

Parameterschätzung

Anne Voormann

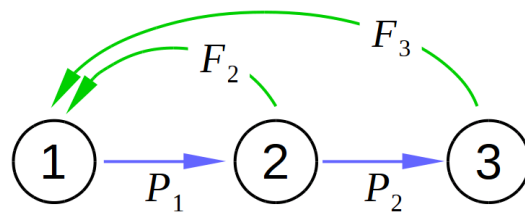
SUMMER SCHOOL KOGNITIVE MODELLIERUNG 2022



Inhalt

1. Daten und Modelle
2. Logik Parameterschätzung
3. Abweichungsfunktionen
4. Minimierung von Abweichungsfunktionen
5. Unsicherheit der Parameterschätzungen

Rückblick



Modell

Simulation

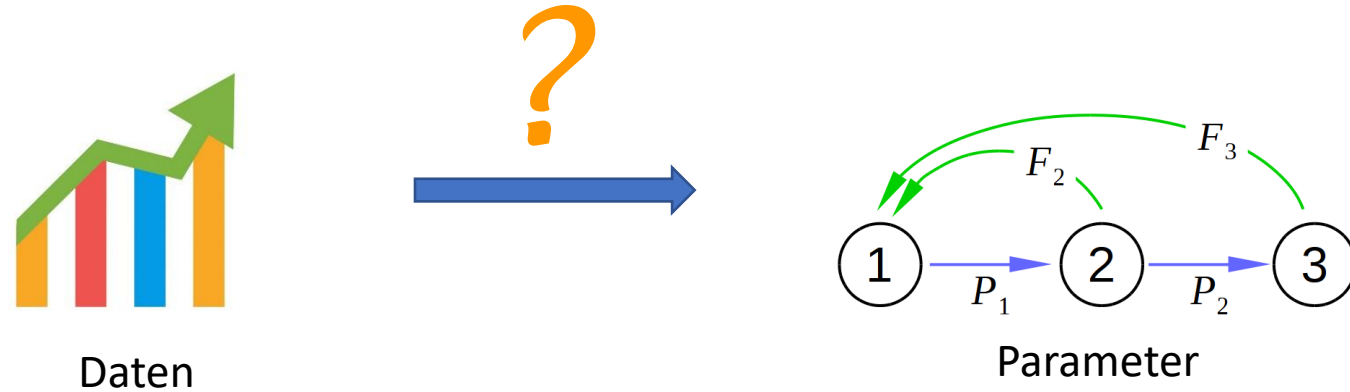


Daten

- Verstehen des Verhalten des Modells
- Ableitung qualitativer Vorhersagen
- Test von vorhandenen qualitativen Vorhersagen

Parameterschätzung

Parameterschätzung



- Überprüfen von Passung zwischen Modell und Daten
- Vergleich von Gruppen
- Testen von Manipulationen

Ablauf Parameterschätzung

- a) Definition eines Modells
- b) Auswahl einer Abweichungsfunktion /
Definition einer Plausibilitätsfunktion
- c) Minimierung der Abweichungsfunktion /
Maximierung der Plausibilitätsfunktion

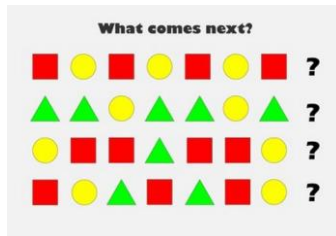
Modelldefinition (1)

- Überlegen des datengenerierenden Prozesses



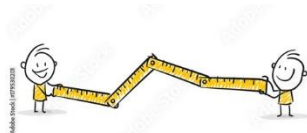
Wie oft muss man eine Münze werfen, damit 5-mal Zahl erscheint?

Pascal-Verteilung



Wie wahrscheinlich ist es, dass eine Person in einem Intelligenztest innerhalb von einer Minute 1, 2, 3, ..., n Aufgaben löst?

Poisson-Verteilung



Wie wahrscheinlich ist es, eine bestimmte Differenz zwischen zwei Gruppen zu beobachten, wenn zur Standardisierung die beobachtete Varianz verwendet wird?

t-Verteilung

Modelldefinition (2)

- Überlegen des datengenerierenden Prozesses
- Definition der mathematischen Zusammenhänge, die zu dem Outcome führen

Wie kommt es zur
Lösungswahrscheinlichkeit in
einem Gedächtnistest?

$$f(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Könnte sich aus einer Erinnerungswahrscheinlichkeit und Ratewahrscheinlichkeit zusammensetzen

Wie kommt es zur
beobachteten Ausprägung von
Motivation?

$$\hat{y}_i = b_0 + b_1 \times x_i$$

- Könnte sich aus einer Grundmotivation und der skalierten Freude an einer Tätigkeit zusammensetzen

Modelldefinition (2)

- Überlegen der datengenerierenden Funktion
- Definition der mathematischen Zusammenhänge, die zu dem Outcome führen

Wie kommt es zur Lösungswahrscheinlichkeit in einem Gedächtnistest?

Wie kommt es zu einer bestimmten Rate?

- Wenn Parameter geschätzt werden sollen:
- Beachten des Verhältnisses zwischen den vorhandenen Freiheitsgraden und den zu schätzenden Modellparametern.
- Könnte sich aus einer Grundmotivation und der skalierten Freude an einer Tätigkeit zusammensetzen



Übungsaufgabe

Erstellt eine Funktion, die euch Vorhersagen für eine Regression mit einem Prädiktor und einem Intercept erstellt.

Ablauf Parameterschätzung

- a) Definition eines Modells
- b) Auswahl einer Abweichungsfunktion /
Definition einer Plausibilitätsfunktion
- c) Minimierung der Abweichungsfunktion /
Maximierung der Plausibilitätsfunktion

Abweichungsfunktionen

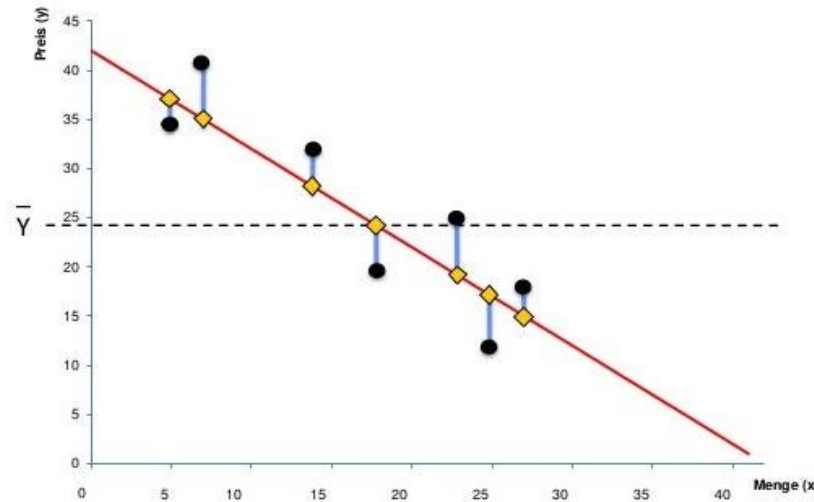
- = Beschreibung der Differenz zwischen beobachteten Daten und erwarteten Daten (aufgrund der Modellvorhersage)
- Nutzung der Höhe der Differenz um eine Aussage über die Passung der Parameter zu treffen
- Arten von Abweichungsfunktionen
 - Für kontinuierliche Daten
 - Für diskrete Daten
 - Deviance

Kontinuierliche Daten

- Methode der kleinsten Quadrate (root-mean-squared error; RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^J (\hat{y}_j - y_j)^2}{J}}$$

\hat{y} = vorhergesagte Daten
 y = beobachtete Daten
 J = Anzahl der
Beobachtungen / Zellen



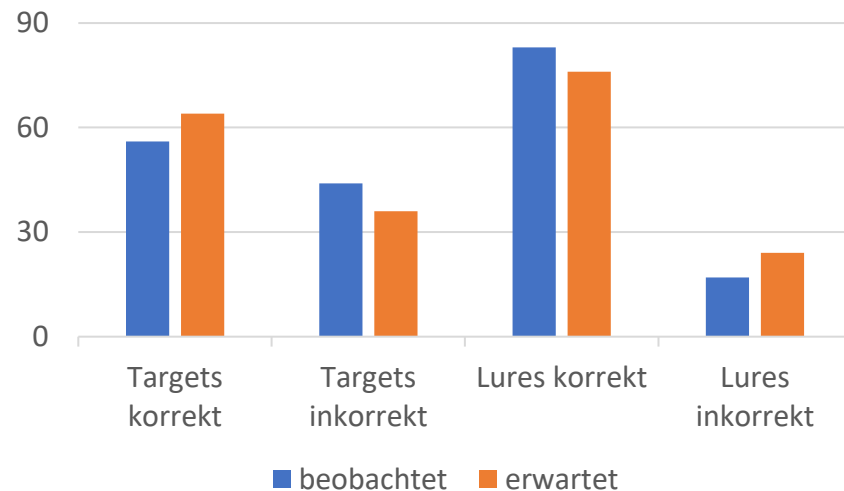
Parameterschätzung

Diskrete Daten

- G^2, X^2

$$X^2 = \sum_{j=1}^J \frac{(O_j - Np_j)^2}{Np_j}$$

$$G^2 = 2 \sum_{j=1}^J O_j \log\left(\frac{O_j}{Np_j}\right)$$



O_j = Anzahl beobachteter Reaktionen in Kategorie j
 N = Summe aller beobachteten Reaktionen
 p_j = Modell Vorhersage (als Wahrscheinlichkeit)
 J = Anzahl der Kategorien / Zellen



Übungsaufgabe

Erstellt eine Funktion zur Berechnung des RMSE

PAUSE



Deviance

- Sowohl für kontinuierliche als auch diskrete Daten
- Basiert auf der **Likelihood**

$$D(y, \hat{y}) = -2\ln(L(\theta|y, M))$$

y = Daten
 θ = Modellparameter
 M = Modell
 L = Likelihood

Likelihood (Plausibilitätsfunktion)

= Wahrscheinlichkeit der Daten unter einem bestimmten Parameterset θ eines Modells M
 $p(y | \theta, M)$

Woran erinnert euch das?

- Probability density function (PDF) – *kontinuierliche Verteilungen*
- Probability mass function - *diskrete Verteilungen*

Achtung: Die Likelihood ist weder eine Wahrscheinlichkeit noch eine PDF!

Exkurs: Warum Likelihood?

- Ziel Parameterschätzung:

Finden der bestmöglichen Parameter für das entsprechende Modell – $p(\theta | d, M)$

- Problem:

Raum aller möglichen Parameter ist unbekannt

- Beste Näherung:

Betrachten der Wahrscheinlichkeit der Daten unter der Annahme eines bestimmten Modells und vielen verschiedenen Parameterwerten – $p(d|\theta, M)$ oder $L(\theta | d, M)$

Likelihood praktisch

Daten: [2, 3, 4, 5]

Modell: Normalverteilung mit Varianz 1

Frage: Welcher Mittelwertparameter ist für den 1., 2., 3. bzw. 4. Datenpunkt am plausibelsten?

➤ Praktisch austesten in R unter Verwendung der `dnorm()` Funktion

Mögliche Mittelwerte [1, 2, 3, 4, 5, 6]

Likelihood praktisch

Daten: [2, 3, 4, 5]

Modell: Normalverteilung mit Varianz 1

Frage: Welcher Mittelwert ist für den 1., 2., 3. bzw. 4. Datenpunkt am wahrscheinlichsten?

➤ Praktisch austesten in R unter Verwendung der `dnorm()` Funktion

Mögliche Mittelwerte [1, 2, 3, 4, 5, 6]

```
> d <- c(2, 3, 4, 5)
> for (i in 1:6) {
+   print(i)
+   print(dnorm(d, i))
+ }
[1] 1
[1] 0.2419707245 0.0539909665 0.0044318484 0.0001338302
[1] 2
[1] 0.398942280 0.241970725 0.053990967 0.004431848
[1] 3
[1] 0.24197072 0.39894228 0.24197072 0.05399097
[1] 4
[1] 0.05399097 0.24197072 0.39894228 0.24197072
[1] 5
[1] 0.004431848 0.053990967 0.241970725 0.398942280
[1] 6
[1] 0.0001338302 0.0044318484 0.0539909665 0.2419707245
```

Likelihood (2)

- Berechnung der Likelihood bei mehreren Datenpunkten

$$L(\theta|d, M) = \prod_{j=1}^J L(\theta|d_j, M)$$
$$L(\theta|d, M) = \prod_{j=1}^J p(d_j|\theta, M)$$

➤ Welcher Mittelwert würde nun am besten passen?

```
> d <- c(2, 3, 4, 5)
> for (i in 1:6) {
+   print(i)
+   print(prod(dnorm(d, i)))
+ }
[1] 1
[1] 7.748596e-09
[1] 2
[1] 2.309824e-05
[1] 3
[1] 0.001261121
[1] 4
[1] 0.001261121
[1] 5
[1] 2.309824e-05
[1] 6
[1] 7.748596e-09
```

Dichtefunktion vs. Likelihood

Dichtefunktion

Ereignisse haben einen Wert zwischen 0 und ∞

Summe/Integral der Ereignisse ist 1

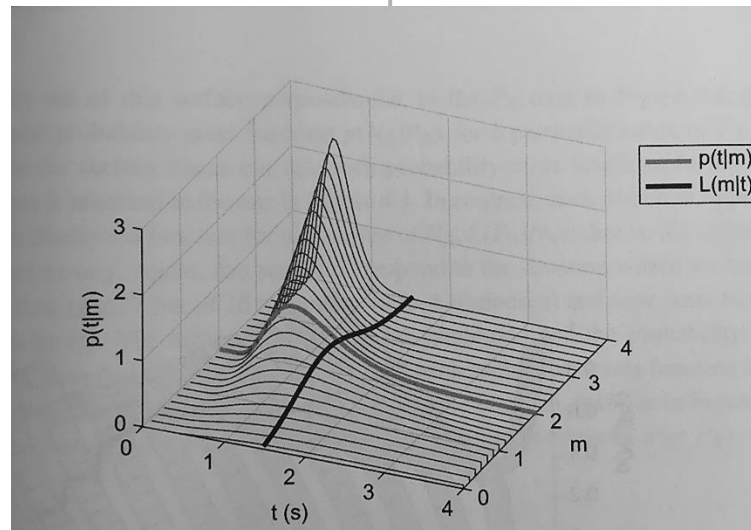
Betrachtung variabler Daten unter konstanten
Parametern und Modell

Likelihood

Ereignisse haben einen Wert zwischen 0 und ∞

Summe/Integral der Ereignisse variabel

Betrachtung konstanter Daten unter variablen
Parameter bei konstantem Modell



Parameterschätzung

Deviance (2)

- Basiert auf der Likelihood

$$D(y, \hat{y}) = -2\ln(L(\theta|y, M))$$

y = Daten
 θ = Modellparameter
 M = Modell
 L = Likelihood

Transformationen:

- $\ln()$:
 - Erleichtert multiplikative Operationen ($\ln(a*b) = \ln(a) + \ln(b)$)
 - Bessere Differenzierbarkeit in kleinen Wertebereichen
- -2 :
 - Erlaubt Minimierung der Funktion anstelle von Maximierung



Übungsaufgabe

Überlegt euch, wie ihr die Deviance für euer Regressionsmodell erstellen könntet und programmiert dies in eine Funktion

PAUSE



Ablauf Parameterschätzung

- a) Definition eines Modells
- b) Auswahl einer Abweichungsfunktion /
Definition einer Plausibilitätsfunktion
- c) Minimierung der Abweichungsfunktion /
Maximierung der Plausibilitätsfunktion

Minimierung der Abweichungsfunktion

- Algebraisch
- Iterativ
 - = Optimierung der Parameter in mehreren Schleifen anhand eines spezifischen Optimierungsalgorithmus

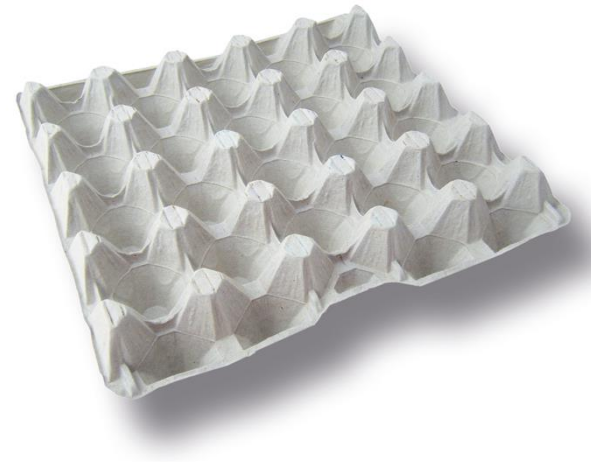
Nelder and Mead's Simplex

- Geometrische Figur aus einer arbiträren Anzahl an Punkten in einer Dimension korrespondierend zum Parameterraum
- **Optimierung** der Parameter durch:
 - Spiegeln des Objektes
 - Verkleinerung des Objektes durch Verschiebung des Punktes mit dem schlechtesten Fit
- **Nachteile:**
 - Nur für kontinuierliche Parameter geeignet
 - Nur für bis zu 5 Parameter geeignet
 - Nur bei deterministischem Zusammenhang zwischen Abweichungsfunktion und Parametern



Simulated Annealing

- Entgegenwirken der Parameterauswahl von lokalen Minima
- Bei Parametersets mit geringerer Abweichung:
 - Annehmen der neuen Parameterkombination
- Bei Parametersets mit höherer Abweichung:
 - Annehmen der neuen Parameterkombination mit einer gewissen Wahrscheinlichkeit abhängig von
 - a) Der Höhe der Verschlechterung
 - b) Der „Temperatur“



Minimierung der Abweichungsfunktion in R

- Iterativ

```
optim(par, fn, gr = NULL, ...,  
      method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN",  
                  "Brent"),  
      lower = -Inf, upper = Inf,  
      control = list(), hessian = FALSE)
```

- Ein paar Hinweise:
 - Optim nimmt nur einen Parametervektor (par) für die zu schätzenden Parameter
 - In der Funktion, die minimiert wird (fn), müssen die Parameter als erstes Argument stehen
 - [...] erlaubt die Zugabe von weiteren Argumenten, die die Funktion (fn) braucht, z.B. Daten



Übungsaufgabe

Minimiert den RMSE und die Deviance und vergleicht die
Ergebnisse

Zeit zum Üben und Ausprobieren



Danke für Eure Aufmerksamkeit!

