Kapitel 4

Stichproben und Schätzungen

4.1 Stichproben, Verteilungen und Schätzwerte

Eine physikalische Messung ist eine endliche Stichprobe aus einer Grundgesamtheit, die endlich oder unendlich sein kann. Im allgemeinen möchte man bei der Weiterverarbeitung der Messergebnisse eine **Reduktion der Daten** auf die wesentliche Information erreichen. Diese Information steckt in der mathematischen Beschreibung der Verteilung der Grundgesamtheit, die durch – hoffentlich endlich viele – Parameter beschrieben werden kann. Man versucht nun die Verteilungen zu bestimmen, indem man Schätzwerte für diese Parameter aus der Messung ableitet. Eine allgemeine Methode zur Schätzung von Parametern ist die Maximum-Likelihood-Methode (Kapitel 6).

Zum Beispiel weiss man beim radioaktiven Zerfall,

$$N(t) = N_0 e^{-\lambda t}, \tag{4.1}$$

dass der einzige Parameter die Zerfallswahrscheinlichkeit (oder mittlere Lebensdauer) λ ist, die man als Mittelwert aus der gemessenen Häufigkeitsverteilung $N(t)$ bestimmt. Die Messwerte haben sonst keine weitere wesentliche Information (wenn man weiss, dass sie einem Zerfallsgesetz folgen).

Eine Stichprobe von n Messungen aus einer Grundgesamtheit mit der Wahrscheinlichkeitsdichte $f(x)$

$$\vec{x} = (x_1, \ldots, x_n) \tag{4.2}$$

kann man als eine n-dimensionale Zufallsvariable auffassen und ihr eine Wahrscheinlichkeitsdichte

$$g(\vec{x}) = g(x_1, \ldots, x_n) \tag{4.3}$$

zuordnen (siehe Beispiel 1 in Abschnitt 3.4). Damit die **Stichprobe zufällig** ist, muss gelten:

(i) Die x_i sind unabhängig

$$\Longrightarrow \quad g(\vec{x}) = g_1(x_1) \cdot g_2(x_2) \ldots g_n(x_n) \tag{4.4}$$

(ii) Jeder Messwert x_i hat die Wahrscheinlichkeitsdichte der Grundgesamtheit:

$$g_i(x_i) = f(x) \tag{4.5}$$

Diese Eigenschaften sind durchaus nicht immer gegeben. Zum Beispiel ändert sich die Wahrscheinlichkeitsdichte, wenn man aus einer endlichen Grundgesamtheit Stichproben entnimmt ohne zurückzulegen (Karten aus einem Kartenstapel usw.).

4.2 Eigenschaften von Schätzwerten

Schätzwerte S sind Funktionen der Messwerte (Stichprobenfunktion):

$$S = S(x_1, \dots, x_n), \tag{4.6}$$

und sind damit selbst wieder Zufallsvariable (die nächste Messreihe ergibt im allgemeinen ein etwas anderes Resultat für S). Als Beispiel hatten wir in Abschnitt 3.4 (Beispiel 1) das arithmetische Mittel als Zufallsvariable behandelt:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.7}$$

Es gibt gewisse Freiheiten Schätzwerte zu definieren. Optimale Eigenschaften von Schätzwerten erhält man mit folgenden Forderungen:

1. **Erwartungstreue:** Unabhängig von der Anzahl der Messwerte soll der Erwartungs des Schätzwerts für einen Parameter λ gleich dem Parameter sein:

$$E(S_\lambda(x_1, \dots, x_n)) = \lambda \tag{4.8}$$

In Abschnitt 3.4 (Beispiel 1) hatten wir gesehen, dass das arithmetische Mittel in diesem Sinne erwartungstreu (unverzerrt, unbiased) ist.

Beispiel: Als weiteres Beispiel wollen wir die Varianz einer Verteilung mit Mittelwert μ und Varianz σ aus einer Stichprobe abschätzen. Dazu betrachten wir zunächst den Erwartungswert der quadratischen Abweichungen vom Mittelwert der Stichprobe:

$$\begin{aligned} E\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) &= E\left(\sum_{i=1}^n (x_i - \mu + \mu - \bar{x})^2\right) \\ &= \sum_{i=1}^n \left[\underbrace{E\left((x_i-\mu)^2\right)}_{\sigma^2} - \underbrace{E\left((\bar{x}-\mu)^2\right)}_{\sigma^2/n}\right] \\ &= n\left[\sigma^2 - \frac{\sigma^2}{n}\right] = (n-1)\,\sigma^2 \end{aligned} \tag{4.9}$$

$$\Longrightarrow \frac{1}{n-1} E\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) = \sigma^2 \tag{4.10}$$

Dabei wurde für die Varianz des Mittelwertes der Stichprobe, σ^2/n, das Ergebnis von (3.45) benutzt. Der Ausdruck

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \tag{4.11}$$

auch empirische Varianz genannt, ist also eine erwartungstreue Schätzung der Varianz der Verteilung, weil für alle n gilt:

$$E(s^2) = \sigma^2. \tag{4.12}$$

Interpretation des Faktors $1/(n-1)$: Aus den n unabhängigen Messungen wurde zunächst der Parameter $\bar{x}$ bestimmt, dadurch geht ein Freiheitsgrad für die Bestimmung weiterer Parameter verloren. Die Anzahl der Freiheitsgrade ist die Anzahl der unabhängigen Messungen minus der Anzahl der bestimmten Parameter, hier also $n_F = n - 1$. Aus der zweiten Zeile in (4.9) sieht man auch, dass die Minderung der Freiheitsgrade mit der Varianz σ^2/n des geschätzten Mittelwertes zusammenhängt.

2. **Konsistenz:** Eine Schätzung wird konsistent genannt, wenn die Varianz des Schätzwertes für große Stichproben gegen Null geht:

$$\lim_{n\to\infty} \sigma^2(S(x_1, \ldots, x_n)) = 0 \tag{4.13}$$

Beispiel: Für die Schätzung der Varianz des arithmetischen Mittels einer Stichprobe hatten wir in Abschnitt 3.4 (Beispiel 1) gefunden:

$$\sigma^2(\bar{x}) = \frac{\sigma^2(x)}{n} \tag{4.14}$$

Das arithmetische Mittel ist damit einen konsistente Schätzung des Mittelwertes der Verteilung.

3. **Effektivität:** Es seien S_1 und S_2 zwei Schätzungen des gleichen Parameters λ. Man sagt, S_1 ist effektiver als S_2, wenn gilt:

$$E\left[(S_1 - \lambda)^2\right] = \sigma^2(S_1) < E\left[(S_2 - \lambda)^2\right] = \sigma^2(S_2) \tag{4.15}$$

Diejenige Schätzung S_i, für die die Varianz minimal wird, nutzt also die vorhandenen Information am effektivsten.

Beispiel: Die Stichprobenfunktionen

$$S = \sum_{i=1}^{n} a_i\, x_i \qquad \text{mit } \sum_{i=1}^{n} a_i = 1 \tag{4.16}$$

sind für sonst beliebige a_i erwartungstreue Schätzungen des Mittelwertes μ:

$$E(S) = E\left(\sum_{i=1}^{n} a_i\, x_i\right) = \sum_{i=1}^{n} a_i\, E(x_i) = \sum_{i=1}^{n} a_i\, \mu = \mu \tag{4.17}$$

Es stellt sich aber heraus, dass S für $a_i = 1/n$ minimale Varianz hat, wenn alle Varianzen gleich sind, $\sigma_i = \sigma$ für alle x_i. Dann ergibt sich:

$$\sigma^2(S) = \sum_{i=1}^{n} a_i^2\, \sigma^2(x_i) = \sigma^2(x) \sum_{i=1}^{n} a_i^2 \tag{4.18}$$

Es bleibt also zu zeigen, dass $A = \sum a_i^2$ für $a_i = 1/n$ minimal wird. Durch die Bedingung $A = 1$ sind nur $n-1$ der a_i unabhängig, so dass sich ein a_i eliminieren lässt:

$$A = \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n-1} a_i^2 + \left(1 - \sum_{i=1}^{n-1} a_i\right)^2 \tag{4.19}$$

Die Extremwertbedingung ergibt:

$$\frac{\partial A}{\partial a_i} = 2\,a_i - 2\underbrace{\left(1 - \sum_{i=1}^{n-1} a_i\right)}_{a_n} = 2\,(a_i - a_n) = 0 \implies a_i = a_n \implies a_i = \frac{1}{n} \quad \forall\, i \tag{4.20}$$

4. **Robustheit:** Die Schätzwerte sollen möglichst gegen Annahmen "falscher" Verteilungen stabil sein.

   Zum Beispiel sind apparative Auflösungen nicht immer gauss-förmig, sondern haben zusätzlich zu einem Gauss-Anteil "nicht-gaussische" Ausläufer. Um stabile Parameter für Mittelwert und Auflösung zu erhalten, hilft häufig ein Abschneiden nach oben und unten (zum Beispiel könnte man die jeweils 20% kleinsten und größten Werte einer Messung wegschneiden). Eine andere Möglichkeit ist, die Verteilung der Messwerte mit einer angenommenen Verteilung in einem begrenzten Bereich anzupassen. Zum Beispiel passt man häufig Gauss-Kurven an Auflösungsverteilungen innerhalb 1 bis 2 Standardabweichungen um den Mittelwert an.

   Beispiel: In den meisten Teilchenexperimenten werden Energieverlustmessungen (dE/dx) zur Identifikation der Teilchen durchgeführt. Da die Fluktuationen sehr groß sein können und die dE/dx-Verteilung ('Landau-Verteilung') lange Ausläufer zu hohen Energien hat, werden sehr viele Messungen gemacht, manchmal einige hundert, und dann gemittelt. Der Mittelwert wird deutlich stabiler, wenn man zum Beipiel die kleinsten 10% und die größten 20% der Messwerte wegschneidet ('truncated mean').

   Robustheit ist schwieriger als die anderen Kriterien für Schätzungen zu behandeln, weil man hier "Unwissen" zu berücksichtigen versucht.

4.3 Stichproben aus Normalverteilungen; χ^2-Verteilung

Wir betrachten Stichproben $(x_1, \ldots, x_n)$ vom Umfang n aus einer Normalverteilung

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \tag{4.21}$$

mit Mittelwert μ und Standardabweichung σ. Dann folgt die Stichprobenfunktion

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \tag{4.22}$$

folgender Verteilung ($\lambda = n/2$):

$$f(\chi^2) = \frac{1}{\Gamma(\lambda)\, 2^\lambda}\, (\chi^2)^{\lambda-1}\, e^{-\frac{\chi^2}{2}}. \tag{4.23}$$

Die Γ-Funktion ist tabelliert zu finden. Mit der Definition

$$\Gamma(x+1) = \int_0^\infty t^x\, e^{-t}\, dt \tag{4.24}$$

findet man folgende Eigenschaften:

$$\begin{aligned} \Gamma(1) &= 1 \\ \Gamma(x+1) &= x\,\Gamma(x) \\ \Gamma(n+1) &= n! \qquad \text{(n ganzzahlig)} \end{aligned} \tag{4.25}$$

Der Beweis, dass die in (4.22) definierte Größe χ^2 der Verteilung (4.23) folgt, ist zum Beispiel in [1] nachzulesen.

Erwartungswert und Varianz der χ^2-Verteilung: Den **Erwartungswert von** χ^2 erhält man aus (4.22) mit $\sigma^2 = E((x_i - \mu)^2)$:

$$E(\chi^2) = n, \tag{4.26}$$

wobei n hier die Anzahl der Messungen und im allgemeinen die Anzahl der Freiheitsgrade ist.

In den meisten Fällen ist der Parameter μ in der χ^2-Funktion (4.22) nicht bekannt und wird durch den Mittelwert der Stichprobe $\bar{x}$ geschätzt. Die χ^2-Funktion wird damit entsprechend definiert:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} \tag{4.27}$$

Mit der empirischen Varianz s^2 ergibt sich:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} = (n-1)\, \frac{s^2}{\sigma^2} \tag{4.28}$$

Da der Erwartungswert von s^2 nach (4.12) gleich σ^2 ist, ist der Erwartungswert der χ^2-Funktion bezüglich $\bar{x}$:

$$E(\chi^2) = n - 1 = n_F \tag{4.29}$$

Im allgemeinen wird in (4.27) $\bar{x}$ der Erwartungswert der Messgröße x_i sein, der eventuell von mehreren geschätzten Parametern abhängt, zum Beispiel wenn an die x_i eine Ausgleichsfunktion angepasst wird (siehe nächstes Kapitel). Die Anzahl der

Freiheitsgrade ist dann allgemein die Anzahl der Messwerte minus die Anzahl n_P der aus der Stichprobe bestimmten Parameter:

$$n_F = n - n_P \tag{4.30}$$

Die **Varianz von** χ^2 ist [1]:

$$\sigma^2(\chi^2) = E\left((\chi^2)^2\right) - \left(E(\chi^2)\right)^2 = 2n. \tag{4.31}$$

Hier wie im folgenden soll $n = n_F$ als Anzahl der Freiheitsgrade, der Parameter der χ^2-Verteilung, verstanden werden.

Eigenschaften der χ^2-Verteilung: Beispiele von χ^2-Verteilungen für verschiedene n sind in Abb. 4.1 gezeigt. Bei $\chi^2 = 0$ findet man folgendes Verhalten:

$$\begin{array}{llcll} n = 1: & f(\chi^2) & = & \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\chi^2}} e^{-\frac{\chi^2}{2}} & \rightarrow \quad \infty \quad \text{für } \chi^2 \rightarrow 0 \\ n = 2: & f(0) & = & \frac{1}{2} & \\ n \geq 3: & f(0) & = & 0 & \end{array} \tag{4.32}$$

Für $n = 1$ hat die χ^2-Verteilung also einen Pol bei $\chi^2 = 0$. Die Verteilungsfunktion $F(\chi^2)$ bleibt aber endlich.

Für große n wird die Verteilung zunehmend symmetrischer und geht, entsprechend dem ‘zentralen Grenzwertsatz’ (Abschnitt 2.6), in eine Normalverteilung mit $\mu = n$ und $\sigma = \sqrt{2n}$ über.

Stichproben aus nicht gleichen Normalverteilungen: Gegenüber (4.22) und (4.28) kann man die χ^2-Funktion auch auf Messwerte mit unterschiedlichen Erwartungswerten μ_i bzw. $\bar{x}_i$ und Standardabweichungen σ_i verallgemeinern:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \tag{4.33}$$

Das ist leicht einzusehen, weil die reduzierten Variablen

$$x_i' = \frac{x_i - \mu_i}{\sigma_i} \tag{4.34}$$

alle der gleichen Normalverteilung $N(0,1)$ mit $\mu = 0$ und $\sigma = 1$ folgen.

Der χ^2-Test: Die Stichprobenfunktion χ^2 wird zum Testen der Zuverlässigkeit (‘confidence level’) einer Messung benutzt. Man erwartet, dass jeder Freiheitsgrad im Mittel eine Varianz σ^2 hat, also eine Einheit zum χ^2 beiträgt:

$$\chi^2/n_f \approx 1 \tag{4.35}$$

Größere Abweichungen von dieser Erwartung deuten darauf hin, dass das angenommenen Gauss-Model oder die Schätzung der Parameter μ, σ für die Daten nicht richtig sind oder dass es einen nicht-gaussischen Untergrund gibt.

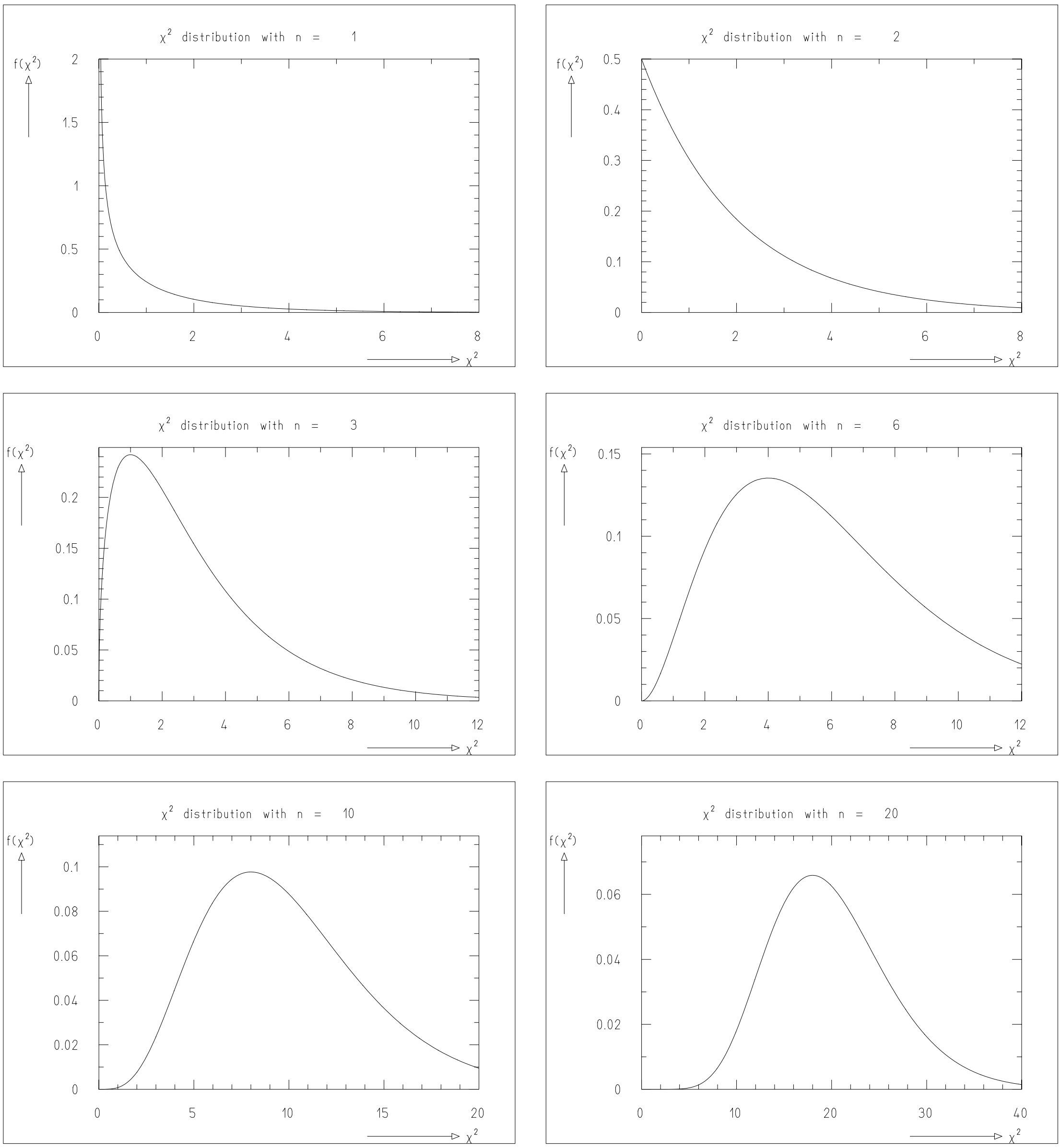

Abbildung 4.1: χ^2-Verteilungen für verschiedene Freiheitsgrade n (erstellt mit dem Programm s2sd [1]).

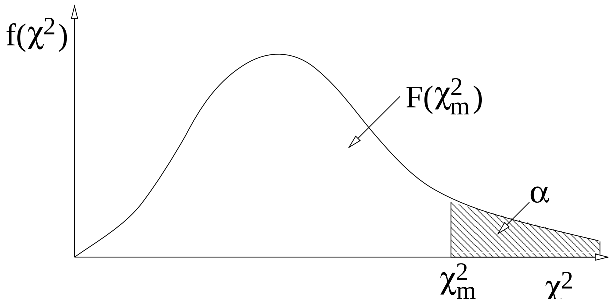

Abbildung 4.2: Definition des Vertrauensniveaus α für einen gemessenen χ^2-Wert χ^2_m.

Quantitativ gibt man die Zuverlässigkeit einer Messung beziehungsweise den Grad der Übereinstimmung mit dem Gauss-Modell durch Angabe des Vertrauensniveaus an. Wie in Abschnitt 2.5 definieren wir das Vertrauensniveau α als das Integral über die χ^2-Verteilung oberhalb des gemessenen χ^2-Wertes χ^2_m (Abb. 4.2):

$$\alpha = 1 - F(\chi^2_m), \tag{4.36}$$

wobei F die Verteilungsfunktion ist. Das Vertrauen in die Messung wird also größer, wenn das gemessene χ^2 kleiner wird. Bei welchem χ^2-Wert ein bestimmtes Vertrauensniveau erreicht wird, hängt von der Anzahl der Freiheitsgrade n_F ab. Man findet diese Angaben in Tabellen und graphischen Darstellungen (Abb. 4.3 und 4.4).

Die Wahrscheinlichkeitsdichte von $F(\chi^2)$ und damit auch von $\alpha = 1 - F(\chi^2)$ ist gleichverteilt zwischen 0 und 1 Die Stichprobenfunktionen $F(\chi^2)$ und α sind dabei als Zufallsvariable zu betrachten. Wenn man sehr viele Messungen gemacht hat, die einen χ^2-Tests erfüllen sollen, kann man die gemessene α-Verteilung graphisch darstellen (Abb. 4.5). Abweichungen von einer Gleichverteilung haben meistens folgende Ursachen:

- das Gauss-Modell ist falsch oder
- die Standardabweichungen σ_i sind zu groß ($\Rightarrow$ Verschiebung zu großen α) oder
- die Standardabweichungen σ_i sind zu klein ($\Rightarrow$ Verschiebung zu kleinen α) oder
- es gibt nicht-gaussischen Untergrund.

Der Untergrund häuft sich bei kleinen Werten von α und kann mit einem Schnitt auf α entfernt werden (typische Forderung: $\alpha > O(1\%)$).

> **Beispiel:** In Teilchenreaktionen werden in der Regel die Impulse und Richtungen der der beobachteten Teilchen mit gewissen Fehlern gemessen. Zusammen mit einer Hypothese für die Massen kann man Impuls- und Energieerhaltung mit einem χ^2-Test überprüfen. Ereignisse, bei denen wenigstens ein Teilchem dem Nachweis entgangen ist, werden sich bei einem kleinen Vertrauensniveau α ansammeln.

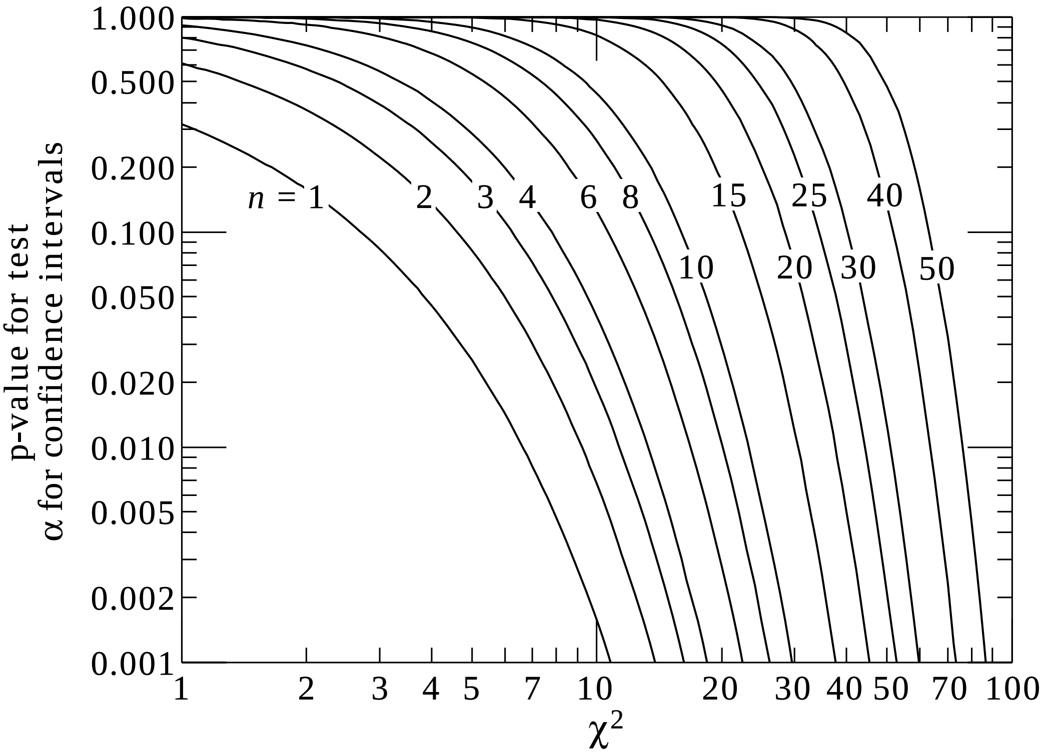

Abbildung 4.3: Das Vertrauensniveau α als Funktion des gemessenen χ^2-Wertes für verschiedene Freiheitsgrade $n = n_F$ (aus PDG [15]).

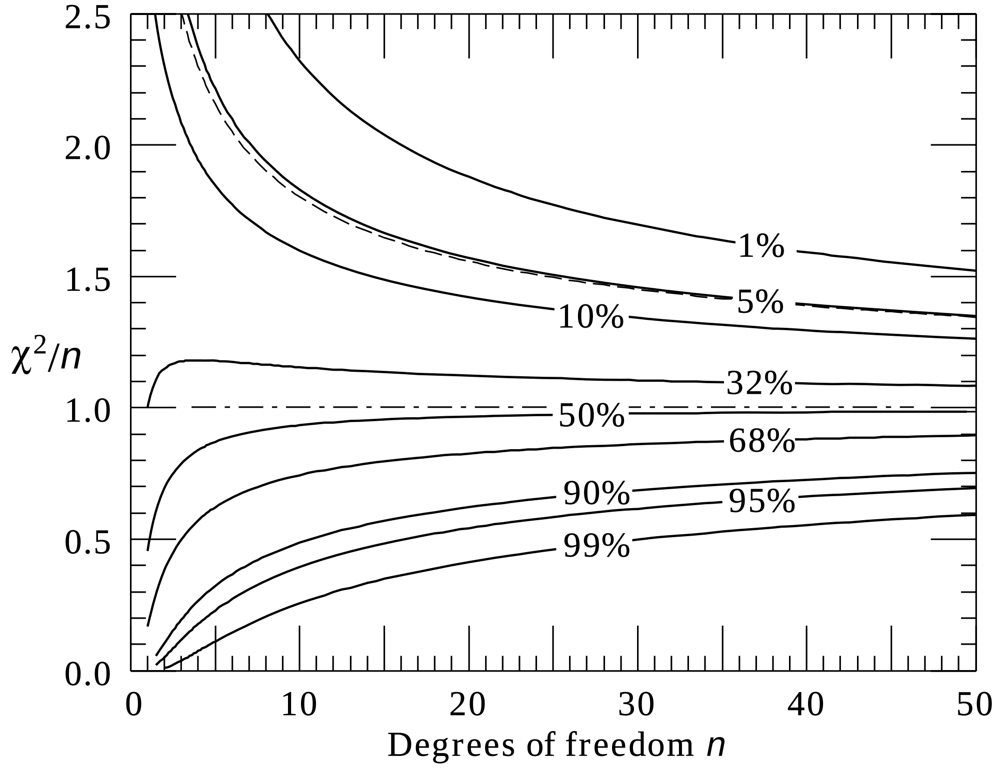

Abbildung 4.4: Das “reduzierte χ^2”, χ^2/n_F, für verschiedene Vertrauensniveaus α als Funktion des Freiheitsgrades $n = n_F$. Für grosse n_F geht die $\alpha = 50\%$-Kurve asymptotisch gegen 1, das heisst, die χ^2-Verteilung wird immer symmetrischer (aus PDG [15]).

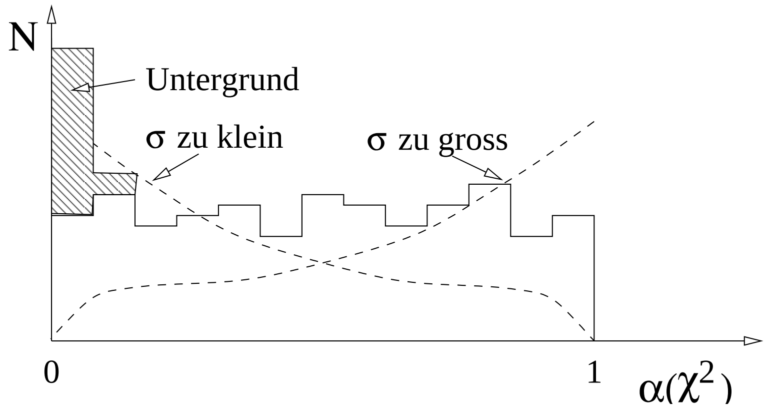

Abbildung 4.5: Typische Verteilung des Vertrauensniveaus α. Über- oder Unterschätzungen der Fehler führen zu Abweichungen von der Gleichverteilung. Der Untergrund sammelt sich nahe $\alpha = 0$

Man sollte sich klar machen, dass grundsätzlich alle Werte von α gleich häufig auftreten. Es ist also nicht von vornherein ein Wert von α nahe 1 besser als einer nahe 0. Selektionsschnitte auf α sollten ausschließlich durch das Untergrundverhalten bestimmt sein.

Die Bestimmung von Vertrauensintervallen wird im Zusammenhang mit Maximum-Likelihood-Schätzungen (Kapitel 6) noch einmal aufgegriffen.