Parameteranpassung mit kleinsten Quadraten und Maximum Likelihood

Armin Burgmeier

27. November 2009

1 Schätzwerte

1.1 Einführung

Physikalische Messungen sind immer fehlerbehaftet. Man misst niemals den wahren Wert, stattdessen werden die einzelnen Messwerte um diesen herum schwanken. Die Messwerte folgen somit einer Verteilung die sich durch ihre *Wahrscheinlichkeitsdichte* $f(\mathbf{x})$ beschreiben lässt. Im Allgemeinen hängt diese Wahrscheinlichkeitsdichte von Parametern $\mathbf{a}$ eines Modells ab das die Messdaten beschreiben soll. Je nach Wahl von $\mathbf{a}$ ändert sie sich. Zum Beispiel kann $\mathbf{a}$ Steigung und Achsenabschnitt einer Geraden sein auf der die Messwerte liegen sollen. Wir notieren dazu $f(\mathbf{x}|\mathbf{a})$, was die Wahrscheinlichkeitsdichte von $\mathbf{x}$ angibt wenn die Modellparameter gerade $\mathbf{a}$ sind.

Im einfachsten Fall ist $f(\mathbf{x}|\mathbf{a})$ symmetrisch um den wahren Wert verteilt, sodass es reicht den *Mittelwert* der Verteilung zu bestimmen um den wahren Wert zu erhalten.

Es gibt auch Situationen wo es gar keinen „wahren“ Wert gibt, sondern nur eine Verteilung. Dies kann bei quantenmechanischen Prozessen der Fall sein, oder auch wenn man sich zum Beispiel für die mittlere Schuhgröße aller Bundesbürger interessiert. Die Schuhgröße folgt einer gewissen Verteilung, aber es gibt keine „wahre Schuhgröße“.

$f(\mathbf{x}|\mathbf{a})$ selbst kann auch nicht gemessen werden. Was man misst ist lediglich eine *Stichprobe* von N Werten die dieser Verteilung folgen. Im Beispiel der Schuhgrößen bedeutet das, dass es kaum möglich sein dürfte jeden Deutschen nach seiner Schuhgröße zu fragen. Stattdessen muss eine möglich repräsentative Gruppe von N Leuten gefunden werden deren Schuhgrößen dann als Stichprobe dienen.

Man bezeichet nun als *Schätzfunktion* eine Funktion die aus solchen Stichproben einen Wert für die Modellparameter $\mathbf{a}$ berechnet. Das Ergebnis bezeichnet man als *Schätzwert* und wird in der Regel mit $\hat{\mathbf{a}}$ bezeichnet.

1.2 Kriterien für Schätzwerte

Es ist klar dass manche Schätzfunktionen besser geeignet sind als andere. Wenn man den Mittelwert einer Verteilung schätzen will ist es intuitiv dass man dazu den Mittelwert der Stichprobe als Schätzwert verwendet. Allerdings wäre es ebenfalls eine gültige Schätzfunktion wenn man einfach den größten Wert der Stichprobe als Schätzwert nehmen würde - dieser ist dann natürlich kein sonderlich guter Schätzwert. Um die Güte von Schätzwerten zu beurteilen gibt es vier Kriterien (die oft nicht alle gemeinsam erfüllt werden können):

1. **Konsistenz**

   Man nennt einen Schätzwert $\hat{\mathbf{a}}$ *konsistent* wenn er sich für größer werdende Stichproben dem wahren Wert $\mathbf{a}_0$ beliebig nahe kommt. Diese Eigenschaft ist so essentiell dass sie praktisch von allen Schätzverfahren gefordert wird.

$$\lim_{n\to\infty} \hat{\mathbf{a}} = \mathbf{a}_0 \tag{1}$$

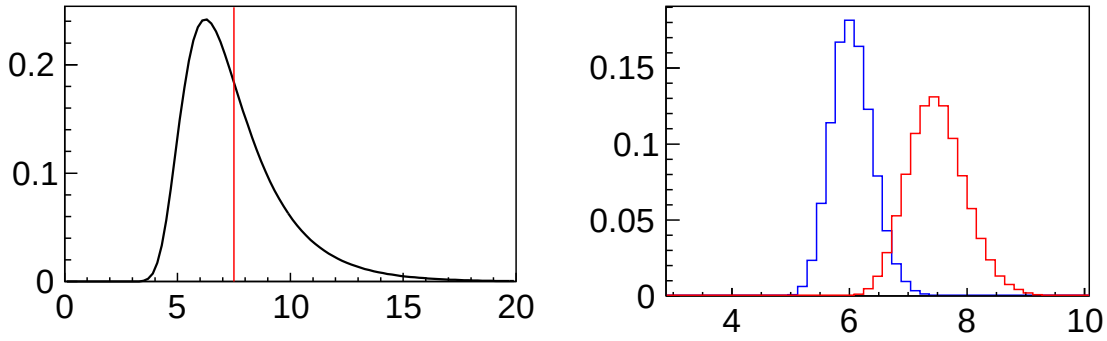

Abbildung 1: Links: Moyal-Verteilung mit Mittelwert 7,5. Rechts: Bei einer Stichprobe von 20 Messwerten zeigt die rote Kurve die Verteilung des Schätzwertes, der durch eine Mittelung der Stichprobe entsteht. Die blaue Kurve entsteht durch Weglassen der 9 größten Werte und Mittelung über die restlichen 11.

2. **Erwartungstreue**

   Ein *erwartungstreuer* Schätzer zeichnet sich dadurch aus, dass sein Erwartungswert gleich dem wahren Wert $\mathbf{a}_0$ ist. Dazu sollte man sich klarmachen, dass jeder Schätzwert wieder einer Verteilung folgt: Eine andere Stichprobe führt in der Regel zu einem anderen Schätzwert.

   $$\mathrm{E}\left[\hat{\mathbf{a}}\right] = \mathbf{a}_0 \tag{2}$$

   Abbildung 1 zeigt ein Beispiel für eine erwartungstreue und eine nicht erwartungstreue Schätzung: Bildet man den Mittelwert einer Stichprobe von 20 Werten um den Mittelwert einer Moyal-Verteilung (hier 7,5) zu schätzen, so ist der Erwartungswert dieses Schätzwertes erwartungstreu (rote Kurve). Lässt man die 9 größten Werte weg und bildet dann den Mittelwert, so erhält man eine verschobene Verteilung.

   Die Größe $b = \mathrm{E}[\hat{\mathbf{a}}] - a_0$ heißt *Verzerrung* (engl. *Bias*) und kann, wenn sie bekannt ist, korrigiert werden. Dies muss allerdings multiplikativ geschehen damit die relative Breite erhalten bleibt:

   $$\hat{\mathbf{a}}' = \frac{\mathrm{E}\left[\hat{\mathbf{a}}\right] - b}{\mathrm{E}\left[\hat{\mathbf{a}}\right]}\hat{\mathbf{a}} \tag{3}$$

3. **Effizienz**

   Die *Varianz* $\mathrm{V}\left[\hat{\mathbf{a}}\right]$ eines Schätzwerts sollte klein sein, sodass die Wahrscheinlichkeit hoch ist, dass der Schätzwert nahe am Erwartungswert $\mathrm{E}\left[\hat{\mathbf{a}}\right]$ liegt. Die *Effizienz* eines Schätzers ist definiert als das Verhältnis der kleinstmöglichen Varianz zu seiner Varianz. Die kleinstmögliche Varianz ist durch die Rao-Cramér-Frechet-Grenze gegeben.

   $$\mathrm{Eff}\left[\hat{\mathbf{a}}\right] = \frac{\mathrm{V}\left[\hat{\mathbf{a}}\right]}{\mathrm{V}_{\mathrm{min}}} \tag{4}$$

   Die blaue Kurve in Abbildung 1 hat eine größere Effizienz als die rote, da ihre relative Breite kleiner ist. Wenn man die Verzerrung also hinreichend gut kennt kann es Sinn machen, den blauen Schätzwert zu verwenden und die Verzerrung im Anschluss zu korrigieren um eine bessere Effizienz zu erhalten.

4. **Robustheit**

   Ein Schätzwert heißt *robust* wenn er von einzelnen Ausreißern (verursacht zum Beispiel durch Fehler im Messgerät) oder anderen falschen Annahmen nicht beeinflusst wird. Oft geschieht dies durch Weglassen von einer bestimmten Prozentzahl der größten und/oder kleinsten Messwerte.

Die beiden Kriterien der Robustheit und Effizienz stehen oft im Widerspruch miteinander, sodass man sich auf einen Kompromiss einigen muss der von Experiment zu Experiment unterschiedlich sein kann.

2 Maximum Likelihood

Die Maximum-Likelihood-Methode ist nun ein spezielles Schätzverfahren welches neben den kleinsten Quadraten häufig zur Anwendung kommt. Es basiert auf der *Likelihood-Funktion* die im Fall unkorrelierter Messwerte wie folgt definiert ist:

$$\mathcal{L}(\mathbf{x}|\mathbf{a}) = f_1(x_1|\mathbf{a}) \cdot f_2(x_2|\mathbf{a}) \cdot \ldots = \prod_i f_i(x_i|\mathbf{a}) \tag{5}$$

Dabei ist $f_i(x|\mathbf{a})$ die Wahrscheinlichkeitsdichte der Verteilung des i-ten Messwerts x_i, die im allgemeinen von den Modellparametern $\mathbf{a}$ abhängt. Die Likelihood-Funktion ist also eine Wahrscheinlichkeitsdichte in $\mathbf{x}$ (nicht in $\mathbf{a}$!), die angibt wie wahrscheinlich es ist einen bestimmten Satz an Messwerten $\{x_i\}$ zu erhalten. Der Schätzwert $\hat{\mathbf{a}}$ wird nach der Maximum-Likelihood-Methode nun so gewählt, dass es am wahrscheinlichsten ist, die Messwerte zu erhalten die man tatsächlich gemessen hat:

$$\mathcal{L}(\{x_i\}|\hat{\mathbf{a}}) = \text{Maximum} \tag{6}$$

In der Praxis stellt sich heraus, dass es oft einfacher ist, die *negative Log-Likelihood-Funktion* $\mathcal{F}(\mathbf{x}|\mathbf{a}) = -(2)\log\mathcal{L}(\mathbf{x}|\mathbf{a})$ zu minimieren. Da der Logarithmus eine monotone Funktion ist ändert sich die Stelle des Maximums nicht. Manchmal (zum Beispiel bei `ROOT`) wird noch ein Faktor 2 zur Definition hinzugezogen (der das Maximum auch nicht verschiebt). Im folgenden werden wir jedoch die einfache Definition ohne diesen Faktor verwenden.

2.1 Fehlerberechnung

Für große Stichproben kann man zeigen dass sich die Likelihood-Funktion einer Gaußkurve annähert. Dann wird die Log-Likelihood-Funktion zu einer Parabel und man kann sie um ihr Minimum entwickeln (wir betrachten der Einfachheit halber den Fall eines einzelnen Parameters):

$$\mathcal{F}(\mathbf{x}|a) = \mathcal{F}(\mathbf{x}|\hat{a}) + \frac{1}{2}\left.\frac{\partial^2 F}{\partial a^2}\right|_{a=\hat{a}} (a-\hat{a})^2 + \ldots \tag{7}$$

Durch den Vergleich mit dem Exponenten $\frac{(a-\hat{a})}{2\sigma^2}$ einer Gaußfunktion findet man

$$\sigma = \left(\left.\frac{\partial^2 F}{\partial a^2}\right|_{a=\hat{a}}\right)^{-1/2} \tag{8}$$

Im Mehrdimensionalen ergibt sich die Inverse der Kovarianzmatrix $\mathbf{V}$ aus der zweiten Ableitung:

$$G_{ij} = \left(\left.\frac{\partial^2 F}{\partial a_i \partial a_j}\right|_{a=\hat{a}}\right) \text{und} \mathbf{V} = \mathbf{G}^{-1} \tag{9}$$

Schaut man sich an, welchen Wert die Log-Likelihood-Funktion bei einer Abweichung von $r\sigma$ vom Schätzwert hat, so ergibt sich

$$\mathcal{F}(\mathbf{a}) = \mathcal{F}(\hat{\mathbf{a}}) + \frac{r^2}{2} \tag{10}$$

Dies definiert Konturen gleicher Likelihood (Intervalle im eindimensionalen) innerhalb derer der wahre Wert mit der zu $r\sigma$ gehörenden Wahrscheinlichkeit liegt. Das funktioniert in sehr guter Nährung auch für endliche Stichproben, ist aber nicht exakt, da dann die Wahrscheinlichkeit dafür, dass der wahre Wert $\mathbf{a}_0$ in der Region enthalten ist von $\mathbf{a}$ selbst abhängt[2].

Neyman-Konstruktion Die *Neyman-Konstruktion* erlaubt es, *Konfidenzintervalle*, oder allgemeiner *Konfidenzregionen*, zu definieren, die bei N Versuchen den wahren Wert a_0 $C \cdot N$ mal einschließen. Man sieht an dieser Definition bereits dass der Neyman-Konstruktion der frequentistische Zugang zur Statistik zugrunde liegt. C nennt man dann *Coverage*. Das Verfahren funktioniert insbesondere auch bei kleinen Stichproben,

hat aber den Nachteil dass sowohl der benötigte Rechenaufwand sehr schnell sehr groß werden kann als auch dass leere Regionen entstehen können die dann keine physikalische Aussage machen.

Für jeden Wert der Parameter **a** bildet man das Integral

$$\int \mathcal{L}(\mathbf{x}|\mathbf{a})\,\mathrm{d}\mathbf{x} = C \tag{11}$$

Liegen die Messwerte $\{x_i\}$ nun innerhalb des Integrationsgebiets, so nimmt man **a** zur Konfidenzregion hinzu, ansonsten nicht. Bei gegebenem **a** ist die Wahrscheinlichkeit dass $\{x_i\}$ im Integrationsgebiet liegt nach Definition der Likelihood-Funktion gerade C.

Es besteht allerdings noch eine Freiheit bei der Wahl des Integrationsgebiets. Dieses sollte für alle **a** auf die gleiche Weise bestimmt werden, ist ansonsten aber nicht festgelegt. Bei nur einem Messwert kann man zum Beispiel ein zentrales Intervall wählen, sodass die Fläche links und rechts davon gerade $(1 - C)/2$ beträgt. Auch üblich ist es Ober- und Untergrenzen zu wählen. Eine Möglichkeit die auch im Mehrdimensionalen funktioniert ist es, die Region so zu wählen, dass sie die höchsten Stellen der Likelihood-Funktion enthält, sodass also immer $\mathcal{L}(\mathbf{x}_{\mathrm{in}}|\mathbf{a}) \geq \mathcal{L}(\mathbf{x}_{\mathrm{out}}|\mathbf{a})$ gilt.

Bayes'sche Statistik Ein anderer Ansatz für die Fehlerberechnung bei kleinen Stichproben liefert der *Satz von Bayes*. Angewandt auf die Likelihood-Funktion ergibt

$$\mathcal{P}(\mathbf{a}|\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}|\mathbf{a}) \cdot f(\mathbf{a})}{g(\mathbf{x})} \tag{12}$$

dabei nennt man $\mathcal{P}$ den *Posterior* und f den Prior. g ist konstant in **a** und dient lediglich der Normierung. Man interpretiert nun **a** als Zufallsvariable und $\mathcal{P}(\mathbf{a})$ als *Glaube*, dass **a** einen bestimmten Wert hat wenn **x** gemessen wurde.

Man fordert für die *Kredibilitätsregion*, dass die Wahrscheinlichkeit, dass sie den wahren Wert nicht enthält, gleich einer *Irrtumswahrscheinlichkeit* α ist.

$$\int \mathcal{P}(\mathbf{a}|\{x_i\}) = 1 - \alpha \tag{13}$$

Eine Freiheit besteht neben dem Prior $f(\mathbf{a})$ auch wieder in der Wahl des Integrationsgebiets.

Diese Methode benötigt weniger Rechenaufwand (nur eine Integration erforderlich) und liefert immer ein nichtleeres Gebiet. Allerdings wird von vielen der Einfluss des Glaubens in die Statistik abgelehnt.

3 Kleinste Quadrate

Für gaußverteilte Fehler kommt man für die Log-Likelihood-Funktion auf die Form

$$S = 2 \cdot \mathcal{F} = \frac{(y_1 - a_1)^2}{\sigma_1^2} + \frac{(y_2 - a_2)^2}{\sigma_2^2} + \ldots = \sum_i \frac{(y_i - a_i)^2}{\sigma_i^2} = \sum_i \frac{\Delta y_i^2}{\sigma_i^2} \tag{14}$$

Die a_i geben dabei den erwarteten Wert für den zugehörigen Messwert y_i an. Sie werden im Normalfall durch ein Modell vorhergesagt das von vergleichsweise wenig zu bestimmenden Parametern abhängt. Die Δy_i heißen *Residuen*. Der Faktor 2 wird manchmal auch bereits in die Definition der Log-Likelihood-Funktion einbezogen (siehe Abschnitt 2).

Das Prinzip der kleinsten Quadrate besagt nun, dass die Summe der quadratischen Abweichungen, also gerade die Größe S, zu minimieren ist. Für gaußverteilte Fehler ist sie damit äquivalent zur Maximum-Likelihood-Methode. Für andere Verteilungen stellt sie ein eigenständiges Schätzverfahren dar. Anstatt der genauen Wahrscheinlichkeitsdichten müssen nun lediglich noch die Varianzen σ_i^2 der Messwerte y_i bekannt sein. Dies beschränkt das Verfahren in seiner Anwendbarkeit auf Verteilungen mit endlichen Varianzen, was in der Praxis jedoch häufig gegeben ist.

Eine weitere Eigenschaft der Methode der kleinsten Quadrate ist, dass auch Korrelationen zwischen den Messwerten einfach berücksichtigt werden können. Mit dem Residuenvektor $\Delta\mathbf{y} = \begin{pmatrix} \Delta y_1 & \Delta y_2 & \ldots \end{pmatrix}^{\mathrm{T}}$ verallgemeinert sich der Ausdruck für S zu

$$S = \Delta\mathbf{y}^{\mathrm{T}}\mathbf{W}\Delta\mathbf{y} \tag{15}$$

mit der *Gewichtsmatrix* $\mathbf{W} = \mathbf{V}^{-1}$.

3.1 Lineare kleinste Quadrate

Ein wichtiger Spezialfall ist der der linearen kleinsten Quadrate. Dabei beschreibt man die erwarteten Messwerte durch ein lineares Modell (linear in den Parametern $\mathbf{a}$) der Form

$$f\left(x|\mathbf{a}\right) = a_1 f_1\left(x\right) + a_2 f_2\left(x\right) + \ldots \tag{16}$$

das einen Messwert y an einer *Stützstelle* x vorhersagt. Dieser Fall tritt auf wenn man eine Ausgleichsgerade an eine Datenmenge anpassen will, ist aber nicht auf diesen Fall beschränkt. Zum Beispiel lassen sich auch allgemein Polynome n-ten Grades in dieser Form darstellen.

Im einfachsten Fall sind die Messwerte y_i unkorreliert und haben die gleiche Varianz σ^2. Minimieren von S führt dann für n Messwert-Paare $\{(x_i, y_i)\}$ und p Parameter $\{a_j\}$ auf die sogenannten *Normalengleichungen*:

$$a_1 \sum_i f_j\left(x_i\right) f_1\left(x_i\right) + a_2 \sum_i f_j\left(x_i\right) f_2\left(x_i\right) + \ldots = \sum_i y_i f_j\left(x_i\right) \tag{17}$$

Mit den Definitionen

$$\mathbf{a} := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}, \quad \mathbf{y} := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{A} := \begin{pmatrix} f_1\left(x_1\right) & f_2\left(x_1\right) & \ldots & f_p\left(x_1\right) \\ f_1\left(x_2\right) & f_2\left(x_2\right) & \ldots & f_p\left(x_2\right) \\ \vdots & \vdots & \ddots & \vdots \\ f_1\left(x_n\right) & f_2\left(x_2\right) & \ldots & f_p\left(x_n\right) \end{pmatrix} \in \mathbb{R}^{n\times p} \tag{18}$$

kann man die Normalengleichungen kompakt in *Matrixschreibweise* schreiben:

$$\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{a} = \mathbf{A}^{\mathrm{T}}\mathbf{y} \tag{19}$$

Neben der kompakten Darstellung hat dies den Vorteil, dass sämtliche Errungenschaften der linearen Algebra auf dieses Problem angewandt werden können und dass viele effiziente Algorithmen zu Matrixoperationen existieren.

Die Lösung kann nun einfach geschrieben werden als

$$\mathbf{a} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y} \tag{20}$$

Da $\mathbf{A}$ keine quadratische Matrix ist lässt sich dies nicht noch weiter vereinfachen wie es auf den ersten Blick scheint. Im allgemeinen Fall mit belieben (Ko)varianzen ist die Lösung

$$\mathbf{a} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y} \tag{21}$$

Fehlerfortpflanzung Für eine lineare Beziehung $\mathbf{y} = \mathbf{B}\mathbf{x}$ gilt im Allgemeinen das *Fehlerfortpflanzungsgesetz*

$$\mathbf{V}_{\mathrm{y}} = \mathbf{B}\mathbf{V}_{\mathrm{x}}\mathbf{B}^{\mathrm{T}} \tag{22}$$

Angewandt auf die Lösung der linearen Quadrate ergibt sich die Kovarianzmatrix der Parameter $\mathbf{V}_{\mathrm{a}}$ aus der Gewichtsmatrix $\mathbf{W} = \mathbf{V}^{-1}$ der Messwerte:

$$\mathbf{V}_{\mathrm{a}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\right)^{-1} \tag{23}$$

3.2 χ^2-Test

Die bisherigen Fehlerbetrachtungen haben alle die Fortpflanzung der Fehler der Messwerte in die Fehler der Parameter zum Ziel gehabt. Dabei wurde jedoch nicht berücksichtigt, wie gut das gewählte Modell zu den Daten passt. Eine Möglichkeit dies bei gaußförmigen Fehlern (die in der Praxis auch bei nicht allzu großen Abweichungen von der Form einer Gaußkurve funktioniert) zu testen bietet der χ^2-Test.

Dazu macht man sich klar, dass die Summe der Quadrate von n *standard-normalverteilten* (d.h. gaußverteilten mit Mittelwert 0 und Varianz 1) Zufallsvariablen einer χ^2-Verteilung mit n Freiheitsgraden folgt. Die Größe S ist gerade eine solche Summe, da die einzelnen Summanden durch die Verschiebung um a_i und die Division durch σ_i^2 entsprechend normiert werden. Sie folgt einer χ^2-Verteilung mit $n - p$ Freiheitsgraden, da p der n Summanden bereits verwendet wurden um die Parameter zu schätzen. Dies ist am Beispiel einer Ausgleichsgeraden leicht einzusehen: Bei nur zwei Messwerten geht die Gerade genau durch die beiden Werte hindurch, die quadratischen Abweichungen verschwinden also und es ist immer $S = 0$.

Man überlegt sich nun wie wahrscheinlich es ist, den gemessenen Wert für S oder einen *noch schlechteren*, d.h. eine noch größere Abweichung von den Modellparametern zu erhalten. Mit der Wahrscheinlichkeitsdichte $f_n\left(\chi^2\right)$ der χ^2-Verteilung mit n Freiheitsgraden ergibt sich diese Wahrscheinlichkeit durch

$$P\left(S\right) = \int_S^\infty f_{n-p}\left(\chi^2\right) \mathrm{d}\chi^2 \tag{24}$$

Man sollte sich nun vorher überlegen welche Wahrscheinlichkeit man noch akzeptiert (Eine Wahrscheinlichkeit von 1% hieße zum Beispiel dass man nur in einem von 100 Fällen so „schlechte“ Messwerte erhalten wird wie die die man vorliegen hat). Genauso sollte man sich Gedanken machen wenn die Wahrscheinlichkeit dafür, den gemessenen Wert für S oder einen *noch besseren* zu erhalten sehr klein ist, da $f_n\left(\chi^2\right)$ für $n > 2$ für $\chi^2 \to 0$ gegen 0 geht. Der Erwartungswert der χ^2-Verteilung ist $\mathrm{E}\left[\chi_n^2\right] = n$.

Kann man das Experiment mehrmals durchführen oder mit Monte-Carlo-Methoden simulieren, so ist es empfehlenswert in einem Histogramm die relative Häufigkeit über $P\left(S\right)$ (siehe Gleichung 24) aufzutragen. Bei gaußförmigen Fehlern und einem Modell das die Daten richtig beschreibt erwartet man eine Gleichverteilung. Andernfalls ist das Modell falsch, die Messwerte anders als angenommen verteilt oder bestehende Korrelationen nicht berücksichtigt.

4 Zusammenfassender Vergleich

	Maximum Likelihood	**Kleinste Quadrate**
Input	PDFs	Varianzen
Anwendbarkeit	PDFs bekannt	Daten unverzerrt, Varianzen endlich
Konsistent	Ja	Ja
Erwartungstreu	nur für große Stichproben	Ja im linearen Fall
Effizient	asymptotisch maximal	maximal (im linearen Fall)
Robust	Nein	Nein
Goodness-of-fit	Nein	Für gaußsche Fehler
Rechenaufwand	im Allgemeinen hoch	Im linearen Fall relativ gering

Tabelle 1: Vergleich der Eigenschaften von mit Maximum Likelihood und kleinsten Quadraten erhaltenen Schätzwerten

Literatur

[1] Volker Blobel and Erich Lohrmann. *Statistische und numerische Methoden der Datenanalyse.* Teubner Verlag, 1 edition, 1998.

[2] C. Amsler et al. Review of particle physics. *Phys. Lett.*, B667:1, 2008.