

# Konzepte für Experimente an zukünftigen Hadroncollidern II

PD Dr. Oliver Kortner

21.05.2021

## Wahrscheinlichkeitsverteilungen

- Eine physikalische Messung ist ein **Zufallsprozess**.
- Eine Messgröße  $x$ , die den Ausgang eines Zufallsprozesses angibt, bezeichnet man als eine **Zufallsvariable** oder **Zufallsgröße**.
- Jede Funktion von  $x$  ist ebenfalls eine Zufallsgröße.
- Falls die Zufallsgröße nur diskrete Werte annehmen kann, gibt es für das Auftreten jedes dieser Werte eine Wahrscheinlichkeit, was die **Wahrscheinlichkeitsfunktion** ist.
- Bei Zufallsvariablen mit kontinuierlichem Wertebereich ersetzt die **Wahrscheinlichkeitsdichte**  $p(x)$  die Wahrscheinlichkeitsfunktion. Es sei  $\Omega$  eine messbare Menge möglicher Werte von  $x$ , deren Maß größer Null ist. Dann ist

$$\int_{\Omega} p(x) dx$$

die Wahrscheinlichkeit dafür, einen Wert  $x \in \Omega$  zu beobachten.

## Axiomatische Definition des Wahrscheinlichkeitsbegriffs

Das mathematische Fachgebiet der Wahrscheinlichkeitstheorie fußt auf Kolmogorovs Axiomen.

### Kolmogorovaxiome

$\Sigma$  bezeichne eine Ereignismenge.

1. Für jedes Ereignis  $A \in \Sigma$  ist die Wahrscheinlichkeit für das Auftreten von  $A$  eine reelle Zahl  $p(A) \in [0, 1]$ .
2. Das sichere Ereignis  $S \in \Sigma$  hat die Wahrscheinlichkeit  $p(S) = 1$ .
3. Die Wahrscheinlichkeit einer Vereinigung abzählbar vieler inkompatibler Ereignisse ist gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse. Dabei heißen Ereignisse  $A_k$  **inkompatibel**, wenn sie paarweise disjunkt sind, also  $A_k \cap A_\ell = \emptyset$  für alle  $k \neq \ell$  gilt.

## Kenngößen von Wahrscheinlichkeitsverteilungen

Nomenklatur.  $D$ : Wertebereich einer Zufallsgröße  $x = (x_1, \dots, x_n)$ .

$p(x)$ : Wahrscheinlichkeitsdichte von  $x$ .

( $D$  ist der Definitionsbereich von  $p$ .)

## Definitionen

Der **Erwartungswert** von  $x$ ,  $E(x)$  (auch  $\langle x \rangle$ ) ist definiert als

$$E(x) := \int_D x \cdot p(x) dx.$$

Die **Kovarianzmatrix**  $cov(x_k, x_l)$  ist definiert als

$$cov(x_k, x_l) := \langle (x_k - \langle x_k \rangle) \cdot (x_l - \langle x_l \rangle) \rangle .$$

Das Diagonalelement  $cov(x_k, x_k)$  nennt man die **Varianz von  $x_k$** ,  $Var(x_k)$ ,  $\sqrt{Var(x_k)}$  die **Standardabweichung**  $\sigma(x_k)$ .

## Erwartungswert einer Funktion einer Zufallsgröße

- Eine Funktion  $f(x)$  ist ebenfalls eine Zufallsgröße.

$$\langle f \rangle = \int_D f(x)p(x)dx.$$

- Wenn  $f(x) = f(x - \langle x \rangle + \langle x \rangle)$  nur für kleine Werte von  $|x - \langle x \rangle|$  deutlich von 0 verschieden ist, kann man  $f(x)$  durch

$$f(\langle x \rangle) + \left. \frac{df}{dx} \right|_{\langle x \rangle} \cdot (x - \langle x \rangle)$$

annähern. Dann ist

$$\begin{aligned} \langle f \rangle &\approx \left\langle f(\langle x \rangle) + \left. \frac{df}{dx} \right|_{\langle x \rangle} \cdot (x - \langle x \rangle) \right\rangle \\ &= \langle f(x) \rangle + \left\langle \left. \frac{df}{dx} \right|_{\langle x \rangle} \cdot (x - \langle x \rangle) \right\rangle \\ &= f(\langle x \rangle) + \left. \frac{df}{dx} \right|_{\langle x \rangle} \cdot (\langle x \rangle - \langle x \rangle) = f(\langle x \rangle). \end{aligned}$$

## Varianz einer Funktion einer Zufallsgröße

Sonderfall:  $f(x) \in \mathbb{R}$ .

$$\begin{aligned} \text{Var}(f) &= \langle (f - \langle f \rangle)^2 \rangle = \langle [f - f(\langle x \rangle)] \rangle \\ &\approx \left\langle \left[ \sum_{k=1}^n \frac{df}{dx_k} \Big|_{\langle x \rangle} \cdot (x_k - \langle x_k \rangle) \right]^2 \right\rangle \\ &= \left\langle \left[ \sum_{k,\ell=1}^n \frac{df}{dx_k} \Big|_{\langle x \rangle} \frac{df}{dx_\ell} \Big|_{\langle x \rangle} \cdot (x_k - \langle x_k \rangle) \cdot (x_\ell - \langle x_\ell \rangle) \right] \right\rangle \\ &= \sum_{k,\ell=1}^n \frac{df}{dx_k} \Big|_{\langle x \rangle} \frac{df}{dx_\ell} \Big|_{\langle x \rangle} \cdot \langle (x_k - \langle x_k \rangle) \cdot (x_\ell - \langle x_\ell \rangle) \rangle \\ &= \sum_{k,\ell=1}^n \frac{df}{dx_k} \Big|_{\langle x \rangle} \frac{df}{dx_\ell} \Big|_{\langle x \rangle} \cdot \text{cov}(x_k, x_\ell), \end{aligned}$$

was die bekannte Fehlerfortpflanzungsformel ist.

## Beispiele wichtiger Wahrscheinlichkeitsverteilungen

# Die Binomialverteilung

- Die **Binomialverteilung** gibt die Wahrscheinlichkeit an,  $n_k$  Ereignisse aus insgesamt  $N$  Ereignissen zu beobachten, wenn  $\nu_k$  Ereignisse erwartet werden:

$$p(n_k; \nu_k) = \binom{N}{n_k} \left(\frac{\nu_k}{N}\right)^{n_k} \left(1 - \frac{\nu_k}{N}\right)^{N-n_k}.$$

- Mit  $p := \frac{\nu_k}{N}$  erhält man aus

$$\begin{aligned} 0 &= \frac{d}{dp} 1 = \frac{d}{dp} \sum_{n_k=0}^N \binom{N}{n_k} p^{n_k} (1-p)^{N-n_k} \\ &= \sum_{n_k=0}^N \binom{N}{n_k} [n_k p^{n_k-1} (1-p)^{N-n_k} - (N-n_k) p^{n_k} (1-p)^{N-n_k-1}] \\ &= \frac{1}{p} \langle n_k \rangle - \frac{1}{1-p} \langle N - n_k \rangle = \left( \frac{1}{p} + \frac{1}{1-p} \right) \langle n_k \rangle + \frac{N}{1-p} \\ &= \frac{1}{p(1-p)} \langle n_k \rangle + \frac{N}{1-p} \Leftrightarrow \langle n_k \rangle = N \cdot p = N \cdot \frac{\nu_k}{N} = \nu_k. \end{aligned}$$

- Mit dem gleichen Rechenrick erhält man  $\text{Var}(n_k) = \nu_k(1 - \frac{\nu_k}{N})$ .

# Übergang zur Poissonverteilung

Wenn  $\nu \gtrsim 10$ ,  $\nu \ll N$  und  $N$  groß ist, kann man die Binomialverteilung durch die **Poissonverteilung** nähern. Grundlage ist die Stirlingformel:

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \text{ für } n \rightarrow \infty.$$

$$\begin{aligned} p(n_k; \nu_k) &= \frac{N!}{n_k!(N - n_k)!} p^{n_k} (1 - p)^{N - n_k} \\ &\approx \frac{1}{n_k!} p^{n_k} \left(\frac{N}{e}\right)^N \sqrt{2\pi N} \frac{1}{\left(\frac{N - n_k}{e}\right)^{N - n_k} \sqrt{2\pi(N - n_k)}} (1 - p)^{N - n_k} \\ &= \frac{1}{n_k} p^{n_k} e^{-n_k} \underbrace{\sqrt{\frac{N}{N - n_k}}}_{\rightarrow 1 \text{ f. } N \rightarrow \infty} \frac{N^N}{(N - n_k)^{N - n_k}} (1 - p)^{N - n_k} \\ &\approx \frac{1}{n_k!} e^{-n_k} p^{n_k} N^{n_k} N^{N - n_k} (1 - p)^{N - n_k} \frac{1}{(N - n_k)^{N - n_k}} \\ &= \frac{\nu_k^{n_k}}{n_k!} e^{-n_k} \frac{(N - \nu_k)^{N - n_k}}{(N - n_k)^{N - n_k}} \approx \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k} \text{ (Poissonverteilung)}. \end{aligned}$$

## Poissonverteilung

$$p(n_k; \nu_k) = \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k}.$$

## Normierung

$$\sum_{n_k=0}^{\infty} p(n_k; \nu_k) = e^{-\nu_k} \sum_{n_k=0}^{\infty} \frac{\nu_k^{n_k}}{n_k!} = e^{-\nu_k} \cdot e^{\nu_k} = 1.$$

Erwartungswert der Poissonverteilung:  $\nu_k$ , was aus  $0 = \frac{d}{d\nu_k} \sum_{n_k=0}^{\infty} p(n_k; \nu_k)$  folgt.

Varianz der Poissonverteilung:  $\nu_k$ , was aus  $0 = \frac{d^2}{d\nu_k^2} \sum_{n_k=0}^{\infty} p(n_k; \nu_k)$  folgt.

# Poissonverteilung für $\nu_k \rightarrow \infty$

Wenn  $\nu_k$  groß wird, dann ist die Wahrscheinlichkeit für das Auftreten kleiner Werte von  $n_k$  klein. Dann kann man  $n_k$  als groß betrachten und für  $n_k!$  in der Poissonverteilung die Stirlingformel verwenden:

$$\begin{aligned} \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k} &\approx \frac{\nu_k^{n_k}}{n_k^{n_k}} \frac{1}{\sqrt{2\pi n_k}} e^{n_k - \nu_k} \\ &\approx \frac{1}{\sqrt{2\pi \nu_k}} \exp\left(\ln \frac{\nu_k^{n_k}}{n_k^{n_k}}\right) \exp(n_k - \nu_k) \\ &= \frac{1}{\sqrt{2\pi \nu_k}} \exp\left(n_k \ln \frac{\nu_k}{\nu_k + n_k - \nu_k}\right) \exp(n_k - \nu_k) \\ &= \frac{1}{\sqrt{2\pi \nu_k}} \exp\left(n_k \ln \frac{1}{1 - \frac{n_k - \nu_k}{\nu_k}}\right) \exp(n_k - \nu_k) \\ &\approx \frac{1}{\sqrt{2\pi \nu_k}} \exp\left[n_k \cdot \left(-\frac{n_k - \nu_k}{\nu_k} - \frac{1}{2} \frac{(n_k - \nu_k)^2}{\nu_k^2}\right)\right] \exp(n_k - \nu_k) \\ &\quad \underbrace{\hspace{10em}}_{\approx -(n_k - \nu_k) - \frac{(n_k - \nu_k)^2}{2\nu_k}} \\ &\approx \frac{1}{\sqrt{2\pi \nu_k}} e^{-\frac{(n_k - \nu_k)^2}{2\nu_k}}. \end{aligned}$$

Normalverteilung einer eindimensionalen Zufallsgröße  $x \in \mathbb{R}$

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

- $\langle x \rangle = \mu$ ,  $Var(x) = \sigma^2$ .
- Die Poissonverteilung nähert sich also im Grenzfall  $\nu_k \rightarrow \infty$  einer Normalverteilung mit dem Erwartungswert  $\nu_k$  und der Varianz  $\nu_k$  an.

Normalverteilung einer  $d$ -dimensionalen Zufallsgröße  $x \in \mathbb{R}^d$

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\det(\Sigma)} \exp\left(-\frac{1}{2}(x - \mu)^t \Sigma (x - \mu)\right).$$

$\Sigma \in^d \mathbb{R}^d$ ,  $\mu \in \mathbb{R}^d$ .

- $\langle x \rangle = \mu$ .
- $cov(x_k, x_l) = \Sigma_{k,l}$ .

$w_n :=$  Wahrscheinlichkeit, einen Wert  $x \in [\mu - n\sigma, \mu + n\sigma]$  zu beobachten.

| $n$ | $w_n$                   |
|-----|-------------------------|
| 1   | 0,6827                  |
| 2   | 0,9545                  |
| 3   | 0,9973                  |
| 4   | $1 - 6,3 \cdot 10^{-5}$ |
| 5   | $1 - 5,7 \cdot 10^{-7}$ |

| $w_n$ | $n$   |
|-------|-------|
| 0,900 | 1,645 |
| 0,950 | 1,960 |
| 0,990 | 2,576 |
| 0,999 | 3,290 |

$(t_n)$  sei eine Folge von Zufallsgrößen und  $T$  ebenfalls eine Zufallsgröße. Man sagt,  $t_n$  **konvergiere stochastisch gegen  $T$** , wenn es zu jedem  $p \in [0, 1[$  und  $\epsilon > 0$  ein  $N$  so gibt, dass die Wahrscheinlichkeit  $P$ , dass  $|t_n - T| > \epsilon$  ist, kleiner als  $p$  für alle  $n > N$  ist:

$$P(|t_n - T| > \epsilon) < p \quad (n > N).$$

Mit anderen Worten: Die Wahrscheinlichkeit, einen von  $T$  verschiedenen Wert  $t_n$  zu beobachten, verschwindet für  $n \rightarrow \infty$ .

## Das Gesetz großer Zahlen

$(x_n)$  sei eine Folge unabhängiger Zufallsgrößen, wobei jedes  $x_n$  derselben Verteilungsfunktion folgt.  $\mu$  bezeichne den Erwartungswert von  $x_n$ . Dann konvergiert das arithmetische Mittel

$$\frac{1}{N} \sum_{n=1}^N x_n$$

stochastisch gegen  $\mu$ .

## Der zentrale Grenzwertsatz

$(x_n)$  sei eine Folge identisch verteilter Zufallsgrößen mit dem Mittelwert  $\mu$  und der Standardabweichung  $\sigma$ . Dann konvergiert für  $N \rightarrow \infty$  die standardisierte Zufallsgröße

$$Z_N := \frac{\sum_{n=1}^N x_n - N\mu}{\sigma\sqrt{N}}$$

punktweise gegen eine Normalverteilung mit dem Mittelwert 0 und der Standardabweichung 1.

Es sei  $\alpha$  ein Parameter einer Wahrscheinlichkeitsverteilung. Das Ziel der **Punktschätzung** ist es, den besten Schätzwert (die beste Messung in der Sprechweise der Physiker) von  $\alpha$  zu finden.

$x$ : Zufallsgröße, die den experimentellen Messwerten entspricht.

$p(x; \alpha)$ : Wahrscheinlichkeitsdichte für die Messung von  $x$  in Abhängigkeit des Parameters  $\alpha$ .

$x$  und  $\alpha$  können mehrdimensional sein.

**Definition.** Ein **Punktschätzer**  $\mathcal{E}_\alpha$  ist eine Funktion von  $x$ , mit der der Wert des Parameters  $\alpha$  geschätzt werden kann.  $\hat{\alpha}$  bezeichne diesen Schätzwert. Es ist also  $\hat{\alpha} = \mathcal{E}_\alpha(x)$ .

**Ziel** ist es, eine Funktion  $\mathcal{E}_\alpha$  zu finden, bei der  $\hat{\alpha}$  so nah wie möglich am wahren Wert von  $\alpha$  liegt.

Da  $\hat{\alpha}$  eine Funktion von Zufallsgrößen ist, ist  $\hat{\alpha}$  selbst eine Zufallsgröße.

$$p(\hat{\alpha}) = \int_D \mathcal{E}_\alpha(x) p(x; \alpha) dx,$$

wobei  $\alpha$  den wahren Wert des Parameters bezeichnet.

## Konsistenz

$n$ : Anzahl der Messungen, die für die Punktschätzung verwendet werden.

$\hat{\alpha}_n$ : zugehöriger Schätzwert.

$\alpha_0$ : wahrer Wert von  $\alpha$ .

Man bezeichnet  $\mathcal{E}_\alpha$  als **konsistenten Punktschätzer**, falls  $\hat{\alpha}_n$  stochastisch gegen  $\alpha_0$  konvergiert. D.h. die Wahrscheinlichkeit, einen von  $\alpha_0$  verschiedenen Wert zu schätzen, geht für  $n \rightarrow \infty$  gegen 0.

## Erwartungstreue

Der **Bias eines Schätzwerts**  $\hat{\alpha}$  ist definiert als

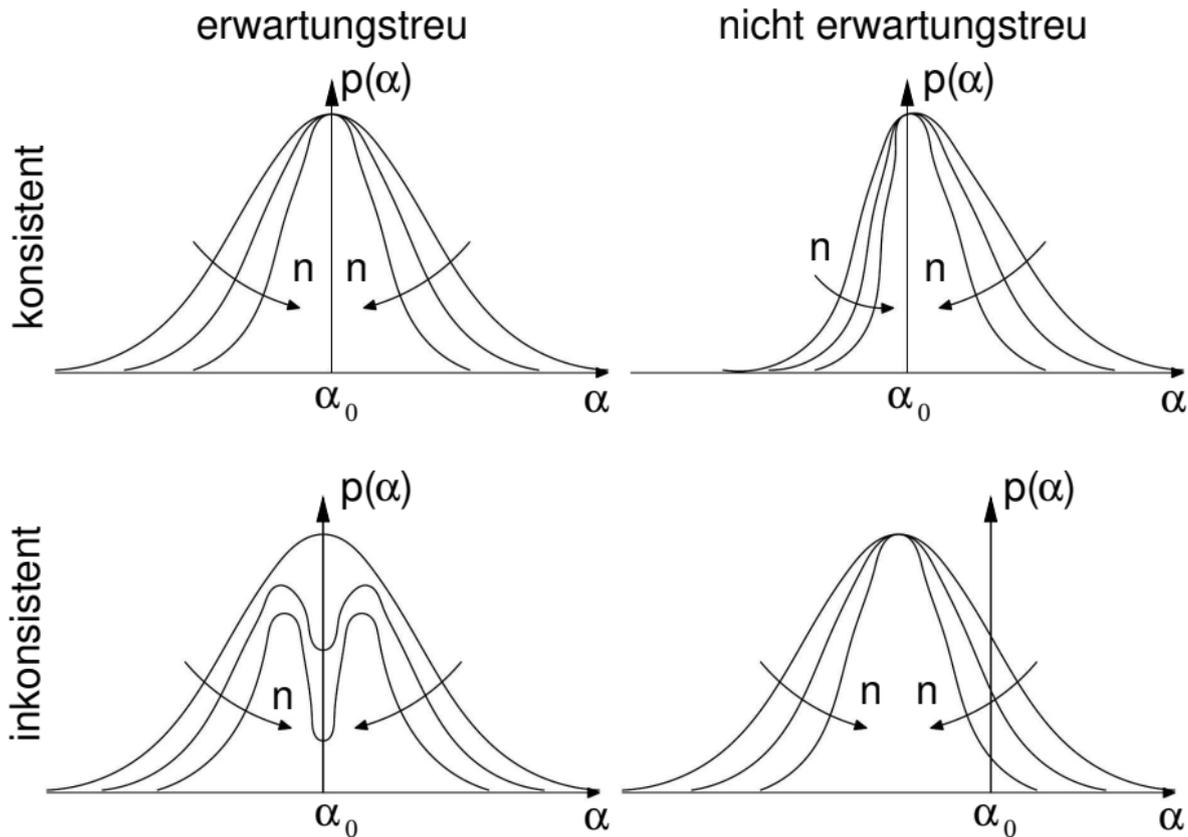
$$b_n(\hat{\alpha}) := E(\hat{\alpha}_n - \alpha_0) = E(\hat{\alpha}_n) - \alpha_0.$$

Der Punktschätzer ist **erwartungstreu**, wenn

$$b_n(\hat{\alpha}) = 0, \text{ also } E(\hat{\alpha}_n) = \alpha_0$$

für alle  $n$  ist.

# Veranschaulichung von Konsistenz und Erwartungstreue



## Effizienz

$V_{min}$  sei die kleinstmögliche Varianz aller Punktschätzer eines reellwertigen Parameters. Die **Effizienz** eines bestimmten Punktschätzers ist als das Verhältnis  $\frac{V_{min}}{Var(\hat{\alpha})}$  gegeben, wobei  $Var(\hat{\alpha})$  die Varianz von  $\hat{\alpha}$  für diesen Punktschätzer ist.

## Suffizienz

Jede Funktion von Messdaten  $x$  nennt man eine **Statistik**. Eine **suffiziente Statistik für  $\alpha$**  ist eine Funktion der Messdaten, die den gesamten Informationsgehalt über  $\alpha$  enthält.

In der Hochenergiephysik verwendete Punktschätzer

$p(x; \alpha)$ : Wahrscheinlichkeit, die Messwerte  $x$  bei einem gegebenem Parameter  $\alpha$  zu erhalten.

- Wenn man nun die gemessenen Werte  $x$  in die Funktion  $p(x; \alpha)$  einsetzt, erhält man eine Statistik von  $x$ , die man **Likelihood** oder **Likelihoodfunktion**  $L(x; \alpha)$  nennt.
- Man verwendet den Begriff des Likelihoods, um die Verwandtschaft mit der Wahrscheinlichkeitsdichte  $p(x; \alpha)$  anzudeuten und gleichzeitig deutlich zu machen, **dass  $L$  keine Wahrscheinlichkeitsfunktion ist.**

$f(x_k; \alpha)$  sei die Wahrscheinlichkeitsdichte für den Ausgang einer einzelnen Messung  $x_k$ . Bei  $n$  unabhängigen Messungen  $x = (x_1, \dots, x_n)$  ist dann

$$L(x_1, \dots, x_n; \alpha) = \prod_{k=1}^n f(x_k; \alpha).$$

Bei der **Methode des maximalen Likelihoods** nimmt man als Schätzwert für  $\alpha$  den Wert von  $\alpha$ , für den  $L(x; \alpha)$  maximal wird.

$n \rightarrow \infty$

- Der Punktschätzer ist konsistent.
- Der Punktschätzer ist effizient.
- $\hat{\alpha}$  ist normalverteilt.
- Wegen der Konsistenz ist der Punktschätzer asymptotisch erwartungstreu.

$n$  endlich

Um das Verhalten des Punktschätzer bei begrenzter Datenmenge  $n$  zu bestimmen, bedient man sich in der experimentellen Praxis Ensembles zufällig erzeugter simulierter Daten, auf die man den Punktschätzer anwendet.

$n$  Messungen  $x_1, \dots, x_n$ .

$E(x_k; \alpha)$ : Erwartungswert von  $x_k$  bei gegebenem  $\alpha$  („theoretische Vorhersage“ für den Wert von  $x_k$ ).

$V = (\text{cov}(x_k, x_\ell))$ : Kovarianzmatrix. Im Allgemeinen ist auch  $V$  eine Funktion von  $\alpha$ .

$$Q^2 := \sum_{k, \ell=1}^n [x_k - E(x_k; \alpha)] V_{k\ell}^{-1}(\alpha) [x_\ell - E(x_\ell; \alpha)].$$

Bei der **Methode der kleinsten Quadrate** wählt man als Schätzwert für  $\alpha$  denjenigen Wert, bei dem  $Q^2$  minimal wird.

**Bemerkung.** Wenn  $V_{k\ell}(\alpha)$  nicht beschränkt ist, kann man unsinnige Ergebnisse für  $\alpha$  erhalten. Falls z.B.  $V_{k\ell}(\alpha) \rightarrow \infty$  für  $\alpha \rightarrow \alpha_{\text{Unsinn}}$  und  $x_k - E(x_k; \alpha)$  beschränkt bleibt, liefert die Minimierung  $\alpha_{\text{Unsinn}}$ . In der Praxis minimiert man daher  $Q^2$  oftmals iterativ. Man fängt mit einem Schätzwert für  $V$  an und variiert  $V$  während der  $Q^2$ -Minimierung nicht. Danach berechnet man  $V$  neu für den erhaltenen Schätzwert von  $\alpha$  und wiederholt die Minimierung bei festgehaltenem  $V$  solange, bis sich  $\hat{\alpha}$  nicht mehr signifikant ändert.