmatrizen: ($A^T W A = C$)

$$\begin{aligned} V(\vec{a}) &= P \underbrace{A^T W A} P = P C P = P (\mathbb{1} - H^T F) \\ &= P - P H^T F = P \end{aligned}$$

$$P \underbrace{A^T W A} F^T = P C F^T = -P H^T R = \emptyset$$

$$V(\vec{\lambda}) = F \underbrace{(A^T W A)} F^T = F C F^T = -R H F^T = -R$$

q. e. d.

$$(3) \quad E(Q_{\min}) = n - p + m$$

Als gegebene Werte \vec{y} :

$$\vec{y} = A \vec{a}$$

Für die Varianz gilt:

$$V(\vec{a}) = A \cdot V(\vec{y}) A^T = A P A^T$$

$$= \underbrace{A C^{-1} A^T}_{\text{wie in Kap. 2}} - \underbrace{A C^{-1} H^T (H C^{-1} H^T)^{-1} H C^{-1} A^T}_{\text{Verkleinerung der Varianz durch zusätzliche Inf. aus den Bedingungsgl.}}$$

6. Direkte Ausgleichung mit nichtlinearen Bedingungs-gleichungen

Iterative Lösungsverfahren notwendig bei nichtlinearem Modell, oder nichtlinearen Bedingungs-gleichungen (Linearisierung für jeden Iterationsschritt)

- Hier:
- direkte Ausgleichung, d.h. $A = 1$ gemessene Variablen $y_i \equiv$ Parameter a_j
 - nichtlineare Bedingungs-gleichungen zwischen den Variablen

Bezeichnungen:

\vec{y}	= Vektor
\vec{y}_m	= Vektor der Messwerte (n Komponenten)
V_y	= W^{-1} Kovarianzmatrix der Messwerte
\vec{f}	= Vektor von m Funktionswerten $f(\vec{y})$

Problem:

Die m Bedingungs-gleichungen

$$f_k(\vec{y}) = 0$$

werden um den Messwerten \vec{y}_m nicht recht erfüllt gesucht sind Korrekturen $\vec{\Delta y}$, bzw. ausgeglichene Werte $\vec{y} = \vec{y}_m + \vec{\Delta y}$, für die

$$f_k(\vec{y}) = 0 \quad \text{gilt.}$$

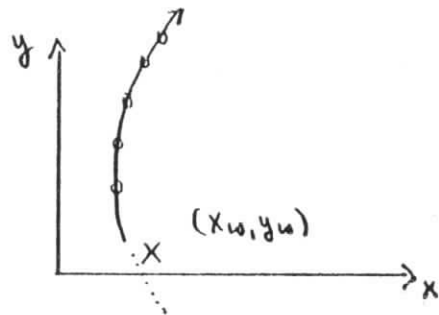
Dabei soll $Q = (\vec{y} - \vec{y}_m)^T W (\vec{y} - \vec{y}_m)$ minimal sein

- Ziel: (1) Ausgleich der Messwerte
 (Verbesserte Genauigkeit durch Bedingungsgl.)
 (2) Test des Modells (Hypothesentest)

Probleme dieser Art in der Hochenergiephysik häufig

1. Beispiel: (Vgl. Beispiel in Kap. 4)

Spurparameter ρ_0, α, r_0 durch Anpassung von Kreisen an Spurenpunkte in Ebene $\perp \vec{\beta}$



Parameter $\begin{pmatrix} \rho_0 \\ \alpha \\ r_0 \end{pmatrix}$

mit Kovarianzmatrix $V = \begin{pmatrix} \dots \\ \dots \\ \dots \end{pmatrix}$

Kreisgleichung:

$$\alpha (r_i^2 + r_0^2) - 2r_i (\lambda + \alpha r_0) \cos(\varphi_i - \varphi_0) + 2r_0 = 0$$

Die Koordinaten x_w, y_w des Ursprungs der Spur (Wechselwirkungspunkt) seien sehr genau bekannt. Wie kann diese Information benutzt werden, um die Spurparameter ρ_0, α, r_0 zu verbessern?

$$\left. \begin{aligned} x_w &= r_i \cos \varphi_i \\ y_w &= r_i \sin \varphi_i \end{aligned} \right\} \text{ in Kreisgleichung einsetzen}$$

$$\alpha (x_w^2 + y_w^2 + r_0^2) - 2(\lambda + \alpha r_0) (x_w \cos \varphi_0 + y_w \sin \varphi_0) + 2r_0 = 0$$

Wenn die Spur mit den Parametern ρ_0, α, r_0 genau durch den Wechselwirkungspunkt (x_w, y_w) geht, ist diese Gleichung recht erfüllt.

→ als Zwangsbedingung benutzen, ergibt Korrekturen an den Parametern = verbesserte Impuls- und Winkelbestimmung.

2. Beispiel: $\pi^0 \rightarrow \gamma \gamma$

Messung von Energie und Richtung des π^0 durch Messung der beiden Zerfallsphotonen
 Schauerzähler



Messung der Photonen
 $\gamma_1: E_1, \theta_1, \phi_1$
 $\gamma_2: E_2, \theta_2, \phi_2$

Wie bestimmt man für das π^0 Energie und Richtung und testet gleichzeitig die Hypothese?

Vierer-Vektor allgemein: $P = (E, \vec{p})$

↳ Impulsvektor
 ↳ Energie

$$P^2 = E^2 - \vec{p}^2 = m^2$$

↳ Masse des Teilchens

(alles in gleichen Einheiten, z.B. GeV, mit $c=1$)

Vierer-Vektor des Photons: $m_\gamma^2 = 0$

$$P = (E, E \cdot \vec{u})$$

↳ Einheitsvektor (Richtung)

Vierer-Vektor des Photons:
$$P = \begin{pmatrix} E \\ E \sin \theta \cos \phi \\ E \sin \theta \sin \phi \\ E \cos \theta \end{pmatrix}$$

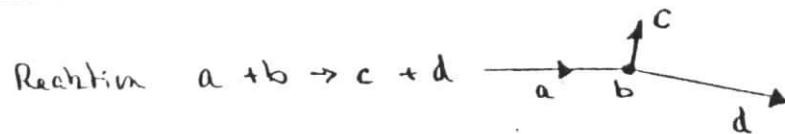
Wenn die Photonen γ_1 und γ_2 aus dem Zerfall des π^0 kommen, muß gelten

$$(P_1 + P_2)^2 - m^2 \pi^0 = 0$$

mit $(P_1 + P_2) = \begin{pmatrix} E_1 + E_2 \\ E_1 \sin \theta_1 \cos \phi_1 + E_2 \sin \theta_2 \cos \phi_2 \\ E_1 \sin \theta_1 \sin \phi_1 + E_2 \sin \theta_2 \sin \phi_2 \\ E_1 \cos \theta_1 + E_2 \cos \theta_2 \end{pmatrix}$

→ Zwangsbedingung für die Photonparameter E, ϕ, θ ergibt genaue Werte für π^0 -Parameter (E, ϕ, θ) und Möglichkeit zum Test (χ^2 -Test)

3. Beispiel:



Massen der Teilchen a und b bekannt

Hypothese über Massen der Teilchen c und d

Zwangsbedingungen: $E_a + E_b = E_c + E_d$

$$\vec{P}_a + \vec{P}_b = \vec{P}_c + \vec{P}_d$$

(4 Gleichungen)

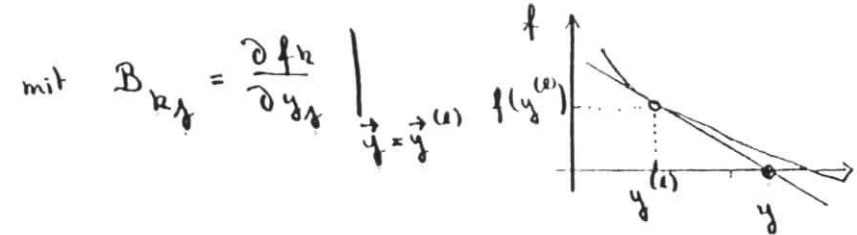
→ genauere Werte der Impulse + Test der Hypothese

Allgemeine Lösung des Problems:

Linearisierung der Bedingungengleichungen

$$\vec{y}^{(k)} = \text{letzte Näherungswert für } \vec{y}$$

$$\Delta f(\vec{y}) \approx f(\vec{y}^{(k)}) + B(\vec{y} - \vec{y}^{(k)})$$



$\vec{\lambda}$ = Vektor von m Lagrange-Multiplikatoren.

Definition von Q:

$$Q(\vec{y}, \vec{\lambda}) = (\vec{y} - \vec{y}_m)^T W (\vec{y} - \vec{y}_m) + 2 \vec{\lambda}^T (f + B(\vec{y} - \vec{y}^{(k)}))$$

Notwendige Bedingung für Minimum von Q ist Verschwinden der Ableitungen (nach \vec{y} und $\vec{\lambda}$):

$$\left| \begin{array}{l} \frac{1}{2} \frac{\partial Q}{\partial \vec{y}} = W(\vec{y} - \vec{y}_m) + B^T \vec{\lambda} \\ \frac{1}{2} \frac{\partial Q}{\partial \vec{\lambda}} = B(\vec{y} - \vec{y}^{(k)}) + f \end{array} \right.$$

Nullsetzen der Ableitungen liefert gehoppelt Gleichungssystem für \vec{y} und λ :

$$\begin{cases} W \vec{y}^{(k+1)} + B^T \vec{\lambda} &= W \vec{y}_m \\ B \vec{y}^{(k+1)} &= -\vec{f} + B \vec{y}^{(k)} \end{cases}$$

als Matrixgleichung schreiben.

$$\begin{pmatrix} W & B^T \\ B & \emptyset \end{pmatrix} \begin{pmatrix} \vec{y}^{(k+1)} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} W \vec{y}_m \\ -\vec{f} + B \vec{y}^{(k)} \end{pmatrix}$$

Für die inverse Matrix gilt

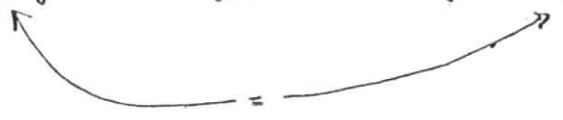
$$\begin{pmatrix} W & B^T \\ B & \emptyset \end{pmatrix}^{-1} = \begin{pmatrix} P & F^T \\ F & R \end{pmatrix}$$

$$\left. \begin{aligned} WP + B^T F &= I \\ WF^T + B^T R &= \emptyset \\ BP &= B \\ BF^T &= I \end{aligned} \right\} \begin{aligned} F &= (BW^{-1}B^T)^{-1} BW^{-1} \\ P &= W^{-1} - W^{-1}B^T(BW^{-1}B^T)^{-1}BW^{-1} \\ R &= -(BW^{-1}B^T)^{-1} \end{aligned}$$

Lösen des Gleichungssystems ergibt $\vec{y}^{(k+1)}$
 $l=1,2,3,\dots$ $\vec{y}^{(l)} \rightarrow \hat{\vec{y}}$

Eigenschaften:

$$(1) E(\hat{\vec{y}}) = E(\vec{y}_m) = E(\vec{y}) = \vec{y}^0$$



Beweis:

$$\hat{\vec{y}} = PW \vec{y}_m + F^T (-\vec{f} + B \vec{y}^{(k)})$$

$$\begin{aligned} E(\hat{\vec{y}}) &= PW E(\vec{y}_m) + F^T B E(\vec{y}^{(k)}) - F^T E(\vec{f}) \\ &= (PW + F^T B) E(\vec{y}) = E(\vec{y}) \cdot \underbrace{0}_{=I} \\ &= \vec{y}^0 \end{aligned}$$

Weitere Eigenschaften siehe unten.

Umformulieren:

$$\vec{y}^{(k)} = \vec{y}_m + \Delta \vec{y}^{(k)}$$

$$\begin{pmatrix} W & B^T \\ B & \emptyset \end{pmatrix} \begin{pmatrix} \Delta \vec{y}^{(k+1)} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} \emptyset \\ -\vec{f}(\vec{y}^{(k)}) + B \Delta \vec{y}^{(k)} \end{pmatrix}$$

$$\begin{pmatrix} \Delta \vec{y}^{(k+1)} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} P & F^T \\ F & R \end{pmatrix} \begin{pmatrix} \emptyset \\ -\vec{f}(\vec{y}^{(k)}) + B \Delta \vec{y}^{(k)} - B \vec{y}_m \end{pmatrix}$$

Weitere Eigenschaften:

$$(2) V(\Delta \vec{y}) = W^{-1} - P$$

Beweis:

$$\Delta \vec{y} = F^T (-\vec{f}(\vec{y}^{(k)}) + B \vec{y}^{(k)} - B \vec{y}_m)$$

$$\begin{aligned} V(\Delta \vec{y}) &= F^T B \cdot V(\vec{y}_m) \cdot B^T F \\ &= \underbrace{F^T B W^{-1} B^T F}_{(1-PW)} \cdot \underbrace{B^T F}_{(1-WP)} \\ &= W^{-1} + PW W^{-1} WP - 2PW W^{-1} \end{aligned}$$

(3) $V(\hat{\vec{y}}) = P$

Beweis: $\vec{y} = \vec{y}_m + \Delta y$
 $= (1 - F^T B) \vec{y}_m + F^T (-\frac{1}{f}(\vec{y}^{(1)}) + B \vec{y}^{(2)})$

$V(\hat{\vec{y}}) = \underbrace{(1 - F^T B)}_{PW} V(\vec{y}_m) \underbrace{(1 - F^T B)^T}_{WP}$
 $= PWP = P(1 - B^T F)$
 $= P - \underbrace{P B^T}_{=0} F = P$

• (4) $V(\hat{\lambda}) = -R$ (ohne Beweis)

• (5) $E(Q) = m$
 $L = \text{Zahl der Bedingungsbedingungen}$

Bei normalverteilten Messwerten (und genau genommen: linearen Zwangsbedingungen) ist Q verteilt nach χ^2 mit m Freiheitsgraden
 \Rightarrow statistischer Test möglich

Weitere statistischer Test einzelner Variablen

$\Delta y = \hat{\vec{y}} - \vec{y}_m$
 $V(\Delta \vec{y}) = V(\vec{y}_m) - V(\hat{\vec{y}})$

Definition:
(Pull_i) = $\frac{(\Delta y)_i}{V(\Delta y)_{ii}}$ nullte $N(0,1)$ verteilt sein
Mittelwert 0, Varianz 1

Diese Aussage kann für jede Variable einzeln geprüft werden. (Verteilungen der Pulls anschauen!)

- liegt der Mittelwert nicht bei 0
 \Rightarrow einseitiger Zug der Bedingungsbedingungen = systematische Fehler in der Messung der i-ten Komponente von \vec{y}_m (oder falsche Hypo)
- ist die Varianz > 1
 \Rightarrow Fehler (σ) der i-ten Komponente von \vec{y}_m war zu klein (oder falsche Hypothese)

Allerdings werden durch systematischen Fehler bzw. falsch Element der Kovarianzmatrix für eine Komponente \rightarrow alle anderen Pulls beeinflusst (über die Zwangsbedingungen)

Die $n \times n$ Matrix $P (= V(\hat{\vec{y}}))$ ist singulär, mit Rangdefekt m , d.h. $\text{Rang}(P) = n - m$. Dies wird durch die m Bedingungsbedingungen zwischen den Variablen verursacht.

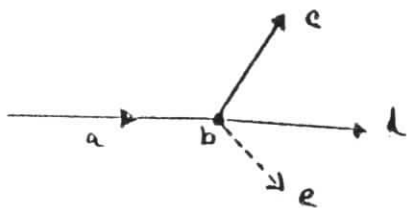
\Rightarrow Transformation auf $(n-m)$ Variable möglich, deren Kovarianzmatrix nichtsingulär ist.

Ungemessene Variable:

Bei vielen Problemen treten in den Bedingungsgleichungen neben den gemessenen Variablen \vec{y} weitere Variable \vec{x} auf, die nicht gemessen sind, jedoch durch die Bedingungsgleichungen bestimmt werden können.

4. Beispiel

Reaktion $a + b \rightarrow c + d + e$



Impulse der Teilchen a, b, c, d gemessen
Impuls von Teilchen e ungemessen

Die 4 Bedingungsgleichungen

$$E_a + E_b = E_c + E_d + E_e$$
$$\vec{p}_a + \vec{p}_b = \vec{p}_c + \vec{p}_d + \vec{p}_e$$

bilden (bei Hypothesen über Massen m_a, m_b, \dots, m_e) immer noch ein (1-fach) überbestimmtes Gleichungssystem

Zwangsbedingungen $f(\vec{y}, \vec{x})$

Für die nicht gemessenen Variablen \vec{x} können (aus geeigneten Zwangsbedingungen) grobe Startwerte bestimmt werden: \vec{x}_s

Linearisierung der Zwangsbedingungen

$$f(\vec{y}, \vec{x}) \approx f(\vec{y}^{(0)}, \vec{x}^{(0)}) + B(\vec{y} - \vec{y}^{(0)}) + A(\vec{x} - \vec{x}^{(0)})$$

Um \vec{x} erweitert Formalismus der Bestimmung von \vec{y}, \vec{x}

$$Q(\vec{y}, \vec{x}, \vec{\lambda}) = (\vec{y} - \vec{y}_m)^T W (\vec{y} - \vec{y}_m) + 2 \vec{\lambda}^T (f + B(\vec{y} - \vec{y}^{(0)}) + A(\vec{x} - \vec{x}^{(0)}))$$

$$\frac{1}{2} \frac{\partial Q}{\partial \vec{y}} = W(\vec{y} - \vec{y}_m) + B^T \vec{\lambda}$$

$$\frac{1}{2} \frac{\partial Q}{\partial \vec{x}} = A^T \vec{\lambda}$$

$$\frac{1}{2} \frac{\partial Q}{\partial \vec{\lambda}} = B(\vec{y} - \vec{y}^{(0)}) + A(\vec{x} - \vec{x}^{(0)}) + f$$

Mit $\Delta \vec{y} = \vec{y} - \vec{y}_m$
 $\Delta \vec{x} = \vec{x} - \vec{x}_s$

Laufen die linearen Gleichungen (Nullsetzen der Able.) in Matrixform

$$\begin{pmatrix} W & B & B^T \\ B & B & A^T \\ B & A & B \end{pmatrix} \begin{pmatrix} \vec{\Delta y} \\ \vec{\Delta x} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ B \\ -\vec{f} + B \vec{\Delta y}^{(1)} + A \vec{\Delta x}^{(1)} \end{pmatrix}$$

Dieses Gleichungssystem kann im Prinzip in dieser Form durch Matrixinversion gelöst werden (Inversionsprogramm mit Pivotwahl, Matrix ist symmetrisch).

In der Praxis ist jedoch die Unterscheidung zwischen gemessenen und ungemessenen Variablen lästig.

Vgl. 2. Beispiel (m⁰ → y y)

Es kann häufig vorkommen, daß eine der Variablen E, θ, φ für die Photonen nicht gemessen ist. Gewünscht ist möglichst geringe Änderung eines Rechenprogramms für diesen Fall.

Einfachste Methode:

Zwischen gemessenen Variablen \vec{y}_m und nichtgemessenen Variablen \vec{x} nicht unterscheiden, d.h. alle als \vec{y}_m bezeichnen, jedoch in der Gewichtsmatrix W, alle Elemente, die sich auf nichtgemessene Variable beziehen, gleich Null setzen ("Gewicht" = 0 für diese Variable)

Dies ergibt:

$$\begin{pmatrix} W & B^T \\ B & B \end{pmatrix} \begin{pmatrix} \vec{\Delta y} \\ \vec{x} \end{pmatrix} = \begin{pmatrix} 0 \\ -\vec{f} + B \vec{\Delta y}^{(1)} \end{pmatrix}$$

mit z.B.

$$W = \begin{pmatrix} w_{11} & 0 & w_{13} \\ 0 & 0 & 0 \\ w_{31} & 0 & w_{33} \end{pmatrix}$$

Die Matrix $\begin{pmatrix} W & B^T \\ B & B \end{pmatrix}$ ist nichtsingulär, und kann invertiert werden, wenn das Problem sinnvoll ist, d.h. wenn die nichtgemessenen Variablen durch die Zwangsbedingungen festgelegt werden.

Dann gilt:

$$\begin{aligned} E(Q) &= \text{Anzahl Zwangsbedingungen (m)} \\ &\quad - \text{Anzahl nicht gemessener Variablen (q)} \\ &= \underline{k = m - q} \end{aligned}$$

bei normalverteilten Messwerten ist

Q χ^2 verteilt mit $k = m - q$ Freiheitsgraden

man spricht von k C - Fit, z.B.

Beispiel 3: 4 C - Fit ohne Teilchen e
1 C - Fit mit Teilchen e

7. Programm - Package CONLES

Für die direkte Ausgleichung mit nichtlinearen Zwangsbedingungen gibt es ein Programm - Paket CONLES.

Dies enthält alle

Matrixoperationen
Konvergenztest
etc.

Zur Anwendung muß nur die für das behandelte Problem spezifische Rechnung programmiert werden.

(maximal 50 Parameter, maximal 20 Zwangsbed.)

Programm - Schema (Skizze):

```
COMMON / / NY, NF, Y ( ), DELT ( ), F(20)
```

```
CALL CONPAR (nr)
    ↳ definiert Problemnr 1...10 für Statistik
```

NY = Zahl der Variablen y

NF = Zahl der Zwangsbedingungen

```
CALL CONLES (1)
```

Y () = Kovarianzmatrix der y, Elemente von ungemessenen Variablen = 0

```
CALL CONLES (2)
```

für ungemessene Variable Startwerte abschätzen

falls numerische Bestimmung der Matrix B gewünscht, Schrittweiten für numerische Diff. definieren

1 Addieren von Korrekturen DELT(J) zu den Meßwerten
↳ ist zu Beginn = 0

F () = ... Berechnen der Zwangsbedingungen

```
CALL CONDER (IREP)
IF (IREP.NE.0) GOTO 1
```

nur, wenn numerische Ableitungen berechnet werden

```
CALL CONCHK (IBR)
IF (IBR.NE.0) GOTO (1, 2, 3), IBR
```

Konvergenztest

```
IF (IBR.NE.0) GOTO (1, 2, 3), IBR
```

↳ keine Konvergenz
↳ Konvergenz erreicht

B () = ... Berechnen der Matrix der Ableitungen
nur wenn nicht numerische Able. berechnet werden

```
CALL CONLES (3)
```

Berechnet neue Korrekturen

```
GOTO 1
```

2 CALL CONLES (4)

Konvergenz erreicht, berechne Kovarianzmatrix der \vec{y} , Pullo etc...

3

Keine Konvergenz möglich

Größtes Praktisches Problem:

Berechnung der NY * NF Ableitungen

z.B. NY = 50, NF = 20

⇒ 1000 Ableitungen

Wenn nur eine der Ableitungen falsch ist, kann es sein, daß trotzdem Konvergenz erfolgt, aber langsam.

⇒ CONLES bietet die Möglichkeit, Ableitungen numerisch zu berechnen. (Rechenzeit größer)

⇒ Programm-Entwicklung

- 1. Testphase: numerische Berechnung
evtl. Verändern der Zwangsbedingungen
- 2. Testphase: analytische Berechnung der Ableitungen
zusätzlich programmieren.
Ausdruck der Matrizen aus numerischer und analytischer Ableitung ⇒ Vergleich
- 3. Produktionsphase: keine numerische Berechnung der Ableitungen, nur analytisch

Anwendung auf Beispiel 2 : $\pi^0 \rightarrow \gamma\gamma$

Gemessene Variable $\left. \begin{array}{l} E_1, \theta_1, \phi_1 \\ E_2, \theta_2, \phi_2 \end{array} \right\} NY = 6$

keine ungemessene Variablen

Zwangsbed. $f(E_1, \theta_1, \phi_1, E_2, \theta_2, \phi_2) - m^2_{\pi^0} = 0$

= 1 C-Fit
NF = 1

Nach erfolgreicher Konvergenz kann der Vektor $\vec{p}(\pi^0)$ durch Variablentransformation berechnet werden; für die Kovarianzmatrix $V(\vec{p}_{\pi^0})$ ist dies aufwendig (wieder viele Ableitungen programmieren).

Andere Möglichkeit der Formulierung des Problems:
mit gleichen Ergebnis

Weitere ungemessene Variablen einführen:

$\left. \begin{array}{l} p_x \\ p_y \\ p_z \\ E \end{array} \right\} \text{ für } \pi^0 \quad NY = 10$
(davon 6 gem.)

Weitere Zwangsbedingungen:

$$p_x - (E_1 \sin \theta_1 \cos \phi_1 + E_2 \sin \theta_2 \cos \phi_2) = 0$$

$$p_y - (E_1 \sin \theta_1 \sin \phi_1 + E_2 \sin \theta_2 \sin \phi_2) = 0$$

$$p_z - (E_1 \cos \theta_1 + E_2 \cos \theta_2) = 0$$

$$E - (E_1 + E_2) = 0$$

NF = 5

+ Zwangsbedingung

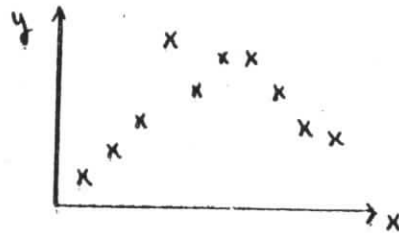
$$E^2 - p_x^2 - p_y^2 - p_z^2 - m^2_{\pi^0} = 0$$

⇒ nach erfolgreicher Konvergenz ist auch die Kovarianzmatrix $V(\vec{p}_{\pi^0})$ bereits vorhanden.

D. Interpolation und Approximation

1.) Glätten von Daten

Messdaten y_i, x_i
 $i = 1, \dots, n$
 liegen vor



Gesucht: interpolierende Funktion.

→ 1. Fall: Parametrisierung der Abhängigkeit $y = y(x)$
 sei bekannt: $f(x, a_1, \dots, a_m)$

→ mit Methode der kleinsten Quadrate $f(x, a_1, \dots)$
 an die Daten anpassen.

Probleme: (insbesondere bei automatisierter Auswertung)
 mögliche Ausreißer mit extrem großen Abweichungen verfälschen das Ergebnis

Häufig bewährte Methode:

Anpassung mit allen Daten y_i
 Größte Abweichung Δy herausuchen
 wenn $|\Delta y| >$ Schranke, Punkt weglassen
 und Anpassung wiederholen etc.

kann bei extremen Ausreißern sparen, u.U. ineffektiv.

→ 2. Fall: Parametrisierung $f(x)$ unbekannt

Probleme: Messfehler
 Ausreißer
 unbekannte Parametrisierung

→ entweder: y_i durch geglättete Punkte, z_i
 ersetzen und zwischen diesen Punkten
 interpolieren

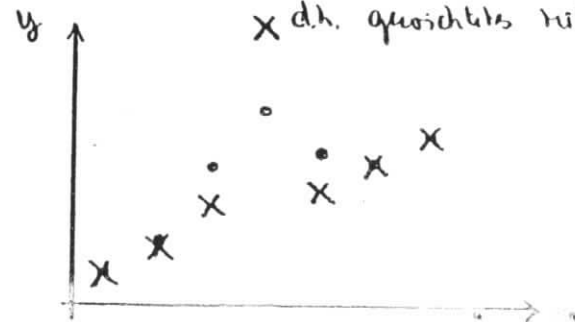
oder: (zusätzlich) Parametrisierung annehmen
 z.B. Polynom m -ten Grades, und
 Polynomgrad optimal wählen

In beiden Fällen kann es sinnvoll sein, die Daten
 zuerst zu "glätten", mit dem Ziel, vor allem Ausreißer
 zu "dämpfen" und statistische Schwankungen zu
 reduzieren.

Mögliche Methode: y_i durch z_i ersetzen mit

$$z_i = \frac{y_{i-1} + 2y_i + y_{i+1}}{4} \quad \text{Methode H (Hanning)}$$

x d.h. gewichtetes Mittel

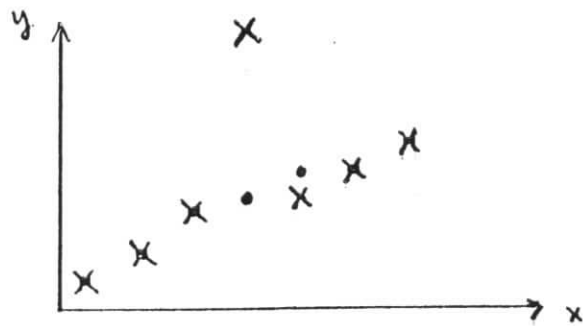


⇒ extreme Abweichungen werden nur wenig gedämpft,
sie wirken sich auch auf die Nachbarn aus.

Andere Möglichkeit: M3

Jeden Wert ersetzen durch den mittleren der
drei Werte y_{i-1}, y_i, y_{i+1}

d.h. $z_i = \text{median}(y_{i-1}, y_i, y_{i+1})$



⇒ der Einfluss einzelner extremer Abweichungen verschwindet
praktisch völlig.

Verfeinerung dieser Methode s.u.

u.U. ist es sinnvoll, vor dem Glätten die Daten
so zu transformieren, dass alle y_i etwa gleiche
statistische Fehler haben.

z.B. bei $y_i = \text{Anzahl Einträge im Histogramm-Interv.}$
(Poisson-Verteilung)

Transformation:

$$z_i = \sqrt{y_i} + \sqrt{y_i + 1}$$

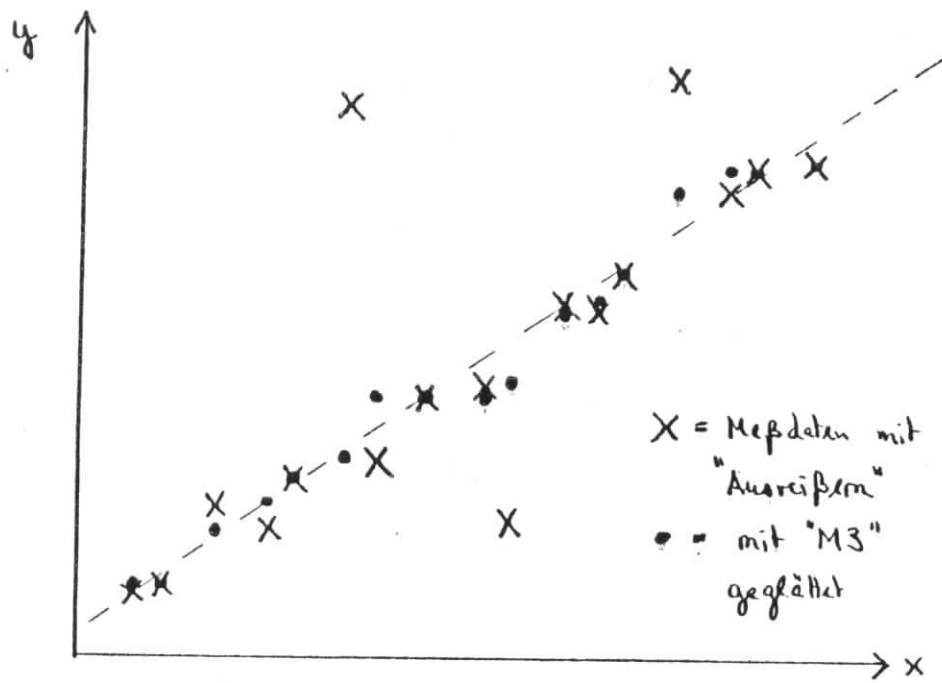
(z_i hat Varianz 1)

$$y_i = \left(\frac{z_i^2 - 1}{2z_i} \right)^2 \quad (\text{Rücktransformation})$$

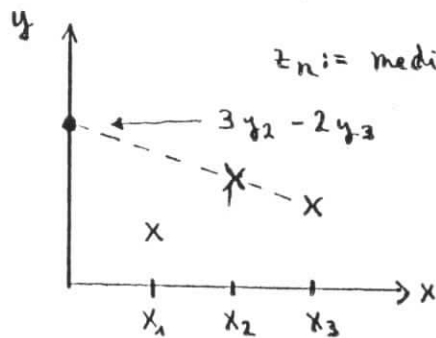
Beispiel für Anwendung von M3:

Durch die Daten (y_i, x_i) ist eine Gerade zu
legen, einzelne Ausreißer sind möglich.

- ① Daten mit M3 glätten ($y_i \rightarrow z_i$)
- ② Durch geglättete Daten z_i Gerade legen.
- ③ Stärke von Gerade abweichende Daten y_i weglassen
und mit verbleibenden Daten wiederum
Gerade anpassen.



Methode M3: $z_i = \text{median}(y_{i-1}, y_i, y_{i+1})$
 $z_1 = \text{median}(y_1, y_2, 3y_2 - 2y_3)$
 $z_n = \text{median}(3y_{n-2} - 2y_{n-1}, y_{n-1}, y_n)$



Eigenschaften: y_i bleibt unverändert, wenn $y_{i-1} < y_i < y_{i+1}$ oder $y_{i-1} > y_i > y_{i+1}$
 • einzelne Ausreißer verschwinden

Verfeinerung der Methode:

Kombinieren verschiedener M-Methoden und der H-Methode

• M3: $z_i = \text{median}(y_{i-1}, y_i, y_{i+1})$
 $z_1 = \text{median}(y_1, y_2, 3y_2 - 2y_3)$
 $z_n = \text{median}(3y_{n-2} - 2y_{n-1}, y_{n-1}, y_n)$

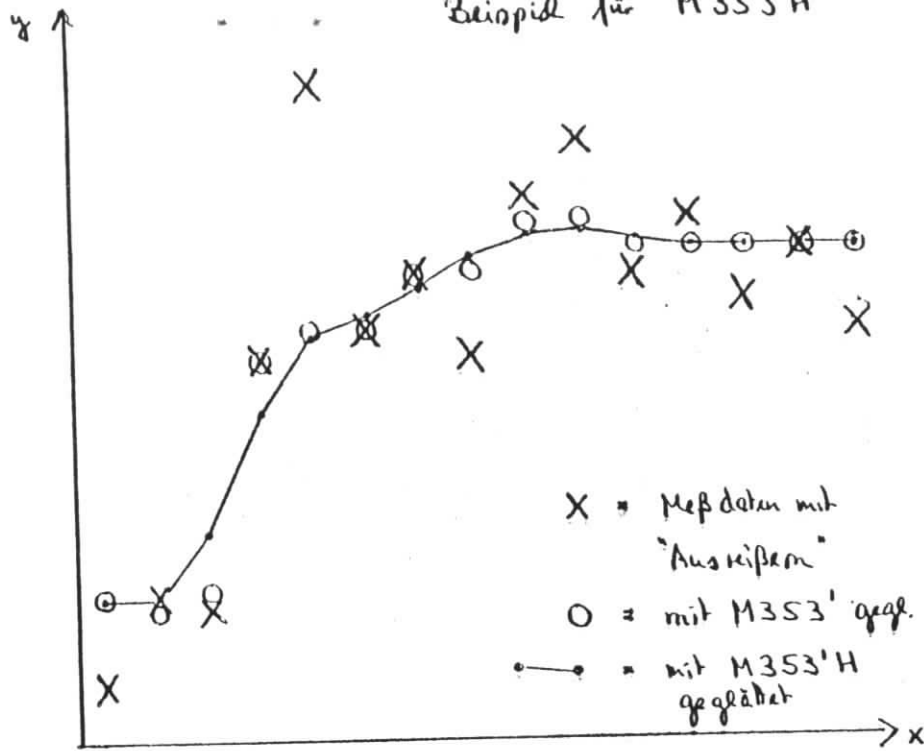
• M5: $z_i = \text{median}(y_{i-2}, y_{i-1}, y_i, y_{i+1}, y_{i+2})$
 $z_1 = y_1$
 $z_2 = \text{median}(y_1, y_2, y_3)$
 $z_{n-1} = \text{median}(y_{n-2}, y_{n-1}, y_n)$
 $z_n = y_n$

• M3': $z_i = \text{median}(y_{i-1}, y_i, y_{i+1})$
 $z_1 = y_1$
 $z_n = y_n$

• H: $z_i = (y_{i-1} + 2y_i + y_{i+1}) / 4$
 $z_1 = y_1$
 $z_n = y_n$

• Kombinieren: $M3 \ 5 \ 3' \ H = M3; M5; M3'; H$
nacheinander
 oder $(M3 \ 5 \ 3' \ H)^2$

Beispiel für M3S3'H



2.) Ausgleichende Interpolation mit orthogonalen Polynomen

Ziel: Anpassung von Polynom m -ten Grades
 an Daten $y_i, x_i \quad i=1, \dots, m$
 mit Standardabw. σ_i
 Grad m optimal bestimmen

m zu klein \rightarrow zu starke Abweichungen der y_i ,
 aber glatt

m zu groß \rightarrow kleine Abw. der y_i , aber

Die beste Methode zur Erreichung dieses Ziels ist
 die Benutzung orthogonaler Polynome (+ Maß der kleinste

\rightarrow numerisch stabil

\rightarrow statistische Methoden zur Bestimmung des optimalen
 Polynom-Grades

Ansatz: $f(x) = a_1 p_1(x) + a_2 p_2(x) + \dots$

mit $p_1(x) = b_{11}$
 $p_2(x) = b_{12} + b_{22}x$
 $p_3(x) = b_{13} + b_{23}x + b_{33}x^2$

Orthogonalität der $p_j(x), p_k(x)$:

$$\sum_{i=1}^n w_i p_j(x_i) \cdot p_k(x_i) = \delta_{jk}$$

$w_i = \frac{1}{\sigma_i^2}$ Die Polynome hängen also von den speziellen Werten x_i und den σ_i ab.

Die Polynome $p_j(x)$ sind durch die Orthogonalitätsbed. definiert.

Konstruktion der ersten beiden Polynome:

durch Einsetzen:

$$b_{11}^2 \sum_{i=1}^n w_i = 1$$

$$b_{22}^2 \left(\sum_{i=1}^n w_i x_i^2 - \frac{(\sum w_i x_i)^2}{\sum w_i} \right) = 1$$

$$b_{12} = -b_{22} \frac{\sum w_i x_i}{\sum w_i}$$

$\Rightarrow b_{11}, b_{12}, b_{22}$ berechnen

Konstruktion der weiteren Polynome durch Rekursionsformel:

$$\lambda p_k(x) = (x - \alpha) p_{k-1}(x) - \beta p_{k-2}(x)$$

$x = x_i$ setzen, mit $w_i p_{k-1}(x_i)$ multiplizieren und über alle i summieren:

Linke Seite:

$$\lambda \sum_i w_i p_k(x_i) p_{k-1}(x_i) = 0$$

Rechte Seite:

$$\sum_i w_i x_i p_{k-1}^2(x_i) - \underbrace{\alpha \sum_i w_i p_{k-1}(x_i)}_{=1} - \beta \underbrace{\sum_i w_i p_{k-1}(x_i) p_{k-2}(x_i)}_{=0} = 0$$

$$\Rightarrow \alpha = \sum_i w_i x_i p_{k-1}^2(x_i)$$

Entsprechend erhält man β durch Multiplikation des Ansatzes mit $w_i p_{k-2}(x_i)$, Summieren über alle i :

$$\sum x_i w_i p_{k-1}(x_i) p_{k-2}(x_i) - \beta \underbrace{\sum w_i p_{k-2}^2(x_i)}_{=1} = 0$$

$$\Rightarrow \beta = \sum x_i w_i p_{k-1}(x_i) p_{k-2}(x_i)$$

λ ergibt sich aus der Normierungsvorgabe:

$$\sum w_i p_k^2(x_i) = 1$$

Benutzt man einen solchen Ansatz mit orthogonalem Polynom bei der Methode der kleinsten Quadrate, hat man z.B. bei 3 Parametern eine Matrix $C = (A^T A)$ (vgl. Kapitel: Methode d. kleinsten Qu.)

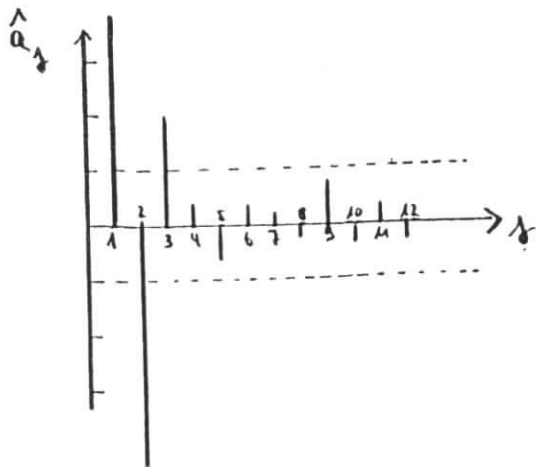
$$\rightarrow C = \begin{pmatrix} \sum w_i p_1^2 & \sum w_i p_1 p_2 & \sum w_i p_1 p_3 \\ \sum w_i p_1 p_2 & \sum w_i p_2^2 & \sum w_i p_2 p_3 \\ \sum w_i p_1 p_3 & \sum w_i p_2 p_3 & \sum w_i p_3^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

und daher die Parameterwerte

$$\hat{a}_j = \sum_{i=1}^n w_i p_j(x_i)$$

\Rightarrow durch einzelne Summenbildung, bei \hat{a}_j wird nur $p_j(x)$ benutzt, also alle \hat{a}_j unabhängig voneinander berechnbar mit Kovarianzmatrix $C^{-1} = I$, d.h. alle Standardabweichungen der \hat{a}_j sind 1.

\Rightarrow die einzelnen \hat{a}_j können nacheinander in der Reihenfolge $\hat{a}_1, \hat{a}_2, \dots$ bestimmt werden.



\Rightarrow Koeffizienten $\hat{a}_1, \hat{a}_2, \hat{a}_3$ signifikant $\neq 0$
alle \hat{a}_j mit $j \geq 4$ mit Null verträglich ("Reinschen")

\Rightarrow einfacher Test zur Bestimmung des maximal notwendigen Polynom-Grades:

Wenn $a_j^0 = 0 \Rightarrow |\hat{a}_j|^2 \chi^2$ verteilt mit einem Freiheitsgrad.

Also: alle Koeffizienten berechnen, die sicher ausreichen
alle Koeffizienten streichen, vom höchsten Grad beginnend, bis $|\hat{a}_j|^2 > \chi^2_{\beta}(1)$

Dann zur bequemeren Anwendung u.U. in gewöhnliches Polynom umrechnen.

Weitere Eigenschaften:

Summe der quadratischen Abweichungen:

$$Q = \sum_{i=1}^n w_i y_i^2 - \sum_{j=1}^m \hat{a}_j^2$$

Beim Übergang von m Parametern zu $(m+1)$ Parametern wird Q um \hat{a}_{m+1}^2 verringert.

Bei normalverteilten Daten ist Q χ^2 verteilt mit $(n - m)$ Freiheitsgraden.

Auch wenn die w_i nicht bekannt sind, kann die optimale Polynomordnung mit statistischen Methoden bestimmt werden (F-Test)

\rightarrow Durchführung mit Programmen ORT(OF) etc.

3.) Spline - Funktionen

Problem: n Punkte (y_i, x_i) exakt interpolierend.

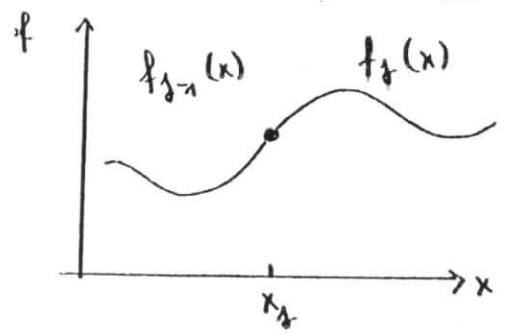
Polynom $(n-1)$ ter Ordnung möglich, wird aber i.a. stark oszillieren

i.a. besser: kubische Spline-Funktion

= Polynom 3. Ordnung

$$f_j(x) = a_1 + a_2(x-x_j) + a_3(x-x_j)^2 + a_4(x-x_j)^3$$

für jedes Intervall x_j bis x_{j+1} .



Bedingungen:

$$\left| \begin{array}{l} f_{j-1}(x_j) = f_j(x_j) = y_j \\ f'_{j-1}(x_j) = f'_j(x_j) \\ f''_{j-1}(x_j) = f''_j(x_j) \end{array} \right.$$

Diese Bedingungen legen die Koeffizienten für alle Intervalle noch nicht fest

↳ 2 weitere Bedingungen

z.B. $f_1''(x_1) = f_{n-1}''(x_n) = 0$

Diese spez. Wahl definiert die sogenannten natürlichen Splines, diese erfüllen die Bedingung

$$\int_{x_1}^{x_n} [f''(x)]^2 dx = \text{Minimum bzgl. aller zweimal stetig diffbaren Funktionen}$$

Vergleiche: elastisches Lineal, an n Punkten fest eingepinnt (Anfangspunkt a , Endpunkt b)

Biegungsenergie $\int_a^b \frac{[f''(x)]^2}{[1+f'(x)^2]^{5/2}} dx$



angenehmt Spline-Funktion minimiert Biegungsenergie

Anderer mögliche Bedingungen

- ① $f_1''(x_1) = c_a$ $f_{n-1}''(x_n) = c_b$
- ② $f'_1(x_1) = c_a$ $f'_{n-1}(x_n) = c_b$
- ③ $\frac{f''_1(x_1)}{f''_2(x_2)} = c_a$ $\frac{f''_{n-1}(x_n)}{f''_{n-2}(x_{n-1})} = c_b$

→ Programm CUBCOF etc...

Zur Bestimmung der Koeffizienten muß ein Gleichungssystem mit tridiagonaler symmetrischer Matrix gelöst werden (einfach und schnell)

Differentiation und Integration:

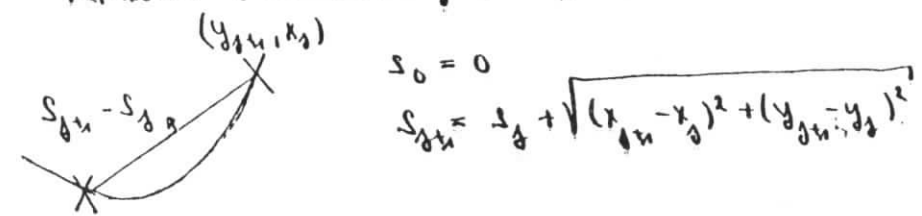
Die Spline-Fkt kann einfach differenziert bzw integriert werden, die dabei entstehenden Funktionswerte können wieder durch eine Spline-Funktion dargestellt werden

⇒ einfache Möglichkeit zur Integration und Diff. von Funktionen.

Interpolation in der x-y-Ebene

(speziell zum Zeichnen von Kurven per Computer)

x_i, y_i als Funktion eines monoton anwachsenden Parameters (Sehnenlänge) auffassen.



(x_j, y_j) Splinefunktionen $x = x(s)$
 $y = y(s)$

entspricht Parametrisierung einer Kurve mit Parameter s .

Literatur

- S. Brandt, Statistische Methoden der Datenanalyse, Bibliographisches Institut Mannheim (1968)
- S. Brandt, Statistical and Computational Methods in Data Analysis, North Holland (1976)
- J. Cramer, Mathematical Methods of Statistics, Princeton University Press (1946)
- C. de Boor, A Practical Guide to Splines, Springer (1978)
- W.T. Eadie, D. Drijard, F. E. James, M. Roos and B. Sadoulet, Statistical Methods in Experimental Physics, North Holland (1971)
- A.G. Frodesen, O. Skjeggstad and H. Tofte, Probability and Statistics in Particle Physics, Universitetsverlaget Bergen (1979)
- D.M. Himmelblau, Applied Nonlinear Programming, McGraw-Hill (1972)
- A.M. Mood and F.A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill (1963)
- J. Stoer and R. Bulirsch, Introduction to Numerical Analysis, Springer (1980)