Statistische Methoden der Datenanalyse

Wintersemester 2012/2013

Albert-Ludwigs-Universität Freiburg

Prof. Markus Schumacher, Dr. Stan Lai

Physikalisches Institut Westbau 2 OG

E-Mail: Markus.Schumacher@physik.uni-freiburg.de

stan.lai@cern.ch

http://terascale.physik.uni-freiburg.de/lehre/ws_1213/statmethoden_ws1213

Kapitel 7

Die Methode der kleinsten Quadrate

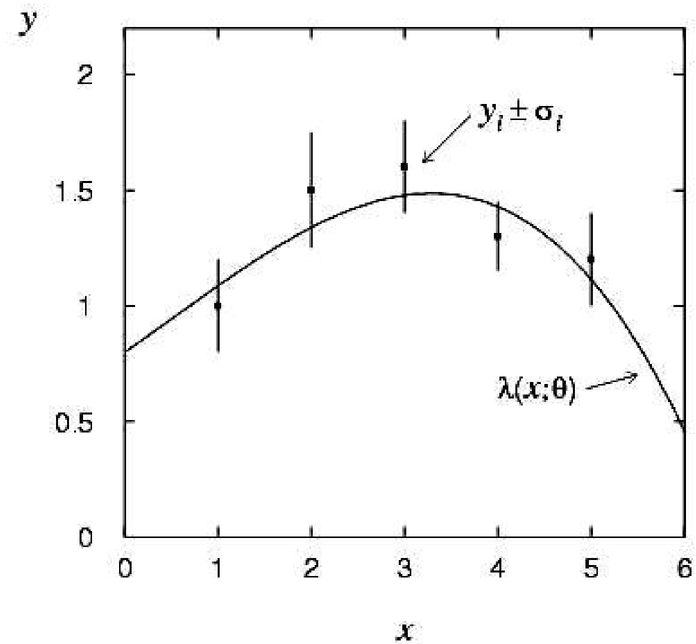
Die Method der kleinsten Quadrate KQ

Betrachte N unabhängige Messungen von Wertepaaren (x_i, y_i)

x_i perfekt bekannt

y_i mit Varianz $V[y_i] = \sigma_i^2$.

aus beliebiger WDF



Theorievorhersage für Zusammenhang λ zwischen x und y

$$E[y_i] = \lambda(x_i;\theta) \ .$$

Anzahl der Parameter θ_i < Anzahl der Messpaare

z.B: $y = s_0 + v_0 x + 0.5 ax^2$ Weg-Zeit-Gesetz x=Zeit

y Häufigkeit des Auftretens eines numerischen Wertes x

Das KQ-Prinzip

Bestimme Schätzwerte für die Parameter aus der Minimierung der folgenden Größe:

$$X^2 \equiv \sum_{i=1}^{N} w_i (y_i - f_i)^2$$

w_i = Gewicht für die i-te Beobachtung

Theoriezusammenhang hier: $f_i = f_i(\theta_1, \theta_2, \ldots, \theta_L; x_i)$

Wenn Varianz bekannt dann: $w_i = 1/\sigma_i^2 \quad X^2 = \sum_{i=1}^{N} \left(\frac{y_i - f_i}{\sigma_i}\right)^2$

Wenn Varianz unbekannt oder gleich dann setze $w_i = 1$: $X^2 = \sum_{i=1}^{N} (y_i - f_i)^2$

Wenn y_i = Zählrate, dann verwende erwartete Poisson-Varianz $\sigma_i^2 \approx f_i$

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{f_i}$$

Nicht erwartete Varianz $\sigma_i^2 \approx y_i$ liefert verzerrte Schätzer $X^2 = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{y_i}$

Zusammenhang mit ML-Methode

Betrachte N unabhängige Messungen von Wertepaaren (x_i, y_i)

x_i perfekt bekannt y_i mit Varianz $V[y_i] = \sigma_i^2$. nun aus Gauss-WDF

Theoriezusammenhang: $E[y_i] = \lambda(x_i;\theta)$.

Die Likelihoodfunktion ist dann gegeben durch:

$$L(\theta) = \prod_{i=1}^{N} f(y_i;\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - \lambda(x_i;\theta))^2}{2\sigma_i^2}\right]$$

und die log-Likelihoodfunktion ergibt sich zu:

$$\ln L(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\theta))^2}{\sigma_i^2} + \text{terms not depending on } \theta$$

Zusammenhang: ML- und KQ-Methoden

$$\ln L(\theta) = -\frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \lambda(x_i;\theta))^2}{\sigma_i^2} + \text{terms not depending on } \theta$$

Vergleich mit: $$\chi^2(\theta) = \sum_{i=1}^{N}\frac{(y_i - \lambda(x_i;\theta))^2}{\sigma_i^2}$$

Es gilt also: - ½ ln L = χ^2

und $\Delta \ln L = -1/2$ entspricht $\Delta\chi^2=1$

d.h. für gaussverteilte y_i gilt: ML-Schätzer = KQ-Schätzer

dann "erben" KQ-Schätzer die Eigenschaften von ML-Schätzern

wenn y_i nicht gaussverteilt, dann ist KQ eine "ad hoc" Methode.

Keine generelle Aussage über Eigenschaften von KQ-Schätzer

KQ für korrelierte Messungen

Wenn die y_i korreliert sind und ihre gemeinsame WDF durch eine Multidimensionale Gauss-Verteilung mit Kovarianzmatrix V gegeben ist:

$$g(\vec{y}, \vec{\lambda}, V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(\vec{y} - \vec{\lambda})^T V^{-1} (\vec{y} - \vec{\lambda})\right]$$

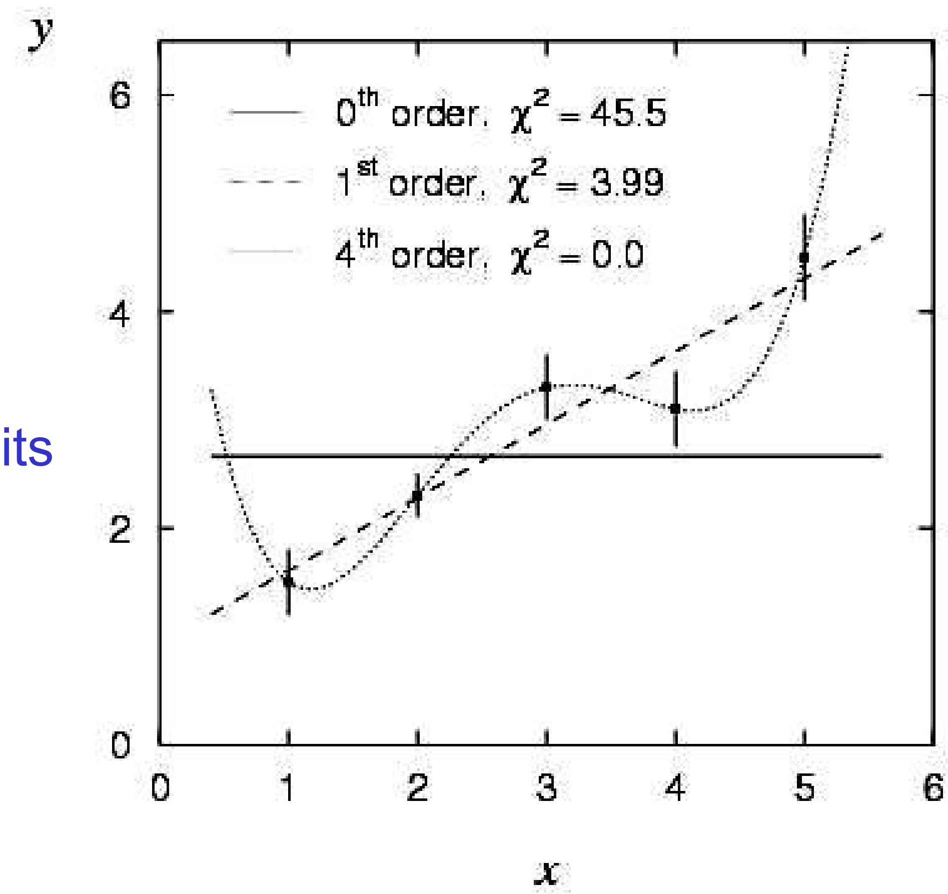
Dann ist die Maximierung der Likelihoodfunktion für korrelierte Messungen äquivalent zur Minimierung des folgenden Ausdrucks:

$$\chi^2(\vec{\theta}) = \sum_{i,j=1}^{N} (y_i - \lambda(x_i; \vec{\theta}))(V^{-1})_{ij}(y_j - \lambda(x_j; \vec{\theta}))$$

Beispiel für eine KQ-Anpassung

Anpassung eines Polynoms der Ordnung *p*:

$$\lambda(x;\theta_0,\ldots,\theta_p) = \sum_{n=0}^{p} \theta_n x^n$$



5 Messpunkte, p+1 Parameter

p=4: perfekte Anpassung
keine Aussage über Güte des Fits

p= 1 und p=0 im Vergleich

Für p=0 ist Anpassung schlecht,
großes χ^2 im Minimum

Für p=1: ist Anpassung ok
5 Messungen -2 Parameter = 3 Freiheitsgrade
χ^2 /FG = 1.33

Varianz für KQ-Schätzer

In den meisten in der Praxis interessanten Fällen erhalten wir die Varianz für die Schätzer ähnlich wie bei der ML-Methode.
Annahme: Schätzer effizient → Schranke Minimaler Varianz
WDF für Schätzer Gauss-WDF χ^2-Fkt. ist Parabel

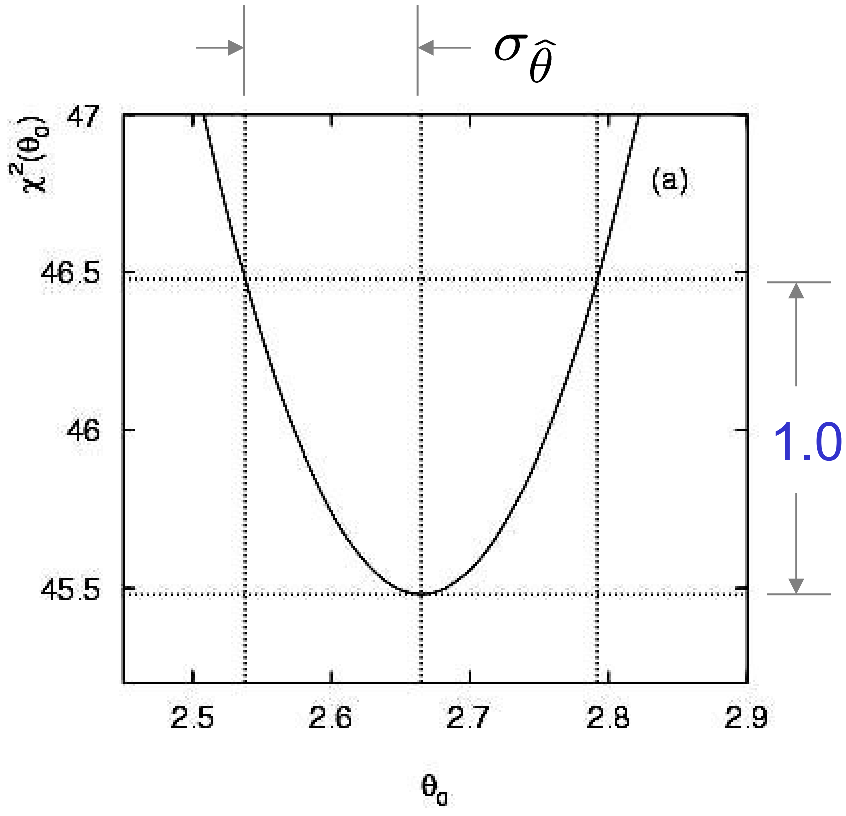
$$\chi^2(\theta) = -2 \ln L(\theta)$$

und daher

$$\widehat{\sigma^2}_{\hat{\theta}} \approx 2 \left[\frac{\partial^2 \chi^2}{\partial \theta^2} \right]^{-1}_{\theta=\hat{\theta}}$$

und in der graphischen Methode lesen wir die Werte ab, für die gilt

$$\chi^2(\theta) = \chi^2_{\text{min}} + 1$$



Zweiparameteranpassung mit KQ-Methode

2-parameter case (line with nonzero slope):

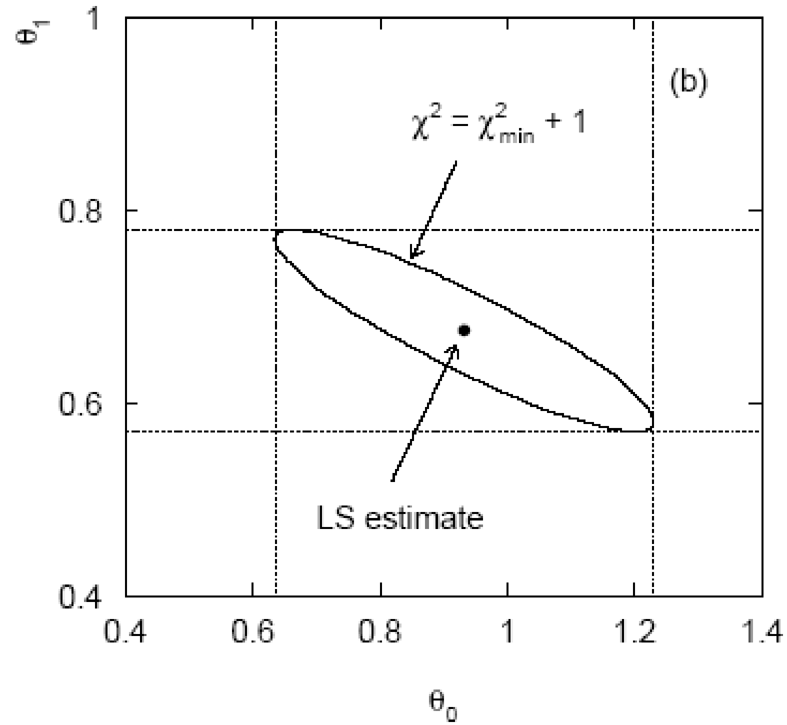
$$\hat{\theta}_0 = 0.93 \pm 0.30,$$

$$\hat{\theta}_1 = 0.68 \pm 0.10$$

$$\widehat{\mathrm{cov}}[\hat{\theta}_0, \hat{\theta}_1] = -0.028$$

$$r = -0.90$$

$$\chi^2 = 3.99$$



Tangent lines $\rightarrow$ $\sigma_{\hat{\theta}_0}$, $\sigma_{\hat{\theta}_1}$.

Angle of ellipse $\rightarrow$ correlation (same as for ML)

Lineare Kleinste Quadrate

Wichtiger Spezialfall: die theoretische Vorhersage ist linear in den zu bestimmenden Parametern (nicht notwendigerweise in Stützwerten)

$$\lambda(x;\vec{\theta}) = \sum_{l=1}^{m} \theta_l \, a_l(x)$$

$a_l(x)$ sind beliebige Funktionen in x (können auch exp, ln, sin. etc. sein)

Einfaches Beispiel: Polynom $\lambda(x;\vec{\theta}) = \sum_{l=1}^{m} \theta_l \, x^l,$

In diesem Fall von linearen kleinsten Quadraten ist
die Lösung analytisch möglich (Matrixinversion und -multiplikation),
Fehler exakt und analytisch berechenbar (Fehlerfortpflanzung)
Die KQ-Schätzer haben gewisse optimale Eigenschaften.

Lineare Kleinste Quadrate: Lösung

Wir definieren die Matrix $A_{il} \equiv a_l(x_i)$

Dann läßt sich die Theorievorhersage für den i-ten Stützwert schreiben als:

$$\lambda(x_i;\vec{\theta}) = \sum_{l=1}^{m} \theta_l\, a_l(x_i) = \sum_{l=1}^{m} A_{il}\,\theta_l = (A\vec{\theta})_i .$$

Der zu minimierende Ausdruck ergibt sich dann zu:

$$\begin{aligned}\chi^2 &= (\vec{y}-\vec{\lambda})^T V^{-1}(\vec{y}-\vec{\lambda}) \\ &= (\vec{y}-A\vec{\theta})^T V^{-1}(\vec{y}-A\vec{\theta})\end{aligned}$$

wobei gilt: $\vec{y} = (y_1, ..., y_N)$ und $\vec{\lambda} = (\lambda_1, ..., \lambda_N)$

Lineare Kleinste Quadrate: Lösung (2)

Nullsetzen der ersten Ableitungen bzgl. der gesuchten Parameter ergibt:

$$\nabla_{\vec{\theta}} \chi^2 = -2\left(A^T V^{-1} \vec{y} - A^T V^{-1} A \hat{\vec{\theta}}\right) = 0$$

Wenn die Matrix $A^T V^{-1} A$ nicht singulär ist, kann sie invertiert werden.

Die zu schätzenden Parametern ergeben sich dann zu:

$$\hat{\vec{\theta}} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} \vec{y} \equiv B \vec{y}$$

d.h. aus Matrixmultiplikation mit dem Vektor der Messwerte.

Es ist keine iterative Minimierung notwendig.

Die Lösung ist exakt.

Reskalierung der Kovarianzmatrix ändern Schätzwerte nicht

Lineare Kleinste Quadrate: Varianz

Die Kovarianzen zwischen den Parametern $U_{l,k} \equiv \mathrm{cov}[\hat{\theta}_l, \hat{\theta}_k]$

erhält man aus der Gausschen Fehlerfortplanzung zu:

$$U = BVB^T = \left(A^T V^{-1} A\right)^{-1}$$

Der rechte Ausdruck ist bereits aus der Bestimmung der Parameter bekannt.

Die linearen KQ-Schätzer sind unverzerrt.

Aus dem Gauss-Markov-Theorem folgt:
unter allen unverzerrten Schätzer, welche lineare Funktionen der Messwerte sind, besitzen die KQ-Schätzer die kleinste Varianz. Dies gilt unabhängig von den WDFs der Messwerte und unabhängig von der Anzahl der Messwerte.

Wenn Messwerte gaussverteilt sind, dass sind die linearen KQ-Schätzer auch effizient.

Lineare KQ: Beispiel Geradenfit

Theorievorhersage: $\lambda(x; m, c) = mx + c$

Zu minimieren ist der Ausdruck: $\chi^2 = \sum_{i=1}^{N} \frac{(y_i - (mx_i + c))^2}{\sigma_i^2}$

Linear in m und c → im Prinzip Matrixformalismus (Übung)
Hier: zu Fuss, weil instruktiver.

Nullsetzen der partiellen Ableitungen nach c:

$$\frac{\partial \chi^2}{\partial c} = -2 \sum_{i=1}^{N} \frac{y_i - (mx_i + c)}{\sigma_i^2} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} = \hat{m} \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \hat{c} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$$

Dividieren durch $1/\sigma_i^2$ zeigt: Gerade geht durch den fehlergewichteten Schwerpunkt

Lineare KQ: Beispiel Geradenfit (2)

Nullsetzen der partiellen Ableitung nach m:

$$\frac{\partial \chi^2}{\partial m} = -2\sum_{i=1}^{N} x_i \frac{y_i - (mx_i + c)}{\sigma_i^2} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} = \hat{m}\sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} + \hat{c}\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}$$

Auflösen der Gleichungen liefert: (Summen laufen von 1 bis N)

$$\hat{m} = \frac{\dfrac{\sum \frac{x_i y_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} - \dfrac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} \dfrac{\sum \frac{y_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}}{\dfrac{\sum \frac{x_i^2}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} - \left(\dfrac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}\right)^2} \qquad \hat{c} = \frac{\sum \frac{y_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} - \hat{m}\frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}$$

Falls σ_i gleich sind, dann findet man dass das χ^2 im Minimum Gegeben ist durch: (Einsetzen der Schätzwerte in χ^2)

$$\chi^2 = \frac{N V[y]}{\sigma^2}\left(1 - \rho_{xy}^2\right)$$

$V[y]$: Varianz der Messwerte y_i

ρ_{xy}: Korrelation in den Wertepaaren (x_i, y_i)

Lineare KQ: Beispiel Geradenfit (2)

Die Varianzen und Kovarianzen ergeben sich zu (z.B. aus Fehlerfortpflanzung):

$$V[\hat{c}] = \frac{\sum \frac{x_i^2}{\sigma_i^2}}{\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2}\right)^2} \qquad V[\hat{m}] = \frac{\sum \frac{1}{\sigma_i^2}}{\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2}\right)^2}$$

$$\mathrm{cov}[\hat{c}, \hat{m}] = -\frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2}\right)^2}$$

Fehler hängen nicht von den y_i ab.

Für den Fehler der Vorhersage des Wertes für x-Wert aus der Geradenanpassung gilt: Kovarianz berücksichtigen!!

$$V[y] = x^2 V[\hat{m}] + V[\hat{c}] + 2x\,\mathrm{cov}[\hat{c}, \hat{m}]$$

Beseitigen der Korrelation durch Variablentransformation:

$x \rightarrow x - \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}$ dann gilt $\frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}} = 0$

Geradenanpassung mit Fehler auf x und y

Sind x_i und y_i mit Fehlern behaftet ist die obige Methode nicht anwendbar

Ein Messpaar (x_i, y_i) kann von verschieden wahren Werten x_i stammen.

Problem kann mit ML-Methode gelöst werden und Wahrscheinlichkeiten können aufintegriert werden.

Im Falle Gausscher Fehler und gleicher Varianzen für alle x_i und y_i Werte und für ein Geradenapassung kann gezeigt werden (siehe Barlow), dass das Integral der Wahrscheinlichkeiten über alle Punkte gleich der Wahrscheinlichkeit ist, von dem wahrscheinlichsten Punkt auf der Gerade. sind die Fehler auf x und y gleich ist dies einfach der Lotpunkt auf die Gerade und der Abstand von der Geraden beträgt:

$$h_i = \frac{y_i - mx_i - c}{\sqrt{1 + m^2}}$$

Die zu minimierende Größe ist dann: $$\chi^2 \sim \sum_{i=1}^{N} \frac{(y_i - mx_i - c)^2}{1 + m^2}$$

Geradenanpassung mit Fehler auf x und y (2)

Sind die Fehler auf x und y nicht gleich kann man dies durch eine Variablentransformation erreichen: $x' = x/\sigma_x, y' = y/\sigma_y$

Nach Variablenrücktransformation ergeben sich die KQ-Schätzer zu:

$$\hat{m} = \frac{\sigma_y}{\sigma_x}\left(A \pm \sqrt{A^2 + 1}\right) \quad \text{mit} \quad A = \frac{\sigma_x^2 V[y] - \sigma_y^2 V[x]}{2\,\sigma_x \sigma_y \mathrm{cov}[x, y]}$$

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

+(-) für cov[x,y] > (<) 0

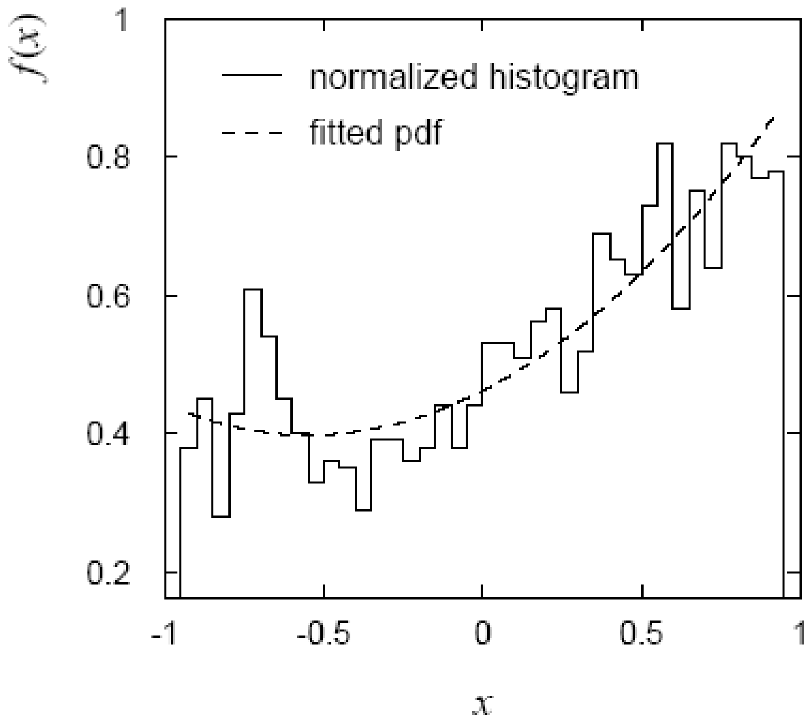
Die Kovarianzen ergeben sich aus Fehlerfortpflanzung

Sind die Fehler für die einzelnen Messwerte in x und y nicht gleich, dann gibt es keine geschlossene analytische Lösung des Problems

KQ für Histogramme

Gegeben: ein Histogramm mit N bins und n Einträgen

anzupassende WDF $f(x;\vec{\theta})$



Wir haben in jedem Bin i:

y_i = Anzahl der Einträge in den Daten

Vorhersage ergibt sich aus Anzahl der Beobachtungen und dem Integral der WDF über Binbreite;

$$\lambda_i(\vec{\theta}) = n \int_{x_i^{\min}}^{x_i^{\max}} f(x;\vec{\theta})dx = np_i(\vec{\theta})$$

Wichtig! Die Gesamterwartung wird auf die beobachte Anzahl von von Messwerten n in den Daten angepasst.

KQ für Histogramme (2)

Anpassung mittels KQ: minimiere

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\vec{\theta}))^2}{\sigma_i^2}$$

wobei $\sigma_i^2 = V[y_i]$, nicht notwendigerweise bekannt.

Behandele y_i als Poisson-verteilte ZV

$$\sigma_i^2 = \lambda_i(\vec{\theta})$$ KQ-Methode

$$\sigma_i^2 = y_i$$ Modifizierte KQ-Methode (MKQ)

MKQ ist manchmal rechentechnisch einfacher zu behandeln.
Aber: χ^2_{min} folgt nicht mehr Chi-Quadrat-WDF (oder ist nicht definiert), wenn einige Bins wenige (oder keine Einträge) haben.

Erweiterte KQ für Histogramme ?

Bisher Normierung der Vorhersage aus Daten n.
Versuch: betrachte n als Poisson-ZV mit Mittelwert ν

Dann wird Theorievorhersage zu: $\lambda_i(\vec{\theta}, \nu) = \nu \int_{x_i^{\min}}^{x_i^{\max}} f(x; \vec{\theta}) dx = \nu p_i(\vec{\theta})$

Problem: $\hat{\nu}$ aus KQ ist verzerrter Schätzer (und n kennen wir ohnehin)

$$\hat{\nu}_{\mathrm{LS}} = n + \frac{\chi^2_{\min}}{2} \qquad \hat{\nu}_{\mathrm{MLS}} = n - \chi^2_{\min}$$

$\chi^2_{\min}$ folgt Chi-Quadrat-Verteilung mit FG= Messwerte-Parameter
Ewartungswert = #FG → Verzerrung der KQ / MKQ (LS/MLS)-Schätzer.

→ Normierung aus Beobachtung in KQ verwenden.
oder EML anwenden → unverzerrte Schätzer für ν

KQ für Histogramme Beispiel

400 Messungen von ZV x im Bereich [0,2], 20 Bins im Histogramm

WDF für x = θ Gauss-WDF + $(1-\theta)$ Exp-WDF

Wahre Werte für Parameter: $\theta=0.5$, $\mu=1$, $\sigma=0.35$, $\tau=4$

Annahme: μ,σ,τ bekannt
Schätzung von θ bzw. $\nu_{Gauss} = \nu\,\theta$ und $\nu_{Exp} = \nu\,(1-\theta)$

a) Anpassung von ν und θ

KQ: $\sigma_i^2 = \lambda_i$ $\chi^2_{min}=17.1$ $\nu_{KQ}= 408.5 \pm 26.2$ zu groß um $\chi^2_{min}/2$
$\theta = 0.498\pm0.056$

MKQ: $\sigma_i^2 = y_i$ $\chi^2_{min}=17.8$ $\nu_{MKQ}= 382.2 \pm 19.5$ zu klein χ^2_{min}
$\theta = 0.551\pm0.062$

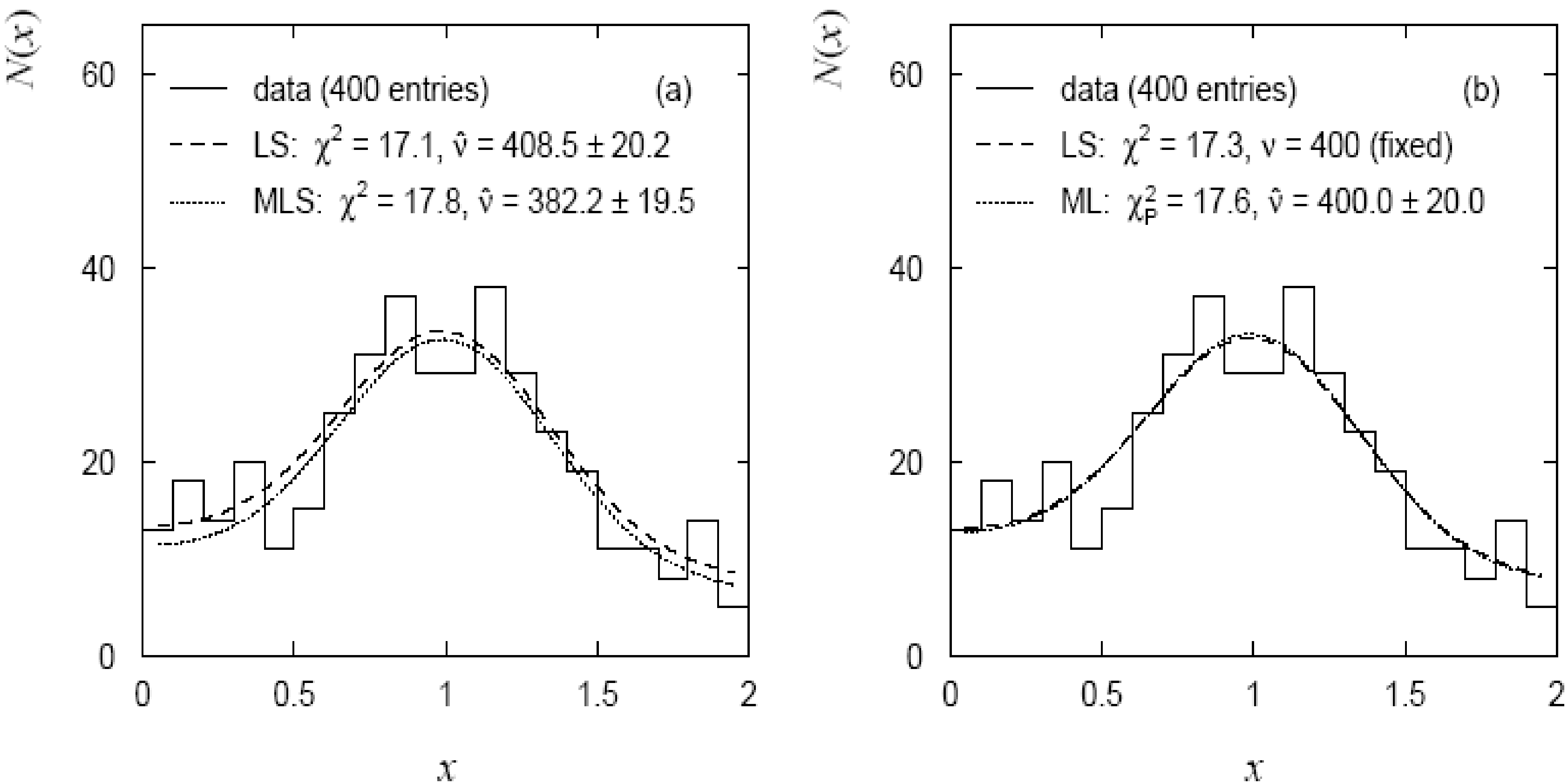
b) Nur Anpassung von θ (ν=n fest für KQ oder aus Fit für EML)

KQ: $\chi^2_{min}=17.3$ $\theta = 0.498\pm0.056$

EML: n=400± 20 $\chi^2_{min}=17.6$ $\theta = 0.514\pm0.057$

Normierung der kleinsten Quadrate: Beispiel

N= 400 Ereignisse, n=20 Bins



Besser: entweder n=konst in KQ oder verwende ML-Methode

Verwendung der KQ zur Kombination von Messungen

Verwende KQ, um gewichteten Mittelwert von N Messungen der Observable λ zu bestimmen.

y_i = Ergebnis der Messung i, $i = 1, \ldots, N$

$\sigma_i^2 = V[y_i]$ Varianz der Messung *i,* als bekannt angenommen

λ = wahrer Wert (Parameter der geschätzt werden soll)

Für unkorrelierte Messungen y_i , minimiere

$$\chi^2(\lambda) = \sum_{i=1}^{N} \frac{(y_i - \lambda)^2}{\sigma_i^2}$$

Setze $\frac{\partial \chi^2}{\partial \lambda} = 0$ und löse nach λ auf:

$$\hat{\lambda} = \frac{\Sigma_{i=1}^{N} y_i/\sigma_i^2}{\Sigma_{j=1}^{N} 1/\sigma_j^2} \qquad V[\hat{\lambda}] = \frac{1}{\Sigma_{i=1}^{N} 1/\sigma_i^2}$$

Kombination korrelierter Messungen mit KQ

Wenn die Kovarianzmatrix gegeben ist durch $\mathrm{cov}[y_i, y_j] = V_{ij}$

dann minimiere $\chi^2(\lambda) = \sum_{i,j=1}^{N} (y_i - \lambda)(V^{-1})_{ij}(y_j - \lambda)$

Dies liefert: $$\hat{\lambda} = \sum_{i=1}^{N} w_i y_i$$

$$w_i = \frac{\Sigma_{j=1}^{N}(V^{-1})_{ij}}{\Sigma_{k,l=1}^{N}(V^{-1})_{kl}}$$

$$V[\hat{\lambda}] = \sum_{i,j=1}^{N} w_i V_{ij} w_j$$

Der KQ-Schätzer für die konstante Funtion λ ist erwartungstreu und effizient. Dies folgt aus dem Gauss-Markov-Theorem.

Beispiel: Mittelung zweier korrelierter Messungen

Suppose we have y_1, y_2, and $V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

$$\rightarrow \quad \hat{\lambda} = w y_1 + (1-w) y_2, \quad w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

$$V[\hat{\lambda}] = \frac{(1-\rho^2)\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} = \sigma^2$$

The increase in inverse variance due to 2nd measurement is

$$\frac{1}{\sigma^2} - \frac{1}{\sigma_1^2} = \frac{1}{1-\rho^2}\left(\frac{\rho}{\sigma_1} - \frac{1}{\sigma_2}\right)^2 > 0$$

$\rightarrow$ 2nd measurement can only help.

Negative Gewichte im KQ-Mittelwert

If $\rho > \sigma_1/\sigma_2$, $\rightarrow$ $w < 0$,

$\rightarrow$ weighted average is not between y_1 and y_2 (!?)

Cannot happen if correlation due to common data, but possible for shared random effect; very unreliable if e.g. ρ, σ_1, σ_2 incorrect.

See example in SDA Section 7.6.1 with two measurements at same temperature using two rulers, different thermal expansion coefficients:

average is outside the two measurements; used to improve estimate of temperature.

Lineare KQ mit linearen Zwangsbedingungen

Oft unterliegen zu schätzende Parameter (eventuell gleich Beobachtungen) gewissen Zwangsbedingungen (ZB).

Messungen erfüllen ZB nicht → verbesserte Messungen sollen ZB erfüllen

2 Methoden: - Eliminierung der überflüssigen Parameter
- Einführung von Lagrange-Multiplikatoren

Beispiel: Messung der Winkel im Dreieck

$y_1 = 63^\circ, \quad y_2 = 34^\circ, \quad y_3 = 85^\circ$

$\sum_{i=1}^{3} y_i - 180^\circ = 2^\circ \qquad \sigma = 1^\circ$

verbesserte Messungen η_i sollen ZB $\Sigma\eta_i = 180$ erfüllen

Lineare KQ mit linearen Zwangsbedingungen (2)

Verbesserte Messungen ableiten aus den Bedingungen

$$\chi^2(\eta_1,\eta_2,\eta_3) = \sum_{i=1}^{3}\left(\frac{y_i-\eta_i}{\sigma_i}\right)^2 = \text{minimum}$$

$$\sum_{i=1}^{3}\eta_i - 180^\circ = 0.$$

1) Eliminierungsmethode: $\eta_3 = 180 - \eta_1 - \eta_2$

Minimiere den Ausdruck:

$$\chi^2(\eta_1,\eta_2) = \left(\frac{y_1-\eta_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\eta_2}{\sigma_2}\right)^2 + \left(\frac{y_3-(180^\circ-\eta_1-\eta_2)}{\sigma_3}\right)^2 = \text{minimum}$$

Lineare KQ mit linearen Zwangsbedingungen (3)

Ergebnis der Minimerung:

$$\hat{\eta}_1 = \frac{1}{3}\,(180^\circ + 2y_1 - y_2 - y_3) = 62\frac{1}{3}^\circ$$

$$\hat{\eta}_2 = \frac{1}{3}\,(180^\circ - y_1 + 2y_2 - y_3) = 33\frac{1}{3}^\circ$$

Und aus der Zwangsbedingung für η_3:

$$\hat{\eta}_3 = 180^\circ - \hat{\eta}_1 - \hat{\eta}_2 = 84\frac{1}{3}^\circ$$

Nun erfüllen die η_i die Zwangsbedingung

Der Überschuss von 2 Grad wird gleichmäßig auf alle Winkel Verteilt (-2/3 Grad für jeden Winkel), da Varianz alle gleich.

Lineare KQ mit linearen Zwangsbedingungen (3)

b) Lagrangsche Multiplikatoren

führe für jede Zwangsbedingung einen Lagrangmultiplikator λ ein

und addiere für jede ZB einen Term im Format 2λ “ZB= 0” zum normalen χ^2

$$X^2(\eta_1,\eta_2,\eta_3,\lambda) = \sum_{i=1}^{3}\left(\frac{y_i-\eta_i}{\sigma_i}\right)^2 + 2\lambda\left(\sum_{i=1}^{3}\eta_i - 180^\circ\right) = \text{minimum}$$

Der Ausdruck enthält nun vier zu schätzende Parameter: $\eta_1,\eta_2,\eta_3,\lambda$

Nullsetzten der Ableitung
Liefert die 4 Gleichungen:

$$0 = \frac{\partial X^2}{\partial \eta_1} = -\frac{2}{\sigma^2}(y_1-\eta_1) + 2\lambda,$$

$$0 = \frac{\partial X^2}{\partial \eta_2} = -\frac{2}{\sigma^2}(y_2-\eta_2) + 2\lambda,$$

$$0 = \frac{\partial X^2}{\partial \eta_3} = -\frac{2}{\sigma^2}(y_3-\eta_3) + 2\lambda,$$

$$0 = \frac{\partial X^2}{\partial \lambda} = 2\left(\sum_{i=1}^{3}\eta_i - 180^\circ\right).$$

Lineare KQ mit linearen Zwangsbedingungen (3)

Multiplikation der ersten drei Gl. mit -1 und der letzten mit $1/\sigma^2$

$$0 = \frac{2}{\sigma^2}\left(\sum_{i=1}^{3} y_i - 180^\circ\right) - 6\lambda \qquad \hat{\lambda} = \frac{1}{3}\frac{1}{\sigma^2}\left(\sum_{i=1}^{3} y_i - 180^\circ\right)$$

Lösen der ersten drei Gleichungen liefert;

$$\hat{\eta}_j = y_j - \sigma^2 \cdot \hat{\lambda} = y_j - \frac{1}{3}\left(\sum_{i=1}^{3} y_i - 180^\circ\right)$$

Selbes Ergebnis wie bei Eliminierung.

Gleichungen symmetrisch in allen Parametern.

Lineare KQ mit linearen Zwangsbedingungen (3)

Bestimmung der Kovarianzmatrix aus Fehlerfortplanzung liefert

$$V(\underline{\hat{\eta}}) = V(\underline{y}) - \frac{1}{3}\sigma^2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1,1,1) = \sigma^2 \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$$

Die verbesserten Messgrößen η_l :

- sind nun negativ korreliert (-1/3)

- besitzen verkleinerte Varianzen (2/3 vgl 1 für y_i)

Lineare KQ mit linearen ZB und Langrange-Mult.

Allgemeine Problemstellung:

$\underline{y}$ N-dimensionaler Vektor der Messwerte

$\underline{\theta}$ L-dimensionaler Vektor der Messwerte

A Lin. Theorievorhersage NxL

$$\chi^2(\underline{\theta}) = (\underline{y}-A\underline{\theta})^T V^{-1}(\underline{y}-A\underline{\theta}) = \text{minimum}$$

$$B\underline{\theta} - \underline{b} = \underline{0}.$$

K Zwangsbedingungen in l Parametern
→ B = KxL Matrix
b = K-dimensionaler Vektor

$$\chi^2(\underline{\theta},\underline{\lambda}) = (\underline{y}-A\underline{\theta})^T V^{-1}(\underline{y}-A\underline{\theta}) + 2\underline{\lambda}^T(B\underline{\theta}-\underline{b}) = \text{minimum}$$

Lineare KQ mit linearen ZB und Langrange-Mult.

K-dim. Vektor von Lagrangemultiplikatoren: $\underline{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_K\}$

Zu minimierender Ausdruck:

$$X^2(\underline{\theta},\underline{\lambda}) = (\underline{y}-A\underline{\theta})^T V^{-1}(\underline{y}-A\underline{\theta}) + 2\underline{\lambda}^T(B\underline{\theta}-\underline{b}) = \text{minimum}$$

Minimierung bzgl. Parameter und Lagrangemultiplikatoren :

$$\nabla_\theta X^2 = -2(A^T V^{-1}\underline{y} - A^T V^{-1} A\underline{\theta}) + 2B^T\underline{\lambda} = \underline{0}$$

$$\nabla_\lambda X^2 = 2(B\underline{\theta}-\underline{b}) = \underline{0}.$$

Einführung der Abkürzungen:

$$C \equiv A^T V^{-1} A \qquad \underline{c} \equiv A^T V^{-1} \underline{y}$$

Lineare KQ mit linearen ZB und Langrange-Mult.

Gleichungen in neuer Form:

$$C\underline{\theta} + B^T\underline{\lambda} = \underline{c}$$

$$B\underline{\theta} = \underline{b}$$

Falls C^{-1} existiert, multipliziere 1. Gl. mit BC^{-1} und ersetze durch 2. Gl.

$$\underline{b} + BC^{-1}B^T\underline{\lambda} = BC^{-1}\underline{c}$$

Abkürzung: $V_B \equiv BC^{-1}B^T$

Lösung für die Lagrangemultiplikatoren:

$$\hat{\underline{\lambda}} = V_B^{-1}(BC^{-1}\underline{c}-\underline{b})$$

Lineare KQ mit linearen ZB und Langrange-Mult.

Einsetzen in obige Gleichung liefert Lösungen für Parameter:

$$\underline{\hat{\theta}} = C^{-1}\underline{c} - C^{-1}B^T V_B^{-1}(BC^{-1}\underline{c}-\underline{b})$$

Lösung ist exakt.

$C^{-1}\underline{c}$ ist die Lösung für Problem ohne ZB, taucht in beiden Lsg. auf

$(BC^{-1}\underline{c}-\underline{b})$ misst Verletzung der Zwangbedingungen

Kovarianzmatrix aus Fehlerfortpflanzung

$$V(\underline{\hat{\theta}}) = \left[C^{-1}A^TV^{-1} - C^{-1}B^TV_B^{-1}BC^{-1}A^TV^{-1}\right]V\left[C^{-1}A^TV^{-1} - C^{-1}B^TV_B^{-1}BC^{-1}A^TV^{-1}\right]^T.$$

Lineare KQ mit linearen ZB und Langrange-Mult.

Oder vereinfacht:

$$V(\hat{\underline{\theta}}) = C^{-1} - (BC^{-1})^T V_B^{-1} (BC^{-1})$$

Erster Summand aus Lösungen ohne Zwangsbedingungen
Zweiter Summand bewirkt Reduktion der Diagonalelemenete

Die verbesserten Messgrößen ergeben sich zu

$$\hat{\underline{\eta}} = A\left[C^{-1}\underline{c} - C^{-1}B^T V_B^{-1}(BC^{-1}\underline{c} - \underline{b})\right]$$

Mit Kovarianzmatrix:

$$V(\hat{\underline{\eta}}) = AC^{-1}A^T - A(BC^{-1})^T V_B^{-1}(BC^{-1})A^T$$

Erster Summand aus Lösungen ohne Zwangsbedingungen