8. Vorlesung, 11. Nov. 2013

Letzte Woche:

- → Methode der kleinsten Fehlerquadrate
- → graphische Methode zur Bestimmung der Unsicherheiten bei der Parameterabschätzung
- → Aufspüren systematischer Fehler durch Untersuchung der Fit-Residuen

Modellvergleich

Bei der Datenanalyse treffen wir häufig auf folgendes Problem: wir haben in einem Experiment einen Datensatz D gemessen, die Theorie bietet uns aber zwei (oder mehrere) Modelle M_1, M_2 zur Analyse der Daten an. Jedes Modell hat seinen eigenen Satz von Parametern (θ_1, θ_2), wobei sich die Modelle auch in der Zahl ihrer Parameter unterscheiden können. Die Abschätzung der Modellparameter nimmt immer die Gültigkeit des entsprechenden Modells als Grundvoraussetzung an. Die Frage lautet nun:
Können wir quantitativ angeben, welches Modell die Daten besser beschreibt?

Beispiele:

- → ist es besser, die gemessenen Daten $y_j(x_j)$ mit $y_j = c$, $y_j = bx_j + c$, oder mit $y_j = ax_j^2 + bx_j + c$ zu beschreiben?
- → Ist eine Resonanz thermisch verbreitert (beschrieben durch Ableitung der Fermiverteilung), Lebensdauerverbreitert (Lorentz-funktion) oder inhomogen verbreitert (Gaussfunktion)?
- → ist der Wert der gemessenen Grösse Null, oder von Null verschieden?
- → ...

Es wäre wünschenswert, die Wahrscheinlichkeit für die Wahrheit des Modells M bei gegebenen Daten

$$\text{prob}(M \mid D, I)$$

angeben zu können. Das Problem ist, dass wir dafür einen vollständigen Satz sich gegenseitig ausschliessender Modelle benötigen würden, um diese Wahrscheinlichkeit normieren zu können. Dies ist in der Regel jedoch nicht der Fall. Formal heisst das (Bayessches Theorem)

$$\text{prob}(M_j \mid D, I) \overset{(5)}{=} \frac{\text{prob}(M_j \mid I)\,\text{prob}(D \mid M_j, I)}{\sum_j \text{prob}(M_j \mid I)\,\text{prob}(D \mid M_j, I)},$$

wobei wir den Normierungsnenner nicht kennen! Als Behelf macht es Sinn, Verhältnisse von Modellwahrscheinlichkeiten zu betrachten, denn

$$\boxed{\ln\underbrace{\left(\frac{\text{prob}(M_1 \mid D, I)}{\text{prob}(M_2 \mid D, I)}\right)}_{\text{"Bayes Faktor"}} = \ln\left(\frac{\text{prob}(M_1 \mid I)\,\text{prob}(D \mid M_1, I)}{\text{prob}(M_2 \mid I)\,\text{prob}(D \mid M_2, I)}\right) = \underbrace{\ln\left(\frac{\text{prob}(M_1 \mid I)}{\text{prob}(M_2 \mid I)}\right)}_{\text{"welches Modell bevorzugen wir, bevor wir die Daten D kennen?"}} + \underbrace{\ln\left(\frac{\text{prob}(D \mid M_1, I)}{\text{prob}(D \mid M_2, I)}\right)}_{\text{"welche Präferenz zu } M_1 \text{ und } M_2 \text{ spricht aus den Daten D?"}}} \quad (73)$$

Im Folgenden werden wir stets davon ausgehen, dass wir vor der Messung keines der beiden Modelle bevorzugen, so dass der erste Term identisch Null ist. Die Ausdrücke $\text{prob}(D \mid M_j, I)$ heissen Evidenz für Modell M_j. Sie ergeben sich aus

$$\boxed{\text{prob}(D \mid M_j, I) \overset{(4)}{=} \int \text{prob}(D, \theta_j \mid M_j, I)\, d\theta_j \overset{(3)}{=} \int d\theta_j\, \underbrace{\text{pdf}(\theta_j \mid M_j, I)}_{\text{Prior prob. für Parameter } \theta_j \text{ des Modells } M_j}\, \underbrace{\text{prob}(D_j \mid \theta_j, M_j, I)}_{\text{Likelihood}}} \quad (74)$$

Anwendung auf Modelle mit additivem unabhängigem normalverteiltem Rauschen

Diese Modelle haben Priorverteilungen $pdf(\theta_j, \sigma | M_j, I)$ und Likelihoodfunktionen der Form

$$\boxed{pdf(D|\theta_j, \sigma, M_j, I) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left[-\frac{N\,Q(\theta)}{2\sigma^2}\right]} \qquad (75)$$

Um die Evidenz für M_j auszurechnen, benötigen wir das Integral (74). Die Integration über σ lässt sich analytisch ausführen, wenn man $pdf(\sigma) = C_\sigma$ = const. annimmt. Das ergibt (bis auf irrelevante Faktoren)

$$\left[\frac{1}{Q_j(\theta)}\right]^{\frac{N-1}{2}} = \left(\frac{1}{Q_{min,j}}\right)^{\frac{N-1}{2}} \left(\frac{Q_{min,j}}{Q_j(\theta)}\right)^{\frac{N-1}{2}}$$

Die Integration über die Parameter θ führen wir numerisch aus. Dabei setzen wir für die Priorwahrscheinlichkeiten jeweils Konstanten C_{θ_j} an. Das numerisch berechnete Integral definieren wir als

$$\boxed{\Omega_j := \int \left(\frac{Q_{min,j}}{Q_j(\theta_j)}\right)^{\frac{N-1}{2}} d\theta_j} \qquad (76)$$

Diese Grösse ist ein Mass für das Volumen im Parameterraum, in dem es eine signifikante Wahrscheinlichkeit für die Parameter θ_j gibt. Dies bezeichnen wir als das relevante Parameterraumvolumen. ↗ Folie

Damit ist die Evidenz (74) gegeben durch

$$pdf(D|M_j, I) \propto \left(\frac{1}{Q_{min,j}}\right)^{\frac{N-1}{2}} C_{\theta_j}\,\Omega_j$$

Die Konstante C_{θ_j} stellt das inverse Volumen im Parameterraum dar, in dem die Priorwahrscheinlichkeit einen signifikanten Wert hat. Das Produkt $C_{\theta_j}\Omega_j$ beschreibt daher, um welchen Faktor sich das relevante Volumen des Parameterraumes durch die Messung der Daten D verkleinert hat.

Der Modellvergleich ergibt also die relevante Kennzahl

$$\boxed{\ln\frac{pdf(D|M_1, I)}{pdf(D|M_2, I)} = \frac{N-1}{2}\ln\left(\frac{Q_{min,2}}{Q_{min,1}}\right) + \ln\frac{C_{\theta_1}\Omega_1}{C_{\theta_2}\Omega_2}} \qquad (77)$$

Wir bevorzugen das Modell M_2 gegenüber M_1, wenn dieser Ausdruck kleiner als Null ist, sonst bevorzugen wir M_1 gegenüber M_2. Der erste Term auf der rechten Seite bevorzugt also das Modell mit dem kleineren $Q_{min,j}$. Der zweite Term bevorzugt dasjenige Modell, das den relevanten Parameterraum weniger verkleinert.

Beispiel Fallversuch: $N = 31$; $Q_{min,1} = 2.899 \times 10^{-7}\,s^2$; $Q_{min,2} = 0.460 \times 10^{-7}\,s^2$; $\Omega_1 = 0.0137\,\frac{m}{s^2}$; $\Omega_2 = 2.281 \times 10^{-6}\,\frac{m^2}{s^2}$

↗ Folien

erster Term

$$\frac{N-1}{2}\log\frac{Q_{min,2}}{Q_{min,1}} = -27.6$$

zweiter Term

$$\ln\frac{C_g\,\Omega_1}{C_g C_{z_0}\,\Omega_2} = \ln\frac{\Omega_1}{C_{z_0}\Omega_2}$$

Um eine Aussage machen zu können, sind wir gezwungen uns auf einen Wert für C_{z_0} festzulegen. Wir erinnern uns, dass z_0 der Offset in der Längenmessung ist. Nehmen wir zum Beispiel an, wir sind sicher, dass $-0.5m \le z_0 \le 0.5m$ ist, dann wäre

$C_{z_0} = 1m$ und $C_{z_0}\Omega_2 = 2.281 \times 10^{-6}\,\frac{m}{s^2} \Rightarrow \frac{\Omega_1}{\Omega_2 C_{z_0}} = 5990.75 \Rightarrow \ln\frac{\Omega_1}{\Omega_2 C_{z_0}} = +8.70$

Der erste Term favorisiert Modell 2 wegen des deutlich besseren Fits. Der zweite Term favorisiert Modell 1, weil das relevante Volumen im Parameterraum weniger stark abnimmt. Das liegt auch daran, dass Modell 2 einen zweidimensionalen Parameterraum besitzt. Der zweite Term 'bestraft' uns daher für den zusätzlichen Parameter z_0. Dennoch ist insgesamt

$$\ln\frac{prob(M_1|D,I)}{prob(M_2|D,I)} = -27.6 + 8.70 = -18.9 \Rightarrow prob(M_2|D,I) = 1.6 \times 10^8\, pdf(M_1|D,I)$$