2.4 Genauigkeit der Schätzwerte

Wir haben Schätzwerte für Parameter gewonnen, indem wir das Maximum der Posteriorverteilung für die Parameter bestimmt haben. Nun interessiert uns die Genauigkeit der Schätzwerte. Generell wird die Genauigkeit der Schätzwerte durch die Breite der Posteriorverteilung für die Parameter bestimmt.

Die Posteriorverteilung für die Parameter hat im hier betrachteten Fall die Form

$$\mathrm{pdf}(\theta|\{(x_i, y_i)\}, \{\sigma_i\}) \propto \exp\left(-\frac{1}{2}\chi^2(\theta)\right).$$

Wir beschreiben die Breite dieser Verteilung (per Konvention) durch die Konturlinie, entlang der diese Funktion auf $e^{-1/2} = 60.6\%$ ihres Peakwertes abgefallen ist. Das ist in Abb. 1 dargestellt.

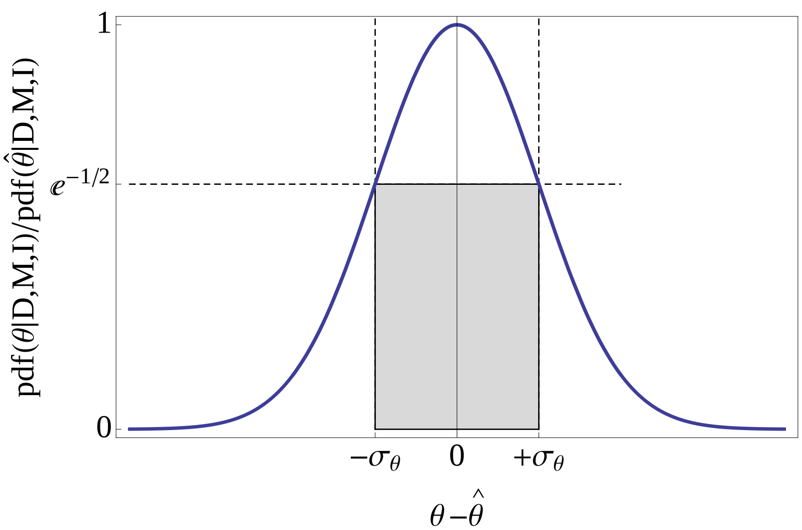

Abbildung 1: Plot der normierten Posteriorverteilung und der daraus zu bestimmenden Unsicherheit des Parameters.

Bemerkung 30. *Besitzt das Modell nur einen Parameter, so besteht die 'Konturlinie' aus zwei Punkten, die sich vom Maximum aus auf unterschiedlichen Seiten befinden. Besitzt das Modell zwei Parameter, so ist die Konturlinie eine geschlossene Linie in der zweidimensionalen Parameterebene. Bei Modellen mit $M > 2$ Parametern ist die 'Konturlinie' eine geschlossene Hyperfläche im M-dimensionalen Parameterraum.*

Ist der Schätzwert der Parameter durch $\hat{\theta}$ gegeben und das entsprechende Minimum $\chi^2(\hat{\theta}) = \chi^2_{\min}$, dann ist das Maximum der Posteriorverteilung

$$\mathrm{pdf}(\hat{\theta}|\{(x_i, y_i)\}, \{\sigma_i\}) \propto \exp\left(-\frac{1}{2}\chi^2_{\min}\right).$$

Parameterwerte θ auf der gesuchten Konturlinie erfüllen die Gleichung

$$\frac{\exp\left(-\frac{1}{2}\chi^2(\theta)\right)}{\exp\left(-\frac{1}{2}\chi^2_{\min}\right)} = \exp\left(-\frac{1}{2}\right) \Rightarrow \boxed{\chi^2(\theta) - \chi^2_{\min} = 1.}$$

Beispiel 6. *Im Fall der konstanten Fitfunktion hatten wir*

$$\chi^2(\mu) = w(\mu - \langle y_i \rangle)^2 + w\mathrm{Var}(y_i),$$

was zum Schätzwert

$$\hat{\mu} = \langle y_i \rangle$$

(gewichteter empirischer Mittelwert) führte. Somit ist

$$\chi^2_{\min} = w\mathrm{Var}(y_i)$$

und die Bestimmungsgleichung für die Genauigkeit des Schätzwertes ist

$$w(\mu - \langle y_i \rangle)^2 = 1.$$

Diese Gleichung besitzt die beiden exakten Lösungen

$$\mu = \langle y_i \rangle \pm \frac{1}{\sqrt{w}}.$$

Bemerkung 31. *Sind in diesem Beispiel alle Standardabweichungen σ_i gleich σ, dann ist*

$$w = \frac{N}{\sigma^2}, \quad \textit{und damit} \quad \boxed{\mu = \langle y_i \rangle \pm \frac{\sigma}{\sqrt{N}}.} \tag{2}$$

Die Grösse $\sigma/\sqrt{N}$ wird als Fehler des Mittelwerts bezeichnet und muss strikt von der Standardabweichung σ der einzelnen Messpunkte unterschieden werden. Das Ergebnis zeigt, dass die Genauigkeit des Schätzwertes $\langle y_i \rangle$ mit zunehmender Zahl N von Messpunkten zunimmt (d.h. die Unsicherheit über den Schätzwert nimmt proportional zu $1/\sqrt{N}$ ab.

Beispiel 7. *Im Fall der linearen Regression hatten wir*

$$\chi^2(\alpha, \beta) = w\mathrm{Var}(x_i)\left[\alpha - \frac{\mathrm{Cov}(x_i, y_i)}{\mathrm{Var}(x_i)}\right]^2 + w(\beta - \langle y_i \rangle)^2 + w\mathrm{Var}(y_i)(1 - \rho^2),$$

mit den Schätzwerten

$$\hat{\alpha} = \frac{\mathrm{Cov}(x_i, y_i)}{\mathrm{Var}(x_i)}, \quad \text{und} \quad \hat{\beta} = \langle y_i \rangle..$$

Somit ist

$$\chi^2_{\min} = w\mathrm{Var}(y_i)(1 - \rho^2)$$

und die Bestimmungsgleichung für die Genauigkeit des Schätzwertes ist

$$w\mathrm{Var}(x_i)\left[\alpha - \frac{\mathrm{Cov}(x_i,y_i)}{\mathrm{Var}(x_i)}\right]^2 + w(\beta - \langle y_i\rangle)^2 = 1$$

Diese Gleichung beschreibt eine Ellipse in der Parameterebene, deren Halbachsen parallel zu den Achsen der beiden Parameter α und β ausgerichtet sind. In diesem Fall bezeichnen wir die beiden Parameter α und β als unkorreliert.

Die Halbachse der Ellipse in α-Richtung hat die Länge $\sigma_\alpha = \sqrt{w\mathrm{Var}(x_i)}$, die Halbachse in β-Richtung hat die Länge $\sigma_\beta = 1/\sqrt{w}$. Wir geben die Fehler der beiden Schätzwerte in der Form

$$\alpha = \frac{\mathrm{Cov}(x_i,y_i)}{\mathrm{Var}(x_i)} \pm \sqrt{\frac{1}{w\mathrm{Var}(x_i)}} \quad \textit{und} \quad \beta = \langle y_i\rangle \pm \sqrt{\frac{1}{w}}$$

an. Dieses Ergebnis ist exakt.

Bemerkung 32. *Wir finden nun eine Variante der obigen Rechnung, die sich verallgemeinern lässt. Im vorangegangenen Beispiel kann man die Rechnung auch folgendermassen durchführen. Wir entwickeln*

$$\chi^2(\theta) - \chi^2_{\min} = \frac{1}{2}\frac{\partial^2\chi^2(\alpha,\beta)}{\partial\alpha^2}(\alpha-\hat\alpha)^2 + \frac{\partial^2\chi^2(\alpha,\beta)}{\partial\alpha\partial\beta}(\alpha-\hat\alpha)(\beta-\hat\beta) + \frac{1}{2}\frac{\partial^2\chi^2(\alpha,\beta)}{\partial\beta^2}(\beta-\hat\beta)^2$$

in eine Taylorreihe um das Mimimum. Die zweiten Ableitungen der Funktion χ^2 an der Stelle des Minimums bilden die Hesse-Matrix

$$H = \frac{1}{2}\left.\begin{pmatrix} \frac{\partial^2\chi^2(\alpha,\beta)}{\partial\alpha^2} & \frac{\partial^2\chi^2(\alpha,\beta)}{\partial\alpha\partial\beta} \\ \frac{\partial^2\chi^2(\alpha,\beta)}{\partial\alpha\partial\beta} & \frac{\partial^2\chi^2(\alpha,\beta)}{\partial\beta^2} \end{pmatrix}\right|_{\hat\alpha,\hat\beta},$$

wobei wir den Faktor $1/2$ aus der Taylorentwicklung in die Hesse-Matrix genommen haben. Dann ist

$$\chi^2(\theta) - \chi^2_{\min} = (\alpha-\hat\alpha, \beta-\hat\beta) H \begin{pmatrix} \alpha-\hat\alpha \\ \beta-\hat\beta \end{pmatrix}.$$

In unserem Beispiel ist die Hesse-Matrix von $\chi^2(\alpha,\beta)$ am Minimum

$$H = w\begin{pmatrix} \mathrm{Var}(x_i) & 0 \\ 0 & 1 \end{pmatrix}.$$

Die Kovarianzmatrix für die Parameter ist als die Inverse der Hesse Matrix definiert, also

$$K = H^{-1}.$$

In unserem Beispiel finden wir

$$K = \frac{1}{w\mathrm{Var}(x_i)}\begin{pmatrix} 1 & 0 \\ 0 & \mathrm{Var}(x_i) \end{pmatrix} = \begin{pmatrix} \sigma^2_\alpha & \sigma^2_{\alpha\beta} \\ \sigma^2_{\alpha\beta} & \sigma^2_\beta \end{pmatrix}.$$

Die Diagonalelemente der Kovarianzmatrix sind die Quadrate der gesuchten Längen der Halbachsen. Die Nebendiagonalelemente nennt man die Kovarianzen der Parameter. In unserem Fall sind sie Null, weil die beiden Parameter α und β unkorreliert sind.

Bemerkung 33. *Auch im Beispiel der linearen Regression kann der Fall auftreten, dass alle σ_i gleich σ sind. Dann ist*

$$\alpha = \frac{\mathrm{Cov}(x_i, y_i)}{\mathrm{Var}(x_i)} \pm \sqrt{\frac{\sigma^2}{N\mathrm{Var}(x_i)}} \quad \textit{und} \quad \beta = \langle y_i \rangle \pm \sqrt{\frac{\sigma^2}{N}}. \tag{3}$$

Auch hier verringert sich die Unsicherheit über die Parameterwerte proportional zu $1/\sqrt{N}$.

Im Fall allgemeiner Fitfunktionen muss die $\chi^2(\theta)$-Funktion nicht notwendigerweise eine quadratische Form der Parameter θ sein. In diesem Fall gibt es zwei Varianten die Genauigkeit der Parameter zu bestimmen.

Variante I: graphische Bestimmung; geeignet bei einem oder zwei Parametern. Im Fall eines einzigen Parameters plottet man die Funktion $\chi^2(\theta) - \chi^2_{\mathrm{min}}$ gegen θ und bestimmt aus dem Graphen die Parameterwerte, bei denen die Funktion Eins wird. Abbildung 2 zeigt, wie ein solcher Plot aussieht.

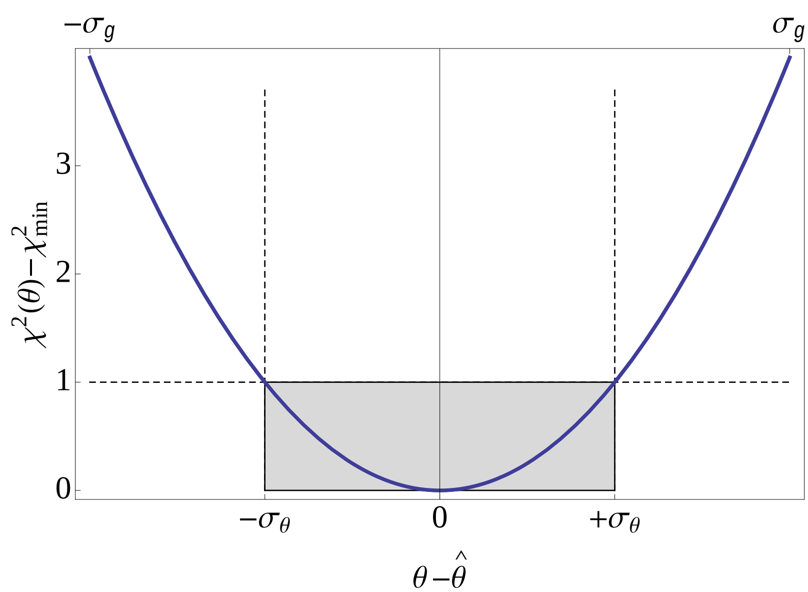

Abbildung 2: Plot der Funktion $\chi^2(\theta)-\chi^2_{\mathrm{min}}$ und der daraus zu bestimmenden Unsicherheit des Parameters.

Im Fall von zwei Parametern plottet man einen zweidimensionalen Konturplot der Funktion, in den man die Konturlinie Eins einzeichnen lässt. Ein Beispiel ist in Abb. 3 dargestellt.

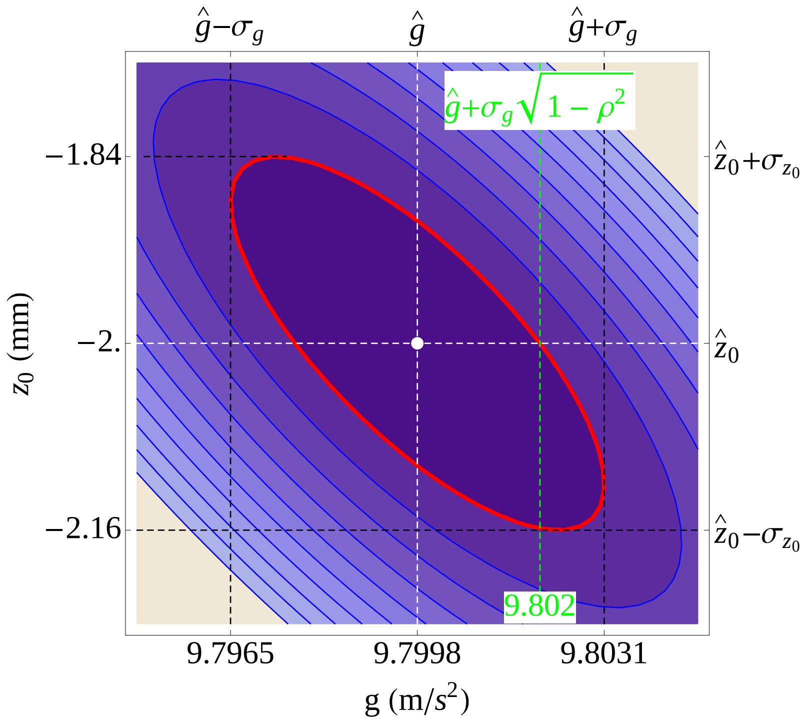

Abbildung 3: Konturplot der Funktion $\chi^2(\theta) - \chi^2_{\text{min}}$ bei zwei Parametern und der daraus zu bestimmenden Unsicherheit der Parameter.

Variante II: Näherungsweise Bestimmung mit Hilfe der Kovarianzmatrix. Alternativ finden wir die Unsicherheit (meist) näherungsweise[3] aus der Entwicklung um das Minimum

$$\chi^2(\theta) \approx \chi^2_{\text{min}} + \frac{1}{2}\sum_{n,m} \left.\frac{\partial^2 \chi^2(\theta)}{\partial\theta_n \partial\theta_m}\right|_{\theta=\hat{\theta}} (\theta_m - \hat{\theta}_m)(\theta_n - \hat{\theta}_n) = \chi^2_{\text{min}} + (\theta - \hat{\theta})H(\theta - \hat{\theta}),$$

wobei H die Hesse-Matrix mit den Elementen

$$H_{mn} = \frac{1}{2}\frac{\partial^2[\chi^2(\theta)]}{\partial\theta_m \partial\theta_n}$$

[3] wir geben in diesem Dokument alle Passagen, die auf numerischen Näherungen beruhen, in blau aus.

ist (beachte, dass wir wie oben den Faktor 1/2 aus der Taylor-Entwicklung in die Hesse-Matrix genommen haben). Das entspricht der gauss'schen Näherung

$$\mathrm{pdf}(\theta|\{(x_i, y_i)\}, \{\sigma_i\}) \propto \exp\left(-\frac{1}{2}(\theta - \hat{\theta})H(\theta - \hat{\theta})\right)$$

für die Posteriorverteilung.

Das Berechnen der zweiten Ableitung könnte man mit der Methode der finiten Differenzen näherungsweise durchführen. Numerisch genauer ist jedoch eine alternative Methode, welche mit der Jacobimatrix arbeitet. Am Minimum gilt nämlich für kleine δ

$$\chi^2(\hat{\theta} + \delta) = \sum_{i=1}^{N}\sum_{j=1}^{N}(y_i - f(x_i; \hat{\theta} + \delta))W_{ij}(y_j - f(x_j; \hat{\theta} + \delta))$$

Wir nähern nun linear

$$f(x_i; \hat{\theta} + \delta) \approx f(x_i; \hat{\theta}) + \sum_{n=1}^{M} \left.\frac{\partial f(x_i; \theta)}{\partial \theta_n}\right|_{\theta=\hat{\theta}} \delta_n,$$

wobei M die Zahl der Parameter ist. Die $N \times M$-Matrix mit den Elementen

$$\boxed{J_{in} = \left.\frac{\partial f(x_i, \theta)}{\partial \theta_n}\right|_{\theta=\hat{\theta}}}$$

nennt man Jacobi-Matrix am Punkt $\hat{\theta}$. In dieser Näherung haben wir

$$\begin{aligned}
\chi^2(\hat{\theta} + \delta) &\approx \sum_{i=1}^{N}\sum_{j=1}^{N}\left(y_i - f(x_i; \hat{\theta}) - \sum_{m=1}^{M} J_{im}\delta_m\right) W_{ij} \left(y_j - f(x_j; \hat{\theta}) - \sum_{n=1}^{M} J_{jn}\delta_n\right) \\
&\approx \chi_{\min} + \sum_{m=1}^{M}\sum_{n=1}^{M} \delta_m (J^T W J)_{mn}\delta_n
\end{aligned}$$

Aus der letzten Zeile der obigen Rechnung identifizieren wir

$$H_{mn} = \frac{1}{2} \left.\frac{\partial^2 \chi^2(\theta)}{\partial \theta_m \theta_n}\right|_{\theta=\hat{\theta}} = (J^T W J)_{mn},$$

so dass

$$H = J^T W J,$$

und die Kovarianzmatrix der Parameter

$$K = (J^T W J)^{-1}.$$

Python: Kovarianzmatrix. Diese Formel wird verwendet, wenn das Minimum von χ^2 numerisch mit Hilfe des Levenberg-Marquardt Algorithmus bestimmt wird, wobei die Jacobi-Matrix ohnehin eine wichtige Rolle spielt. Dies ist zum Beispiel bei der Python-Funktion `scipy.optimize.curve_fit` der Fall, welche die Kovarianzmatrix genau nach dieser Formel berechnet. Man erhält sie mit dem Code

```
covMatrix = result[1].
```

Bemerkung 34. *Die Diagonalelemente σ_{ii}^2 der Kovarianzmatrix sind die quadratischen Unsicherheiten der Parameter θ_i. Die Nebendiagonalelemente σ_{ij}^2 heissen Kovarianzen. Ist eine Kovarianz Null, dann sind die beiden entsprechenden Parameter unkorreliert. Man definiert den linearen Korrelationskoeffizienten zweier Parameter θ_i und θ_j als*

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

Der Wert dieses Koeffizienten liegt im Intervall $[-1, 1]$. Ein Wert von 1 zeigt perfekte Korrelation an, ein Wert von -1 perfekte Antikorrelation. Unkorrelierte Parameter haben $\rho = 0$.

Bemerkung 35. *Sind bei Modellen mit mehreren Parametern bestimmte Paare von Parametern korreliert, so muss bei der Angabe der Genauigkeit der Parameter auch der Wert der Kovarianz, oder alternativ der des Korrelationskoeffizienten angegeben werden.*