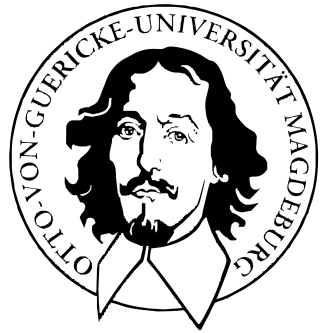


Grundlagen der Testtheorie

BSc Psychologie WiSe 2025/26

Prof. Dr. Dirk Ostwald

(4) Item-Response-Theorie

Grundlagen

Birnbaum-Modell

Rasch-Modell

Selbstkontrollfragen

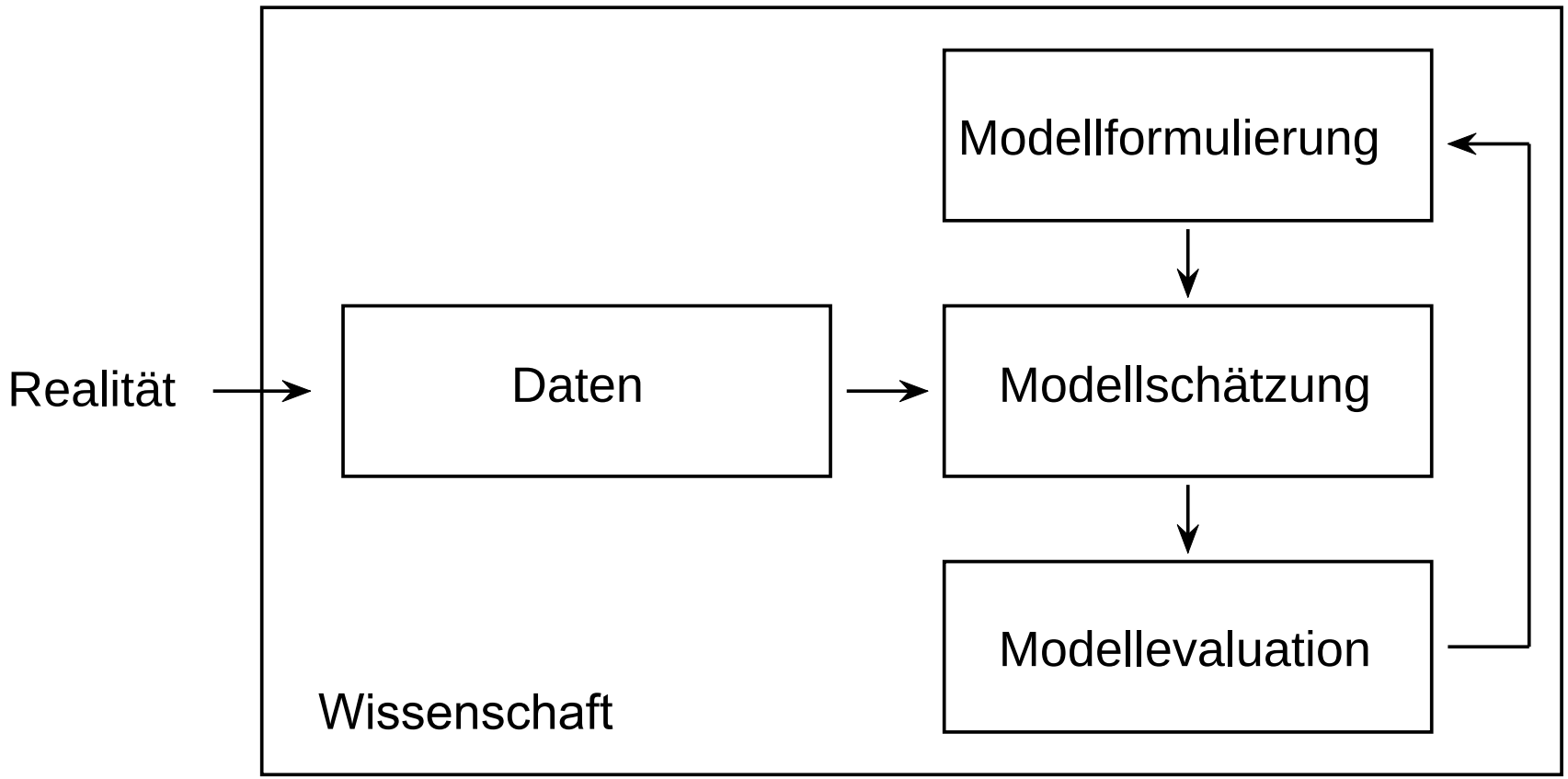
Grundlagen

Birnbaum-Modell

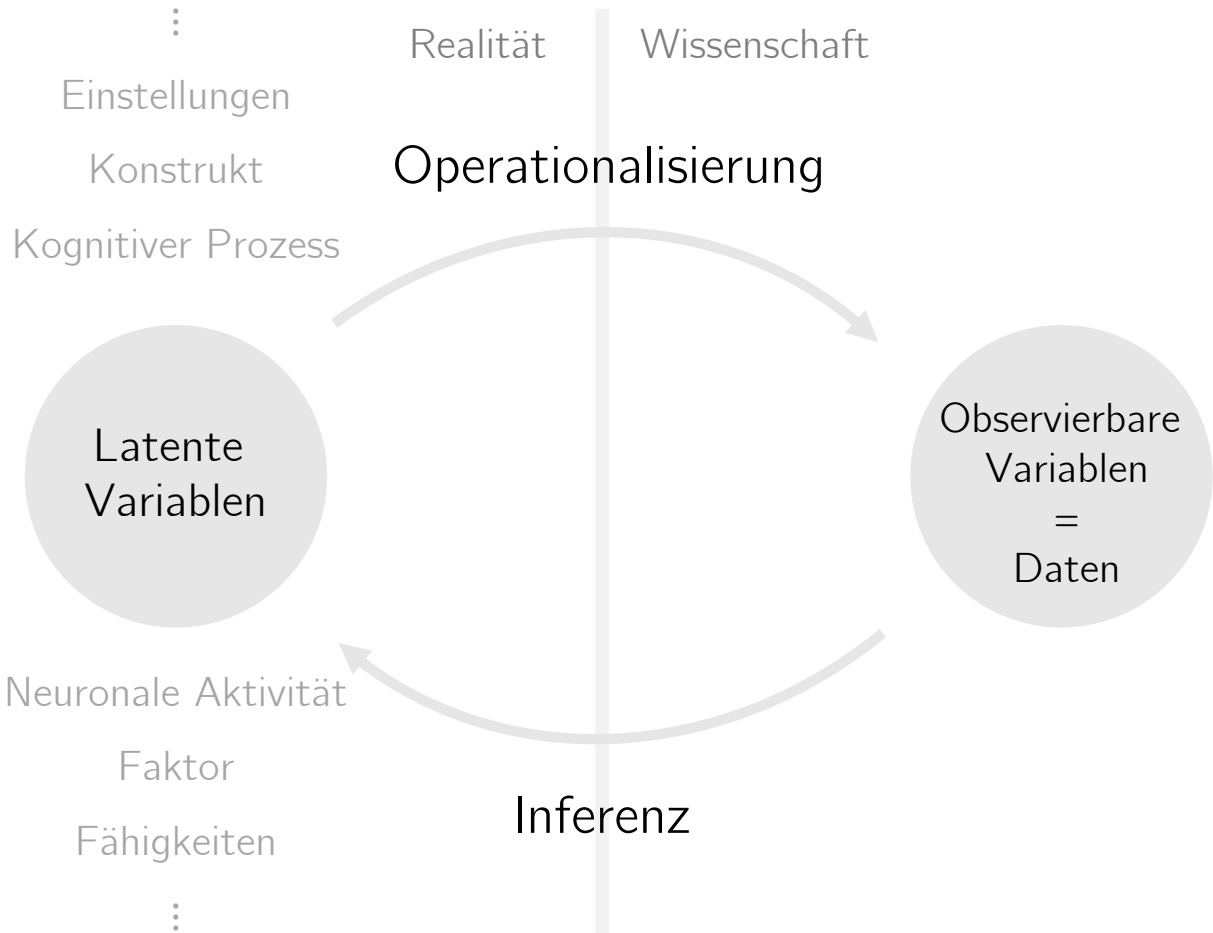
Rasch-Modell

Selbstkontrollfragen

Psychologische Datenwissenschaft



Psychologische Datenwissenschaft



Item-Response-Theorien

Probabilistische Modelle für die funktionalen Beziehungen zwischen

- latenten Personeneigenschaften,
- Itemparametern und
- Itemantwortwahrscheinlichkeiten

Typische latente Personeneigenschaften sind dabei

- Fähigkeiten in der Leistungsmessung
- Einstellungen in der Persönlichkeitsmessung

Typische Itemparameter sind dabei

- Schwierigkeit, Diskriminationsfähigkeit, Ratewahrscheinlichkeit

Item-Response-Theorien bilden eine große Modellfamilie

- Von eindimensionalen Personeneigenschaften zu mehrdimensionalen Personeneigenschaften
- Von dichotomen Antwortformaten zu polytomen Antwortformaten
- Wir betrachten hier nur eindimensionalen Personeneigenschaften und dichotome Antwortformate

Weite Verbreitung in Bildungsforschung (PISA, GRE), Medizin (PROMIS), Psychometrie und Sozialwissenschaften

Theoretische Grundlage für adaptive Testverfahren mit interaktiver Itemauswahl

Eine grundlegende Einführung gibt das dreibändige *Handbook of Item Response Theory* von Linden (2016)

Traditionen

Frühe Ansätze von Item-Response-Theorien unter anderem bei Binet & Simon (1904)

Statistische Tradition

- Datenanalytisch-geprägter USA Ansatz nach Mosier (1940), Richardson (1936), Lawley (1943), Lord (1951)
- Seminale Arbeiten von Allan Birnbaum (1923 - 1976) in Lord & Novick (1968) (vgl. auch Birnbaum (1962))
- 1PL, 2PL, 3PL Modelle als Weiterentwicklung des Normal-Ogiven-Modells
- Marginal-Maximum-Likelihood / Expectation-Maximization
- Frühe Implementationen in Programmen wie LOGIST, BILOG-MG, TESTFACT, MULTILOG
- Recht klar im datenanalytischen Mainstream verortet

Messtheoretische Tradition

- Axiomatisch-geprägter europäischer Ansatz nach Georg Rasch (1901–1980) (Rasch (1960), Rasch (1966))
- Seminale Arbeiten von Erling Bernhard Andersen (1939-2004) in Andersen (1970) und Andersen (1977)
- Popularisierung in den USA unter anderem durch Wright (1977) und Andrich (1978)
- Europäische Weiterführung durch Fischer (1938 -) und Molenaar (1935-2018) (vgl. Fischer (1995))
- Conditional-Maximum-Likelihood / Suffizienz
- Frühe Implementation in Programmen wie WinSteps, ConQuest, RUMM
- Nicht im datenanalytischen Mainstream verortet, aber bei psychologischen Anwendern populär

vgl. Embretson & Reise (2000)

Anwendungsbeispiel

Simulierter Datensatz basierend auf den 18 dichotomen Items der GEB Scale (Kaiser (1998))

Bei den folgenden Handlungen ist nicht die Häufigkeit gefragt; es geht vielmehr darum, was *eher für Sie zutrifft*. Kreuzen Sie *„Keine Angabe“* (KA) dann an, wenn eine Frage auf Ihre momentane Lebenssituation *nicht zutrifft* (beispielsweise können Sie über keine Geschirrspülmaschine einer bestimmten Energieeffizienzklasse verfügen, wenn Sie keine Geschirrspülmaschine besitzen).

		ja	nein	KA
1	Ich verwende Einkaufstüten oder -taschen mehrfach.	☐	☐	☐
2	In meiner Wohnung ist es im Winter so warm, dass man ohne Pullover nicht friert.	☐	☐	☐
3	Ich benutze beim Waschen einen Weichspüler.	☐	☐	☐
4	Leere Batterien werfe ich in den Hausmüll.	☐	☐	☐
5	Breiige Essensreste leere ich in die Toilette.	☐	☐	☐
6	In der Toilette benutze ich chemische Duftsteine für den guten Geruch.	☐	☐	☐
7	Ich bin Mitglied in einer Umweltschutzorganisation.	☐	☐	☐
8	Im Hotel lasse ich täglich die Handtücher wechseln.	☐	☐	☐
9	Ich besitze eine Geschirrspülmaschine der Effizienzklasse A+ oder besser.	☐	☐	☐
10	Ich verlasse nach einem Picknick den Platz genauso, wie ich ihn angetroffen habe.	☐	☐	☐
11	Ich habe eine Solaranlage zur Energie- bzw. Wärmeerzeugung angeschafft.	☐	☐	☐
12	Ich habe mich über Vor- und Nachteile einer Solaranlage informiert.	☐	☐	☐
13	Ich beziehe Strom aus erneuerbarer Energie.	☐	☐	☐
14	Ich verzichte auf ein Auto.	☐	☐	☐
15	Ich bin in einem „Car-Sharing“-Pool.	☐	☐	☐
16	Durch mein Fahrverhalten versuche ich, den Kraftstoffverbrauch so niedrig wie möglich zu halten.	☐	☐	☐
17	Ich besitze ein verbrauchsreduziertes Auto (weniger als 6 Liter Treibstoff pro 100 km).	☐	☐	☐
18	Ich ernähre mich vegetarisch.	☐	☐	☐

Anwendungsbeispiel

Simulierter Datensatz basierend auf den 18 dichotomen Items der GEB Scale (Kaiser (1998))

	I 1	I 2	I 3	I 4	I 5	I 6	I 7	I 8	I 9	I 10	I 11	I 12	I 13	I 14	I 15	I 16	I 17	I 18
P 1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 2	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 3	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P 4	0	1	0	1	0	1	1	1	0	0	1	0	1	0	1	0	1	0
P 5	1	1	0	1	1	1	1	1	0	1	1	0	1	1	0	0	1	1
P 6	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0
P 7	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0
P 8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P 9	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
P 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P 11	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1
P 12	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 13	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0
P 14	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 15	1	1	0	0	1	0	0	1	1	0	0	0	1	0	1	0	1	0
P 16	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0
P 17	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0
P 18	0	1	0	1	1	1	0	1	1	1	1	0	0	1	1	0	1	0
P 19	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1
P 20	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1

$n = 20$ von $n = 1000$ Personen, keine KA Antworten

Überblick

Birnbaum-Modell

- Logistische Funktion
- Modellformulierung
- Modelleigenschaften
- Maximum-Marginal-Likelihood-Schätzung durch ELBO Maximierung
- Expectation-Maximization-Algorithmus nach Woodruff (1996)

Rasch-Modell

- Objektive Messungen und Spezifische Objektivität
- Modellformulierung
- Modelleigenschaften
- Conditional-Marginal-Likelihood-Schätzung
- Suffizienz und Parameterseparierbarkeit

Grundlagen

Birnbaum-Modell

Rasch-Modell

Selbstkontrollfragen

Definition (Logistische Funktion)

Die Funktion

$$f : \mathbb{R} \to]0, 1[, x \mapsto f(x) := \frac{1}{1 + \exp(-a(x - b))} \text{ mit } a, b \in \mathbb{R} \tag{1}$$

heißt *logistische Funktion*.

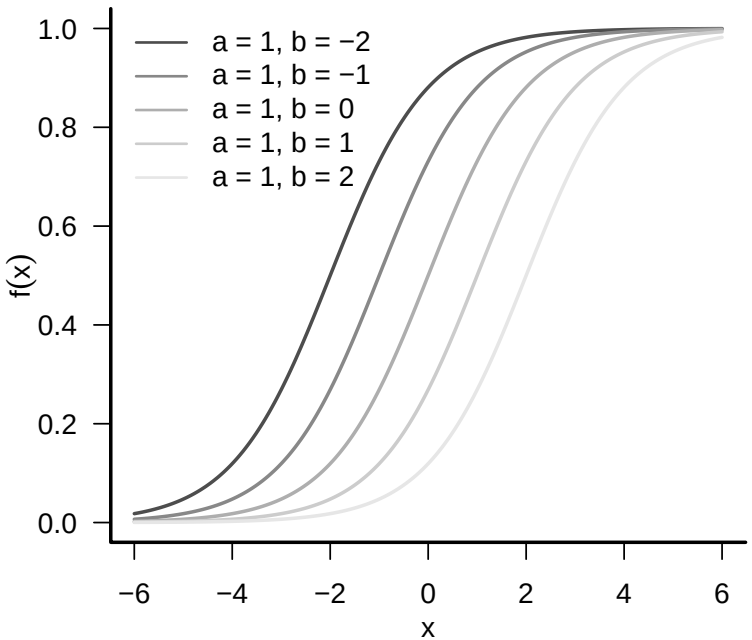
Bemerkungen

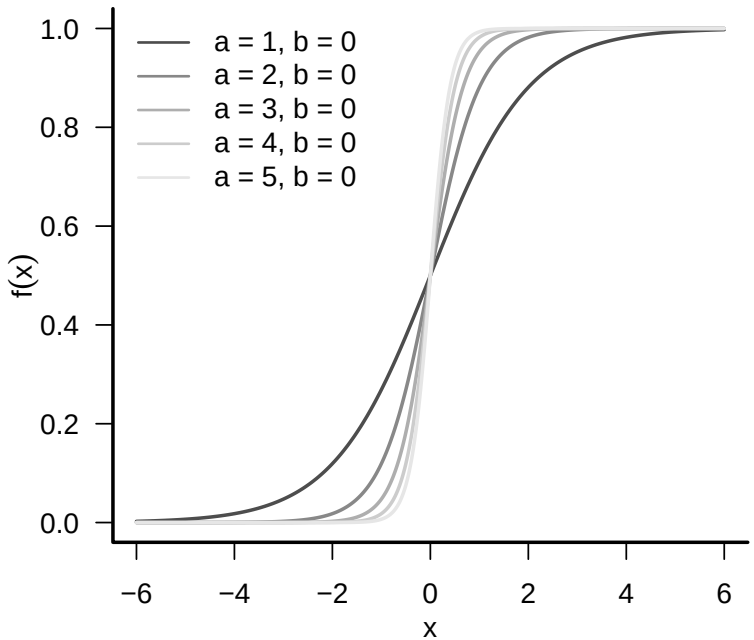
- Die logistische Funktion geht auf Verhulst (1845) zurück und findet breite mathematische Anwendung.
- Sie wird zur Modellierung von Wachstumsprozessen, Kondensatorladung, Pharmakokinetik u.v.a.m genutzt.
- Verwendung findet sie auch in der Logistischen Regression, dem Deep Learning und im Reinforcement Learning.
- Wir betrachten im Folgenden immer den Fall einer ansteigenden logistischen Funktion mit $a > 0$.
- Die **R** Funktion `plogis()` implementiert die *logistische Standardfunktion*

$$h : \mathbb{R} \to]0, 1[, x \mapsto h(x) := \frac{1}{1 + \exp(-x)} \tag{2}$$

- Formal ist `plogis` dabei die KVF der *logistischen Verteilung*.
- Es gilt also insbesondere $h(a(x - b)) = f(x)$.
- Die logistische Funktion ist die zentrale Komponente von Birnbaum- und Rasch-Modell.

Logistische Funktion





$\Rightarrow$ b verschiebt die logistische Funktion bezüglich der x-Achse, a moduliert ihre Steigung

Theorem (Eigenschaften der logistischen Funktion)

Für $a > 0, b \in \mathbb{R}$ sei

$$f : \mathbb{R} \to]0, 1[, x \mapsto f(x) := \frac{1}{1 + \exp(-a(x - b))} \tag{3}$$

die *logistische Funktion*. Dann gelten

(1) $f(x) \to 0$ für $x \to -\infty$ und $f(x) \to 1$ für $x \to \infty$ (Grenzwerte)

(2) $f(x) = \frac{\exp(a(x-b))}{1+\exp(a(x-b))}$ (Alternative Form)

(3) $f(b) = \frac{1}{2}$ (Spezieller Wert)

(4) $f'(x) = a(1 - f(x))f(x)$ (Ableitung)

(5) $f'(x) > 0$ für alle $x \in \mathbb{R}$ (Strenge Monotonie)

(6) $f'(b) = \frac{1}{4}a$ (Spezieller Ableitungswert)

(7) $\arg\max f'(x) = b$ (Maximalstelle der Ableitung)

Beweis

(1) Mit den Eigenschaften der Exponentialfunktion gilt

$$x \to -\infty \Rightarrow \exp(-a(x-b)) \to \infty \Rightarrow \frac{1}{1+\exp(-a(x-b))} \to \frac{1}{1+\infty} = 0 \tag{4}$$

sowie

$$x \to \infty \Rightarrow \exp(-a(x-b)) \to 0 \Rightarrow \frac{1}{1+\exp(-a(x-b))} \to \frac{1}{1} = 1 \tag{5}$$

(2) Mit den Eigenschaften der Exponentialfunktion gilt

$$\frac{1}{1+\exp(-a(x-b))} = \frac{1}{1+\frac{1}{\exp(a(x-b))}} = \frac{\exp(a(x-b))}{\exp(a(x-b)) + \frac{\exp(a(x-b))}{\exp(a(x-b))}} = \frac{\exp(a(x-b))}{1+\exp(a(x-b))} \tag{6}$$

(3) Mit den Eigenschaften der Exponentialfunktion gilt

$$f(b) = \frac{1}{1+\exp(-a(b-b))} = \frac{1}{1+\exp(0)} = \frac{1}{1+1} = \frac{1}{2} \tag{7}$$

Beweis (fortgeführt)

(4) Es gilt

$$
\begin{aligned}
f'(x) &= \frac{d}{dx}\left(1+\exp(-a(x-b))\right)^{-1} \\
&= -\left(1+\exp(-a(x-b))\right)^{-2}\frac{d}{dx}\left(1+\exp(-a(x-b))\right) \\
&= -\left(1+\exp(-a(x-b))\right)^{-2}\exp(-a(x-b))(-a) \\
&= a\left(\frac{\exp(-a(x-b))}{\left(1+\exp(-a(x-b))\right)^2}\right) \\
&= a\left(\frac{1+\exp(-a(x-b))-1}{\left(1+\exp(-a(x-b))\right)^2}\right) \\
&= a\left(\frac{1+\exp(-a(x-b))}{\left(1+\exp(-a(x-b))\right)^2}-\frac{1}{\left(1+\exp(-a(x-b))\right)^2}\right) \\
&= a\left(\frac{1}{1+\exp(-a(x-b))}-\frac{1}{\left(1+\exp(-a(x-b))\right)^2}\right) \\
&= a\left(\frac{1}{1+\exp(-a(x-b))}\left(1-\frac{1}{1+\exp(-a(x-b))}\right)\right) \\
&= af(x)(1-f(x))
\end{aligned}
\tag{8}
$$

Beweis (fortgeführt)

(5) Mit dem Beweis der für die Ableitung von f gilt

$$f'(x) = a \left(\frac{\exp(-a(x-b))}{(1+\exp(-a(x-b)))^2} \right) \tag{9}$$

Dabei ist $a > 0$ nach Voraussetzung, $\exp(-a(x-b)) > 0$ mit den Eigenschaften der Exponentialfunktion und $(1+\exp(-a(x-b)))^2 > 0$ mit den Eigenschaften der Exponential- und Quadratfunktion. Also gilt $f'(x) > 0$ für alle $x \in \mathbb{R}$ und f ist damit streng monoton steigend.

(6) Mit Eigenschaft (4) gilt

$$f'(b) = a(1-f(b))f(b) = a\left(1-\frac{1}{2}\right)\frac{1}{2} = \frac{1}{4}a \tag{10}$$

Beweis (fortgeführt)

(7) Da a konstant ist, nimmt die Ableitung der logistischen Funktion ihr Maximum genau dann an, wenn $f(x)(1-f(x))$ ein Maximum annimmt. Wir setzen also $y := f(x)$ und betrachten die Funktion

$$g : \mathbb{R} \to \mathbb{R}, y \mapsto g(y) := (1-y)y = y - y^2. \tag{11}$$

Dann gilt

$$g'(y) = 1 - 2y \text{ und } g''(y) = -2. \tag{12}$$

Nullsetzen der ersten Ableitung ergibt

$$g'(y^*) = 0 \Leftrightarrow 1 - 2y^* = 0 \Leftrightarrow y^* = \frac{1}{2} \tag{13}$$

und mit $g''(y) < 0$ handelt es sich dabei um ein Maximum. Die Ableitung der logistischen Funktion nimmt also für den Wert $f(x) = \frac{1}{2}$ ein Maximum an. Offenbar gilt mit (3), dass dann $x = b$ gilt. Da f nach Eigenschaft (5) streng monoton steigend ist, nimmt f den Wert $\frac{1}{2}$ nur einmal an.

Definition (Birnbaum-Modell)

Für $i = 1, ..., n$ Personen und $j = 1, ..., m$ dichotome Items seien

- x_i die Zufallsvariable zur Modellierung der Eigenschaft von Person i,
- y_{ij} die Zufallsvariable zur Modellierung der $j$ten Itemantwort von Person i mit Wertebereich $\{0, 1\}$,
- $y_i := (y_{i1}, ..., y_{im})^T$ der Zufallsvektor der Itemantworten von Person i,
- $a_j \in \mathbb{R}_{>0}, b_j \in \mathbb{R}, c_j \in [0, 1]$ die *Itemdiskriminations-, Itemschwierigkeits-, bzw. Itemrateparameter* von Item j
- $\kappa_j := \left(a_j, b_j, c_j\right)$ der Itemparametervektor von Item j,
- $\kappa := (\kappa_1, ..., \kappa_m)$ die Gesamtheit der Itemparameter.

Dann heißt die bedingte Item-Response-Verteilung

$$p_\kappa(y_i|x_i) := \prod_{j=1}^{m} p_{\kappa_j}(y_{ij}|x_i) \tag{14}$$

mit

$$p_{\kappa_j}(y_{ij}|x_i) := \text{Bern}\left(\mu_{ij}^{\kappa_j}\right) \text{ und } \mu_{ij}^{\kappa_j} := c_j + (1 - c_j)\frac{1}{1 + \exp\left(-a_j(x_i - b_j)\right)} \tag{15}$$

Birnbaum-Modell oder *3-Parameter-Logistisches-Modell (3PL-Modell)*. Weiterhin wird das Modell für

- $c_j := 0$ *2PL-Modell* und für
- $c_j := 0$ und $a_j := 1$ *1PL-Modell* oder *Rasch-Modell*.

genannt.

Birnbaum-Modell

Bemerkungen

- Das Modell wird ausführlich in Kapiteln 17 - 20 von Lord & Novick (1968) diskutiert.
- Die Modellierung von Personeneigenschaften x_i als Zufallsvariablen unterstellt MML-Schätzung.
- $p_{\kappa_j}(y_{ij}|x_i)$ wird die *Item-Response-Funktion* oder *Item-Characteristic-Curve (ICC)* des Modells genannt.
- $\prod_{j=1}^{m} p_{\kappa_j}(y_{ij}|x_i)$ enkodiert die bedingte Unabhängigkeit von Itemantworten gegeben die Eigenschaft x_i.
- Diese Faktorisierungseigenschaft des Modells über Items wird klassisch auch "lokale Unkorreliertheit" genannt.
- Für den Parameterraum der Parameter eines Items gilt $(a_j, b_j, c_j) \in]0, \infty[\times]-\infty, \infty[\times[0, 1]$.
- In der Literatur findet man oft die Bezeichnung *ability* und das Symbol θ_i für x_i.
- Die Bedeutung der Itemparameter verdeutlichen wir untenstehend.
- Lord (1951) entwickelte ursprünglich das sogenannte *Normal-Ogiven-Modell* mit

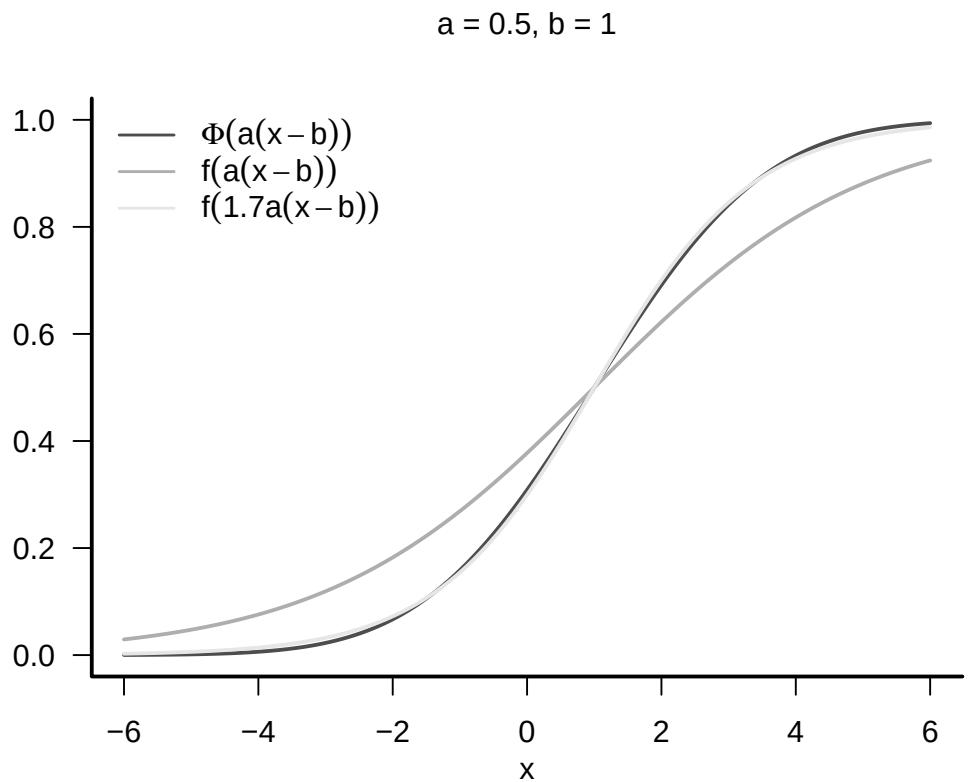
$$\mu_{ij} := \int_{-\infty}^{a_j(x_i - b_j)} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}s^2\right) ds = \Phi(a_j(x_i - b_j)) \tag{16}$$

  bei dem also die kumulative Dichtefunktion der Standardnormalverteilung als Item-Response-Funktion diente.
- Die logistische Funktion approximiert die KDF der Standardnormalverteilung und mathematisch einfacher.
- Birnbaum adaptierte in Lord & Novick (1968) frühere Arbeiten durch z.B. Finney (1952) im Bereich Bioassays.

Birnbaum-Modell

Bemerkungen (Normal-Ogiven-Modell und 2PL-Modell)

- Haley (1952) hat gezeigt, dass $|f(1.7x) - \Phi(x)| < 0.01$ für alle $x \in \mathbb{R}$.



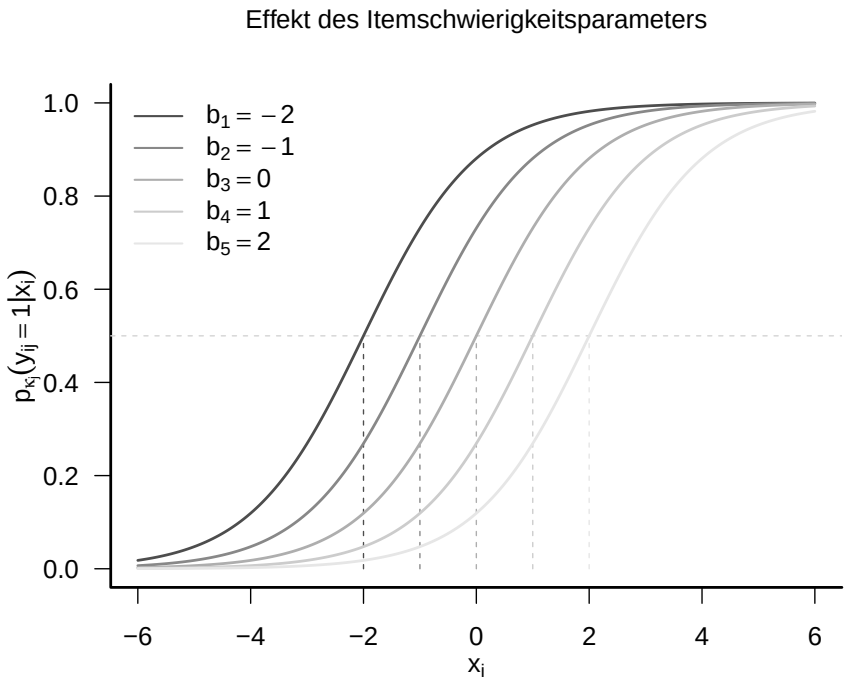
Birnbaum-Modell

Bemerkungen (Itemschwierigkeitsparameter)

- Es seien $c_j = 0$ und $x_i = b_j$. Dann gilt

$$p_{\kappa_j}(y_{ij}|x_i) = 0 + (1-0)\frac{1}{1+\exp(-a_j(b_j - b_j))} = \frac{1}{1+\exp(0)} = \frac{1}{1+1} = \frac{1}{2} \tag{17}$$

- Der Itemschwierigkeitsparameter b_j entspricht hier dem Eigenschaftswert x_i, für den $p_{\kappa_j}(y_{ij} = 1|x_i) = \frac{1}{2}$ ist.
- Höhere Schwierigkeitsparameterb_j implizieren hier höhere Eigenschaften x_i, für die $p_{\kappa_j}(y_{ij} = 1|x_i) = \frac{1}{2}$ ist.



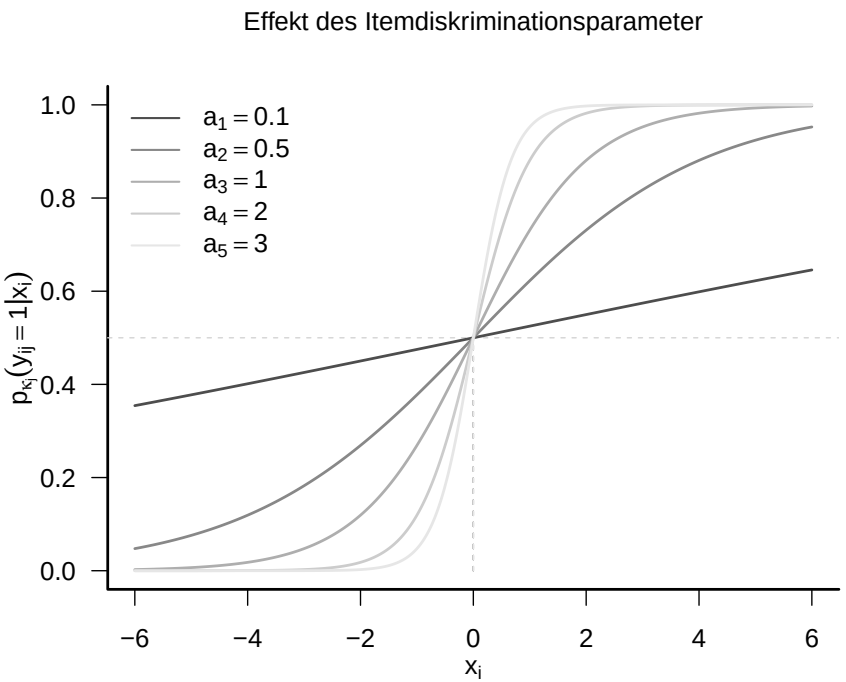
Birnbaum-Modell

Bemerkungen (Itemdiskriminationsparameter)

- Es seien $c_j = 0$. Dann gilt mit Aussage (6) von Theorem Eigenschaften der logistischen Funktion

$$p'_\kappa(y_{ij}|b_j) = \frac{1}{4} a_j. \tag{18}$$

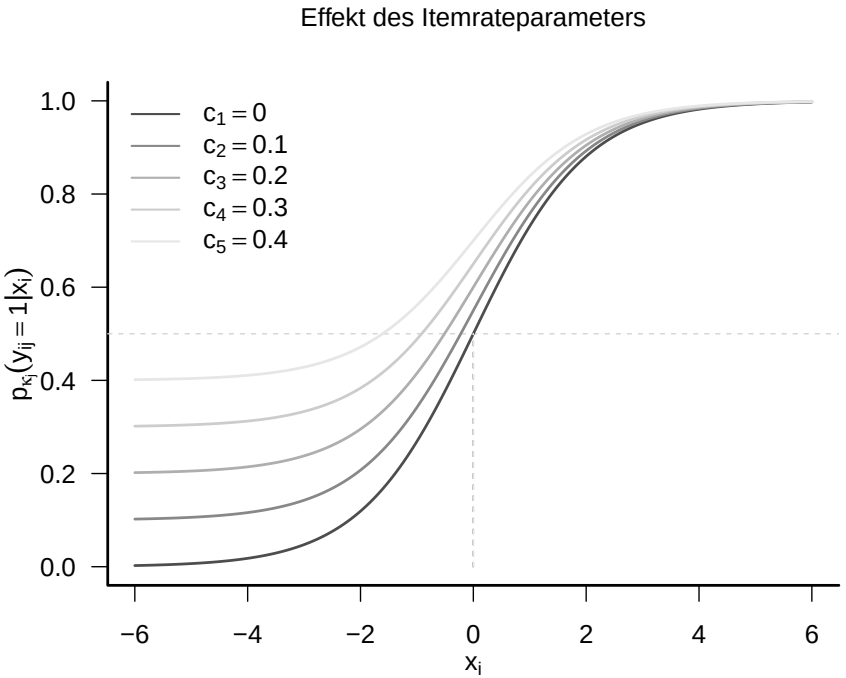
- Für $x_i^{(1)} < b_j < x_i^{(2)}$ gilt also $a_j \uparrow \Rightarrow p_\kappa\left(y_{ij}|x_i^{(1)}\right) - p_\kappa\left(y_{ij}|x_i^{(2)}\right) \uparrow$ bei identischer Differenz $x_i^{(2)} - x_i^{(1)}$
- In diesem Sinne diskriminiert ein Item mit höherem a_j stärker zwischen zwei Eigenschaftswerten $x_i^{(1)}, x_i^{(2)}$.



Birnbaum-Modell

Bemerkungen (Itemrateparameter)

- c_j entspricht $p_{\kappa_j}(y_{ij}|x_i)$ für $x_i \to -\infty$.
- Mit $c_j \neq 0$ ist das 3PL-Modell kein logistisches Modell
- c_j wird sicherlich besser als "Pseudorateparameter" gedacht.



Birnbaum-Modell

Bemerkungen (Itemrateparameter)

- Das 3PL-Modell ist kein "Logistisches Modell".
- Die Parameterinterpretationen sind Parameterinteraktionen unterworfen:
- $\Rightarrow$ Für $c_j \neq 0$ ist die Steigung in b_j auch von c_j abhängig.
- $\Rightarrow$ Für $c_j \neq 0$ ist der Wert x_i für $p_\kappa\left(y_{ij}|x_i\right) = 0.5$ auch von c_j abhängig.
- Die Interpretationen von a_j und b_j sind also nicht absolut.
- Als *Item*parameter c_j modelliert c_j nicht das Raten von individueller *Personen*.
- $c_j > 0$ mag auch daran liegen, dass Distraktoren auffällig sein können.
- $\Rightarrow$ "Welchen Durchmesser hat ein Mitochondrium?" (a) Etwa $1\mu m$ (b) Känguru

Birnbaum-Modell Datengeneration

```
# 3PL Modell Bernoulli-Erwartungswertparameter
mu = function(x,kappa){
  a    = kappa[1]                                          # Diskriminationsparameter
  b    = kappa[2]                                          # Schwierigkeitsparameter
  c    = kappa[3]                                          # Rateparameter
  mu   = c + (1-c)*plogis(a*(x-b))                         # Erwartungswertparamer
}

# Datengeneration
set.seed(0)                                                # Reproduzierbarkeit
n     = 1000                                               # Anzahl Personen
m     = 18                                                 # Anzahl Items
r     = 11                                                 # Anzahl Personeneigenschaftswerte (odd)
x     = -((r-1)/2):((r-1)/2)                               # Personeneigenschaftswerte
kappa = rbind(runif(m,0,2), runif(m,-3,3), runif(m,0,1))   # Itemparameter
pi    = dnorm(x,0,r/3)/sum(dnorm(x,0,r/3))                 # Personenverteilungsparameter
X     = matrix(rep(NaN, n),   nrow = n)                    # Latente PersonenEigenschaften
Y     = matrix(rep(NaN, n*m), nrow = n)                    # Observierbare Item-Responses
for(i in 1:n){                                             # Personeniterationen
  X[i]      = sample(x, size = 1, prob = pi)               # x_i \sim \prod_{s=1}^r \pi_s^{\ib{x_i = s}}
  for(j in 1:m){                                           # Itemiterationen
    Y[i,j]  = rbinom(1,1, mu(X[i], kappa[,j]))             # y_{ij} \sim Bern(\mu_{ij}^\kappa)
  }
}
colnames(y) = paste("I", 1:m)                              # Itemlabel
write.csv(Y, "./4-Daten/4-irt.csv", row.names = FALSE)     # Datenspeicherung
```

Anwendungsbeispiel

Simulierter Datensatz basierend auf den 18 dichotomen Items der GEB Scale (Kaiser (1998))

	I 1	I 2	I 3	I 4	I 5	I 6	I 7	I 8	I 9	I 10	I 11	I 12	I 13	I 14	I 15	I 16	I 17	I 18
P 1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 2	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 3	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P 4	0	1	0	1	0	1	1	1	0	0	1	0	1	0	1	0	1	0
P 5	1	1	0	1	1	1	1	1	0	1	1	0	1	1	0	0	1	1
P 6	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0
P 7	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0
P 8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P 9	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
P 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P 11	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1
P 12	0	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 13	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0
P 14	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0
P 15	1	1	0	0	1	0	0	1	1	0	0	0	1	0	1	0	1	0
P 16	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0
P 17	0	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0
P 18	0	1	0	1	1	1	0	1	1	1	1	0	0	1	1	0	1	0
P 19	1	1	0	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1
P 20	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1

$n = 20$ von $n = 1000$ Personen, keine KA Antworten

Überblick zur Parameterschätzung

Joint-Maximum-Likelihood-Schätzung (JML)

- Eher historische Anwendung nach Birnbaum in Lord & Novick (1968) für 1PL und 2PL, selten 3PL
- Personeneigenschaften x_i als feste unbekannte Parameter, nicht als latente Zufallsvariablen konzeptualisiert
- Iterative Maximierung der Likelihood-Funktion bezüglich Personeneigenschaften und Itemparameter
- Personeneigenschaftsparameteranzahl $x_1, ..., x_n$ steigt mit Stichprobenumfang
- $\Rightarrow$ Keine Konsistenz der Itemparameterschätzer

Marginal-Maximum-Likelihood-Schätzung (MML)

- 1PL (weniger zentral bei Rasch-Modellen), 2PL, 3PL
- Personeneigenschaften als latente Zufallsvariablen konzeptualisiert
- Marginalisierung der Likelihood über Personeneigenschaften zur sogenannten *Marginalen Likelihood*
- Itemparameterschätzung durch Maximierung der Marginalen Likelihood
- $\Rightarrow$ Konsistente und asymptotisch unverzerrte Itemparameterschätzungen

Überblick zur Marginal-Maximum-Likelihood-Schätzung

Im Rahmen der Marginal-Maximum-Likelihood (MML)-Schätzung wird das Birnbaum-Modell als parametrisiertes Modell mit latenten Zufallsvariablen konzipiert, wobei die Personeneigenschaften $x_i, i = 1, \ldots, n$ die Rolle der latenten Zufallsvariablen einnehmen. Seien formal dazu

- y die Gesamtheit aller nm beobachtbaren Datenpunkte,
- x die Gesamtheit aller n latenten Personeneigenschaften,
- θ die Gesamtheit aller Item- und Personenpopulationsparameter.

Dann entspricht ein Birnbaum-Modell einer gemeinsamen Wahrscheinlichkeitsverteilung $p_\theta(x, y)$. Grundlegende Idee der MML-Schätzung ist es "über die latenten Variablen zu mitteln" und analog zur Maximum-Likelihood-Schätzung zu fragen, welche Parameterwerte die sogenannte *Marginal-Likelihood-Funktion*

$$L : \theta \to \mathbb{R}_{\geq 0}, \theta \mapsto L(\theta) := p_\theta(y) = \int p_\theta(x, y)\, dx, \tag{19}$$

oder, äquivalent, die *Log-Marginal-Likelihood-Funktion*

$$\ell : \theta \to \mathbb{R}_{\geq 0}, \theta \mapsto \ell(\theta) := \ln p_\theta(y) = \ln \int p_\theta(x, y)\, dx \tag{20}$$

für einen festen beobachteten Datensatz y maximieren. Für das Birnbaum-Modell nutzt man dazu Varianten des sogenannten (exakten) *Expectation-Maximization-Algorithmus* (vgl. Bock & Aitkin (1981), Thissen (1982), Rigdon & Tsutakawa (1983)). Der exakte EM-Algorithmus ist eine Spezialform der koordinatenweisen ELBO-Maximierung der variationalen Inferenz (vgl. Ostwald et al. (2014), Blei & Smyth (2017), Starke & Ostwald (2017)). Wir führen zunächst den exakten EM-Algorithmus ein und betrachten dann seine Spezialisierung nach Woodruff (1996).

Theorem (Evidence Lower Bound)

Für einen Datensatz $y \in \mathbb{R}^{m \times n}$ sei $\ln p_\theta(y)$ die Log Marginal-Likelihood eines latenten Variablenmodells $p_\theta(x, y)$. Dann gilt für eine beliebige Verteilung $q(x)$ der latenten Variablen, dass

$$\ln p_\theta(y) \geq \int q(x) \ln p_\theta(x, y)\, dx - \int q(x) \ln q(x)\, dx =: \text{ELBO}(q(x), \theta).$$

$\text{ELBO}(q(x), \theta)$ heißt *Evidence Lower Bound*.

Beweis

Mit der Jensenschen Ungleichung (Theorem Jensensche Ungleichung) gilt

$$\ln p_\theta(y) := \ln \int p_\theta(x, y)\, dx = \ln \int q(x) \left(\frac{p_\theta(x, y)}{q(x)} \right) dx \geq \int q(x) \ln \left(\frac{p_\theta(x, y)}{q(x)} \right) dx.$$

Damit aber folgt

$$\begin{aligned} \ln p_\theta(y) \geq \int q(x) \ln \left(\frac{p_\theta(x, y)}{q(x)} \right) dx &= \int q(x) \left(\ln p_\theta(x, y) - \ln q(x) \right) dx \\ &= \int q(x) \ln p_\theta(x, y)\, dx - \int q(x) \ln q(x)\, dx. \end{aligned}$$

Birnbaum-Modell

Bemerkungen

- Für einen festen Datensatz y ist ELBO$(q(x), \theta)$ eine Funktion von $q(x)$ und θ.
- Die Bezeichnung Evidence Lower Bound geht auf den Begriff "Evidence" für $p_\theta(y)$ zurück.
- In den kognitiven Neurowissenschaften ist die ELBO weiterhin als "Freie Energie" bekannt.
- Die Signifikanz der ELBO geht weit über das Birnbaum-Modell hinaus.
- Die ELBO ist für die Variational Inference zentral.
- Variational Inference (VI) ist für zeitgenössische Gehirntheorien zentral (Parr et al. (2022)).
- Es gibt auch einen expliziten VI Zugang zum Birnbaum-Modell (vgl. Linden (2016) II.14, Wu et al. (2020)).
- Wir fokussieren hier auf den etwas traditionelleren MML/EM Ansatz

Definition (Expectation-Maximization-Algorithmus)

Die iterative koordinatenweise Maximierung der ELBO heißt *Expectation-Maximization (EM)-Algorithmus*.

Der Algorithmus hat die allgemeine Form:

EM-Algorithmus

0. Initialisierung von $q^{(0)}(x)$ und $\theta^{(0)}$

Für $k = 1, 2, \ldots$

1. E-Schritt: Setze $q^{(k)}(x) := \arg\max_{q(x)} \text{ELBO}\left(q(x), \theta^{(k-1)}\right)$
2. M-Schritt: Setze $\theta^{(k)} := \arg\max_{\theta} \text{ELBO}\left(q^{(k)}(x), \theta\right)$

Nach Konvergenz, nutze $\hat{\theta} := \theta^{(k)}$ als Parameterschätzer.

Bemerkungen

- "Expectation Schritt" ist eine Fehlbezeichnung, es handelt sich auch um einen Maximization Schritt...
- ... allerdings ergibt die Bezeichnung im sogenannten "exakten" EM-Algorithmus einigermaßen Sinn.

Theorem (Exakter Expectation-Maximization Algorithmus)

Das Setzen von

$$q^{(k)}(x) := p_{\theta^{(k-1)}}(x|y) \text{ für alle } k = 1, 2, \ldots \tag{21}$$

im E-Schritt des EM-Algorithmus maximiert die ELBO hinsichtlich $q(x)$ und heißt *exakter E-Schritt*.

Der Algorithmus hat dann die folgende Form

Exakter EM-Algorithmus

0. Initialisierung von $\theta^{(0)}$

Für $k = 1, 2, \ldots$

1. E-Schritt $q^{(k)}(x) := p_{\theta^{(k-1)}}(x|y)$
2. M-Schritt $\theta^{(k)} := \arg\max_\theta \int p_{\theta^{(k-1)}}(x|y) \ln p_\theta(x, y)\, dx$

Nach Konvergenz, nutze $\hat{\theta} := \theta^{(k)}$ als Parameterschätzer.

Bemerkungen

- Im Ansatz von Woodruff (1996) wird $p_{\theta^{(k-1)}}(x|y)$ "nichtparametrisch" evaluiert.
- Der M-Schritt des exakten EM-Algorithmus nach Woodruff (1996) benötigt numerische Optimierung.

Beweis

Wir zeigen zunächst, dass die ELBO $q^{(k)}(x) := p_{\theta^{(k-1)}}(x|y)$ den maximalen Wert $\ln p_{\theta^{(k-1)}}(y)$ annimmt:

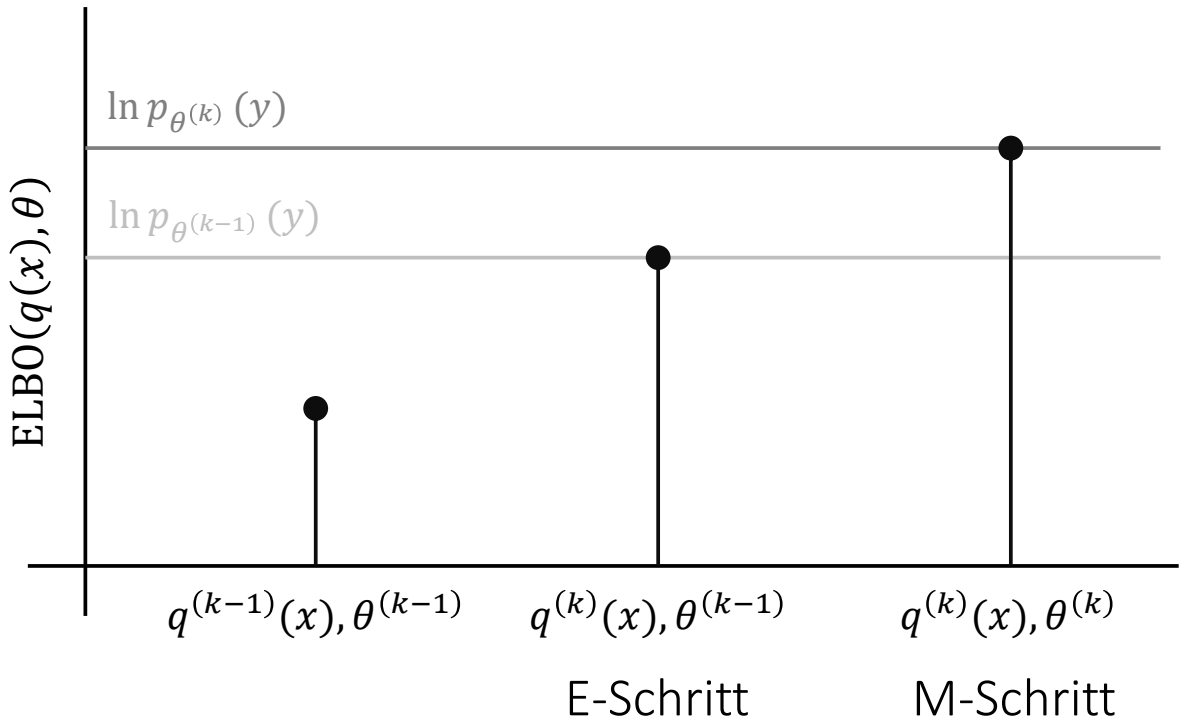
$$\begin{aligned}
\text{ELBO}(p_{\theta^{(k-1)}}(x|y), \theta) &= \int p_{\theta^{(k-1)}}(x|y) \ln \left(\frac{p_{\theta^{(k-1)}}(X,Y)}{p_{\theta^{(k-1)}}(x|y)} \right) dx \\
&= \int p_{\theta^{(k-1)}}(x|y) \ln \left(\frac{p_{\theta^{(k-1)}}(y) p_{\theta^{(k-1)}}(x|y)}{p_{\theta^{(k-1)}}(x|y)} \right) dx \\
&= \int p_{\theta^{(k-1)}}(x|y) \ln p_{\theta^{(k-1)}}(y)\, dx \\
&= \ln p_{\theta^{(k-1)}}(y) \int p_{\theta^{(k-1)}}(x|y)\, dx \\
&= \ln p_{\theta^{(k-1)}}(y).
\end{aligned}$$

Der M-Schritt hat dementsprechend die Form

$$\begin{aligned}
\theta^{(k)} &= \arg\max_{\theta} \text{ELBO} \left(p_{\theta^{(k-1)}}(x|y), \theta \right) \\
&= \arg\max_{\theta} \int p_{\theta^{(k-1)}}(x|y) \ln p_{\theta}(x,y)\, dx - \int p_{\theta^{(k-1)}}(x|y) \ln p_{\theta^{(k-1)}}(x|y)\, dx.
\end{aligned}$$

Der M-Schritt des exakten EM-Algorithmus folgt dann damit, dass der zweite Integralterm hier nicht von θ abhängt.

Visuelle Intuition



Bemerkungen

- Der M-Schritt der $k$ten Iteration des exakten EM-Algorithmus entspricht der Maximierung des Erwartungswertes der logarithmierten WDF der gemeinsamen Datenverteilung $p_\theta(x, y)$ hinsichtlich θ, wobei der Erwartungswert hinsichtlich der WDF der bedingten Datenverteilung von X gegeben y basierend auf der Parameterschätzung $\theta^{(k-1)}$, die in der $(k-1)$ten Iteration des exakten EM-Algorithmus gewonnen wurde.

- Überraschenderweise garantiert aufgrund der inhärenten Logik des EM-Algorithmus die Maximierung des Erwartungswertes

$$\mathbb{E}_{p_{\theta^{(k-1)}}(x|y)} \left(\ln p_\theta(x, y) \right) = \int p_{\theta^{(k-1)}}(x|y) \ln p_\theta(x, y) \, dx \tag{22}$$

  auch die Maximierung der tatsächlichen Funktion von Interesse, also von $\ln p_\theta(y)$.

- Für konkrete Algorithmen und für das Birnbaum-Modell muss obiger Erwartungswert als Funktion von $\theta^{(k-1)}$ ausgewertet werden und dann hinsichtlich θ maximiert werden, um die Parameterschätzung $\theta^{(k)}$ zu erhalten.

Birnbaum-Modell

Bock & Aitkin (1981) haben die MML-EM-Methode für das Birnbaum-Modell eingeführt

- Behandlung der Personeneigenschaften $x_1, ..., x_n$ als kontinuierliche normalverteilte latente Zufallsvariablen
- Evaluation des M-Schritt Likelihood-Integrals durch numerische Integration (Quadratur)

Woodruff (1996) hat die MML-EM-Methode für das Birnbaum-Modell neu konzeptualisiert

- Behandlung der Personeneigenschaften $x_1, ..., x_n$ als diskrete latente Zufallsvariablen
- Bock & Aitkin (1981) ist dann direkt äquivalent zum MML-EM-Algorithmus für endliche Mischungsmodelle
- Eindeutige Verortung von IRT innerhalb von endlichen Mischungsmodellen (vgl. Titterington et al. (1994))
- Es ist keine spezielle "Item-Response-Theorie-spezifische" EM-Theorie erforderlich (vgl. Murphy (2023))
- Viele Konvergenzeigenschaften des IRT-EM werden aus der Theorie endlicher Mischungsmodelle übernommen
- Wird das Modell als Mischung aufgefasst können nicht-normale latente Verteilungen geschätzt werden
- Insgesamt erleichtert Woodruff (1996) das Verständnis

Aktuelle **R** Pakete wie `ltm` oder `mirt` implementieren Bock & Aitkin (1981) und damit äquivalent Woodruff (1996)

Definition (Verteilungsform des Birnbaum-Modells nach Woodruff (1996))

Für $i = 1, ..., n$ Personen und $j = 1, ..., m$ dichotome Items seien

- x_i die Zufallsvariable zur Modellierung der Eigenschaft von Person i mit Wertebereich $\{1, ..., r\}$,
- $x := (x_1, ..., x_n)$ der Zufallsvektor der x_i,
- y_{ij} die Zufallsvariable zur Modellierung der $j$ten Itemantwort von Person i mit Wertebereich $\{0, 1\}$,
- $y_i := (y_{i1}, ..., y_{im})^T$ der Zufallsvektor der Itemantworten von Person i,
- $y := (y_1, ..., y_n)$ die Zufallsmatrix der y_i,
- $\kappa := (a_j, b_j, c_j)_{j=1,...,m}$ die Itemparameter des Birnbaum-Modells,
- π mit $\pi_s \geq 0$ und $\sum_{s=1}^{r} \pi_s = 1$ der Parameter der Personeneigenschaftsverteilung,
- $\theta := (\pi, \kappa)$ der Gesamtparameter.

Dann ist die *Verteilungsform des Birnbaum-Modells nach Woodruff (1996)* gegeben durch

$$p_\theta(x, y) := \prod_{i=1}^{n} p_\theta(x_i, y_i) := \prod_{i=1}^{n} p_\pi(x_i) p_\kappa(y_i|x_i) := \prod_{i=1}^{n} p_\pi(x_i) \prod_{j=1}^{m} p_{\kappa_j}(y_{ij}|x_i) \tag{23}$$

mit der *Personeneigenschaftsverteilung*

$$p_\pi(x_i) = \prod_{s=1}^{r} \pi_s^{[\![x_i=s]\!]} \tag{24}$$

und der *Item-Response-Verteilung*

$$p_{\kappa_j}(y_{ij}|x_i) := \text{Bern}\left(\mu_{ij}^\kappa\right) \text{ mit } \mu_{ij}^\kappa := c_j + (1 - c_j) \frac{1}{1 + \exp\left(-a_j(x_i - b_j)\right)}. \tag{25}$$

Birnbaum-Modell

Bemerkungen

- Die Personeneigenschaft nimmt hier die r diskreten Werte $s = 1, \ldots, r$ an.
- $[\![x_i = s]\!]$ bezeichnet die Iverson-Klammer mit $[\![x_i = s]\!] = 1$ für $x_i = s$ und $[\![x_i = s]\!] = 0$ sonst.
- Die Wahrscheinlichkeit, dass die Personeneigenschaft für Person i den Wert s annimmt ist also

$$p_\pi(x_i = s) = \prod_{s=1}^{r} \pi_s^{[\![x_i=s]\!]} = \pi_1^0 \cdot \ldots \cdot \pi_s^1 \cdot \ldots \cdot \pi_r^0 = 1 \cdot \ldots \cdot \pi_s \cdot \ldots \cdot 1 = \pi_s. \tag{26}$$

- Die Wahrscheinlichkeit dafür, dass x_i den Wert s annimmt ist also der $s$te Eintrag von π, π_s.
- Die Definition

$$p_\theta(x, y) := \prod_{i=1}^{n} p_\theta(x_i, y_i) \tag{27}$$

  enkodiert die Unabhängigkeitsannahme über Personen.
- Die Definition

$$p_\theta(x_i, y_i) := p_\pi(x_i) p_\kappa(y_i | x_i) \tag{28}$$

  enkodiert die partielle Wirksamkeit der Parameter π und κ.
- Die oft als "lokale Unkorreliertheit" bezeichnete Definition

$$p_\pi(x_i) p_\kappa(y_i | x_i) := p_\pi(x_i) \prod_{j=1}^{m} p_{\kappa_j}(y_{ij} | x_i) \tag{29}$$

  enkodiert die bedingte Unabhängigkeit der Itemantworten von Person i gegeben ihre Eigenschaft x_i.
- Die Verteilungsform des Birnbaum-Modells nach Bock & Aitkin (1981) nimmt normalverteilte Eigenschaften.
- Die dann im EM-Algorithmus nötige numerische Integration wird bei Woodruff (1996) vermieden.

Theorem (Birnbaum-Modell-EM-Algorithmus nach Woodruff (1996))

Gegeben sei die Verteilungsform des Birnbaum-Modells nach Woodruff (1996). Dann hat der exakte EM-Algorithmus zur Schätzung der Personeneigenschaftsverteilungparameter π und der Itemparameter κ die Form

0. Initialisierung von $\theta^{(0)} = \left(\pi^{(0)}, \kappa^{(0)}\right)$

Für $k = 1, 2, \ldots$

1. E-Schritt. Setze $q^{(k)}(x) := p_{\theta^{(k-1)}}(x|y)$ mit

$$p_{\theta^{(k-1)}}(x|y) = \prod_{i=1}^{n} p_{\theta^{(k-1)}}(x_i|y_i) \text{ mit } p_{\theta^{(k-1)}}(x_i = s|y_i) = \frac{\pi_s^{(k-1)} p_{\kappa^{(k-1)}}(y_i|x_i = s)}{\sum_{\tilde{s}=1}^{r} \pi_{\tilde{s}}^{(k-1)} p_{\kappa^{(k-1)}}(y_i|x_i = \tilde{s})} \tag{30}$$

   für alle $i = 1, \ldots, n$ und $s \in \{1, \ldots, r\}$.

2. M-Schritt. Es seien $\nu_s^{(k-1)} := \sum_{i=1}^{n} p_{\theta^{(k-1)}}(x_i = s|y_i)$ und $\rho_{js}^{(k-1)} := \sum_{i=1}^{n} y_{ij} p_{\theta^{(k-1)}}(x_i = s|y_i)$.
   Dann gilt für $\theta^{(k)} = \left(\pi^{(k)}, \kappa^{(k)}\right)$, dass

$$\pi_s^{(k)} := \frac{\nu_s^{(k-1)}}{\sum_{s=1}^{r} \nu_s^{(k-1)}} \text{ für } s = 1, \ldots, r \tag{31}$$

   und

$$\kappa^{(k)} = \text{argmax}_{\kappa} \sum_{s=1}^{r} \sum_{j=1}^{m} \left(\ln p_{\kappa_j}(y_{ij} = 1|x_i = s) \rho_{js}^{(k-1)} + \ln p_{\kappa_j}(y_{ij} = 0|x_i = s) \left(\nu_s^{(k-1)} - \rho_{js}^{(k-1)} \right) \right) \tag{32}$$

Beweis

(1) E-Schritt

Zur Vereinfachung der Notation setzen wir $\theta := \theta^{(k-1)}$. Wir halten dann zunächst fest, dass durch wiederholte Anwendung der Linearitätseigenschaften von Summen folgt, dass

$$\begin{aligned} p_\theta(y) &= \sum_X p_\theta(x, y) \\ &= \sum_{x_1} \cdots \sum_{x_n} p_\theta(x_1, ..., x_n, y_1, ..., y_n) \\ &= \sum_{x_1} \cdots \sum_{x_n} \left(\prod_{i=1}^{n} p_\theta(x_i, y_i) \right) \\ &= \prod_{i=1}^{n} \left(\sum_{x_i} p_\theta(x_i, y_i) \right) \\ &= \prod_{i=1}^{n} p_\theta(y_i). \end{aligned} \tag{33}$$

Dann aber gilt

$$p_\theta(x|y) = \frac{p_\theta(x, y)}{p_\theta(y)} = \frac{\prod_{i=1}^{n} p_\theta(x_i, y_i)}{\prod_{i=1}^{n} p_\theta(y_i)} = \prod_{i=1}^{n} \frac{p_\theta(x_i, y_i)}{p_\theta(y_i)} = \prod_{i=1}^{n} p_\theta(x_i|y_i). \tag{34}$$

Beweis (fortgeführt)

Weiterhin gilt

$$p_\theta(x_i|y_i) = \frac{p_\theta(x_i, y_i)}{p_\theta(y_i)} = \frac{p_\pi(x_i)p_\kappa(y_i|x_i)}{\sum_{x_i} p_\pi(x_i)p_\kappa(y_i|x_i)} \tag{35}$$

Insgesamt ergibt sich also

$$p_{\theta^{(k-1)}}(x|y) = \prod_{i=1}^{n} p_{\theta^{(k-1)}}(x_i|y_i) = \prod_{i=1}^{n} \frac{p_{\pi^{(k-1)}}(x_i)p_{\kappa^{(k-1)}}(y_i|x_i)}{\sum_{x_i} p_{\pi^{(k-1)}}(x_i)p_{\kappa^{(k-1)}}(y_i|x_i)} \tag{36}$$

Speziell ergibt sich für einen Wert $s \in \{1, \ldots, r\}$ der möglichen Werte von x_i, dass

$$p_{\theta^{(k-1)}}(x_i = s|y_i) = \frac{\pi_s^{(k-1)} p_{\kappa^{(k-1)}}(y_i|x_i = s)}{\sum_{\tilde{s}=1}^{r} \pi_{\tilde{s}}^{(k-1)} p_{\kappa^{(k-1)}}(y_i|x_i = \tilde{s})} \tag{37}$$

Beweis (fortgeführt)

(2) M-Schritt

Wir werten zunächst das im M-Schritt bezüglich θ zu maximierende Integral

$$\int p_{\theta^{(k-1)}}(x|y) \ln p_\theta(x, y)\, dx, \tag{38}$$

das aufgrund der diskreten Natur der Personeneigenschaften im vorliegenden Fall einer Summe über die möglichen Werte von X entspricht, aus. Dabei ergibt sich mit dem Ergebnis des E-Schritts und der Definition von $p_\theta(x, y)$, dass sich dieses Integral als Summe einer Funktion der Personeneigenschaftsparameter $f(\pi)$ und einer Funktion der Itemparameter $g(\kappa)$ schreiben lässt, welche sich dann in der Folge unabhängig voneinander bezüglich der Komponenten von θ maximieren lassen. Es ergibt sich

Beweis (fortgeführt)

$$
\begin{aligned}
\int p_{\theta(k-1)}(x|y) \ln p_\theta(x,y)\, dx &= \sum_{x_1} \cdots \sum_{x_n} \left(p_{\theta(k-1)}(x_1,...,x_n|y_1,...,y_n) \ln \left(\prod_{i=1}^{n} p_\theta(x_i,y_i) \right) \right) \\
&= \sum_{x_1} \cdots \sum_{x_n} \left(p_{\theta(k-1)}(x_1,...,x_n|y_1,...,y_n) \sum_{i=1}^{n} \ln p_\theta(x_i,y_i) \right) \\
&= \sum_{x_1} \cdots \sum_{x_n} \sum_{i=1}^{n} p_{\theta(k-1)}(x_1,...,x_n|y_1,...,y_n) \ln p_\theta(x_i,y_i) \\
&= \sum_{i=1}^{n} \left(\sum_{x_1} \cdots \sum_{x_n} p_{\theta(k-1)}(x_1,...,x_n|y_1,...,y_n) \ln p_\theta(x_i,y_i) \right) \\
&= \sum_{i=1}^{n} \sum_{x_i} p_{\theta(k-1)}(x_i|y_i) \ln p_\theta(x_i,y_i) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta(k-1)}(x_i=s|y_i) \ln p_\theta(x_i=s,y_i) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta(k-1)}(x_i=s|y_i) \ln \left(p_\pi(x_i=s) p_\kappa(y_i|x_i=s) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta(k-1)}(x_i=s|y_i) \left(\ln p_\pi(x_i=s) + \ln p_\kappa(y_i|x_i=s) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta(k-1)}(x_i=s|y_i) \ln p_\pi(x_i=s) + \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta(k-1)}(x_i=s|y_i) \ln p_\kappa(y_i|x_i=s) \\
&=: f(\pi) + g(\kappa)
\end{aligned}
$$

Beweis (fortgeführt)

Dabei ergibt sich die fünfte Gleichung beispielhaft für $n := 3$ und $i := 1$ durch

$$\begin{aligned}
&\sum_{x_1}\sum_{x_2}\sum_{x_3} p_{\theta^{(k-1)}}(x_1, x_2, x_3|y_1, y_2, y_3) \ln p_\theta(x_1, y_1) \\
&= \sum_{x_1} \ln p_\theta(x_1, y_1) \sum_{x_2}\sum_{x_3} p_{\theta^{(k-1)}}(x_1, x_2, x_3|y_1, y_2, y_3) \\
&= \sum_{x_1} \ln p_\theta(x_1, y_1) p_{\theta^{(k-1)}}(x_1|y_1, y_2, y_3) \\
&= \sum_{x_1} p_{\theta^{(k-1)}}(x_1|y_1) \ln p_\theta(x_1, y_1)
\end{aligned} \tag{39}$$

wobei sich die letzte Gleichung mit der bedingten Unabhängigkeit von x_1 von y_2, y_3 gemäß des Beweises des E-Schritts ergibt.

Es verbleiben also die Maximierung von f bezüglich π und von g bezüglich κ. Dabei kann die Maximierung von f bezüglich der Personeneigenschaftsparameter π mit einem Lagrangeansatz auf analytische Weise durchgeführt werden, wohingegen die Maximierung von g bezüglich der Itemparameter κ eine Auslagerung in die numerische Optimierung benötigt.

Beweis (fortgeführt)

Wir vereinfachen die Schreibweisen der Funktionen f und g noch etwas. Speziell gelten

$$\begin{aligned} f(\pi) &= \sum_{i=1}^{n}\sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i)\ln p_\pi(x_i = s) \\ &= \sum_{s=1}^{r}\sum_{i=1}^{n} p_{\theta^{(k-1)}}(x_i = s|y_i)\ln p_\pi(x_i = s) \\ &= \sum_{s=1}^{r}\sum_{i=1}^{n} p_{\theta^{(k-1)}}(x_i = s|y_i)\ln(\pi_s) \qquad (40) \\ &= \sum_{s=1}^{r}\ln(\pi_s)\sum_{i=1}^{n} p_{\theta^{(k-1)}}(x_i = s|y_i) \\ &= \sum_{s=1}^{r}\ln(\pi_s)\nu_s^{(k-1)} \end{aligned}$$

mit

$$\nu_s^{(k-1)} := \sum_{i=1}^{n} p_{\theta^{(k-1)}}(x_i = s|y_i). \qquad (41)$$

Beweis (fortgeführt)

Weiterhin gilt mit $\mu_{sj}^{\kappa} = \mu_{ij}^{\kappa}$ für $x_i = s$, dass

$$
\begin{aligned}
g(\kappa) &= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \ln p_{\kappa}(y_i|x_i = s) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \ln \left(\prod_{j=1}^{m} p_{\kappa_j}(y_{ij}|x_i = s) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \left(\sum_{j=1}^{m} \ln \left(p_{\kappa_j}(y_{ij}|x_i = s) \right) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \left(\sum_{j=1}^{m} \ln \left(\left(\mu_{sj}^{\kappa} \right)^{y_{ij}} \left(1 - \mu_{sj}^{\kappa} \right)^{1-y_{ij}} \right) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \left(\sum_{j=1}^{m} \left(y_{ij} \ln \left(\mu_{sj}^{\kappa} \right) + \left(1 - y_{ij} \right) \ln \left(1 - \mu_{sj}^{\kappa} \right) \right) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \left(\sum_{j=1}^{m} \left(y_{ij} \ln \left(\mu_{sj}^{\kappa} \right) \right) + \sum_{j=1}^{m} \left(1 - y_{ij} \right) \ln \left(1 - \mu_{sj}^{\kappa} \right) \right) \\
&= \sum_{i=1}^{n} \sum_{s=1}^{r} p_{\theta^{(k-1)}}(x_i = s|y_i) \left(\sum_{j=1}^{m} \left(y_{ij} \ln \left(\mu_{sj}^{\kappa} \right) \right) + \sum_{j=1}^{m} \ln \left(1 - \mu_{sj}^{\kappa} \right) - \sum_{j=1}^{m} y_{ij} \ln \left(1 - \mu_{sj}^{\kappa} \right) \right)
\end{aligned}
$$

Beweis (fortgeführt)

und damit

$$
\begin{aligned}
g(\kappa) = & \sum_{i=1}^{n}\sum_{s=1}^{r} p_{\theta(k-1)}(x_i = s|y_i) \sum_{j=1}^{m} \left(y_{ij} \ln\left(\mu_{sj}^{\kappa}\right)\right) + \\
& \sum_{i=1}^{n}\sum_{s=1}^{r} p_{\theta(k-1)}(x_i = s|y_i) \sum_{j=1}^{m} \ln\left(1-\mu_{sj}^{\kappa}\right) - \\
& \sum_{i=1}^{n}\sum_{s=1}^{r} p_{\theta(k-1)}(x_i = s|y_i) \sum_{j=1}^{m} y_{ij} \ln\left(1-\mu_{sj}^{\kappa}\right) \\
= & \sum_{i=1}^{n}\sum_{s=1}^{r}\sum_{j=1}^{m} p_{\theta(k-1)}(x_i = s|y_i) \left(y_{ij} \ln\left(\mu_{sj}^{\kappa}\right)\right) + \\
& \sum_{i=1}^{n}\sum_{s=1}^{r}\sum_{j=1}^{m} p_{\theta(k-1)}(x_i = s|y_i) \ln\left(1-\mu_{sj}^{\kappa}\right) - \\
& \sum_{i=1}^{n}\sum_{s=1}^{r}\sum_{j=1}^{m} p_{\theta(k-1)}(x_i = s|y_i) y_{ij} \ln\left(1-\mu_{sj}^{\kappa}\right) \\
= & \sum_{s=1}^{r}\sum_{j=1}^{m}\sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) y_{ij} \ln\left(\mu_{sj}^{\kappa}\right) + \\
& \sum_{s=1}^{r}\sum_{j=1}^{m}\sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) \ln\left(1-\mu_{sj}^{\kappa}\right) - \\
& \sum_{s=1}^{r}\sum_{j=1}^{m}\sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) y_{ij} \ln\left(1-\mu_{sj}^{\kappa}\right)
\end{aligned}
$$

Beweis (fortgeführt)

und weiter

$$
\begin{aligned}
g(\kappa) = & \sum_{s=1}^{r}\sum_{j=1}^{m} \ln\left(\mu_{sj}^{\kappa}\right) \sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) y_{ij} + \\
& \sum_{s=1}^{r}\sum_{j=1}^{m} \ln\left(1-\mu_{sj}^{\kappa}\right) \sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) - \\
& \sum_{s=1}^{r}\sum_{j=1}^{m} \ln\left(1-\mu_{sj}^{\kappa}\right) \sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) y_{ij} \\
= & \sum_{s=1}^{r}\sum_{j=1}^{m} \ln\left(\mu_{sj}^{\kappa}\right) \rho_{js}^{(k-1)} + \sum_{s=1}^{r}\sum_{j=1}^{m} \ln\left(1-\mu_{sj}^{\kappa}\right) \nu_{s}^{(k-1)} - \sum_{s=1}^{r}\sum_{j=1}^{m} \ln\left(1-\mu_{sj}^{\kappa}\right) \rho_{js}^{(k-1)} \\
= & \sum_{s=1}^{r}\sum_{j=1}^{m} \left(\ln\left(\mu_{sj}^{\kappa}\right) \rho_{js}^{(k-1)} + \ln\left(1-\mu_{sj}^{\kappa}\right) \left(\nu_{s}^{(k-1)} - \rho_{js}^{(k-1)}\right)\right)
\end{aligned}
$$

mit

$$
\nu_{s}^{(k-1)} := \sum_{i=1}^{n} p_{\theta(k-1)}(x_i = s|y_i) \text{ und } \rho_{js}^{(k-1)} := \sum_{i=1}^{n} y_{ij} p_{\theta(k-1)}(x_i = s|y_i) \tag{42}
$$

Beweis (fortgeführt)

(3) M-Schritt Optimierung von $f(\pi)$

Wir sind auf das restringierte Optimierungsproblem

$$\max f(\pi) \text{ unter der Nebenbedingung } \sum_{s=1}^{r} \pi_s = 1 \tag{43}$$

für

$$f : \Pi^r \to \mathbb{R}, \pi \mapsto f(\pi) := \sum_{s=1}^{r} \ln(\pi_s)\nu_s^{(k-1)} \tag{44}$$

geführt, wobei Π^r den r-dimensionalen Wahrscheinlichkeitssimplex

$$\Pi^r := \left\{ \pi \in \mathbb{R}^r \middle| \pi_s \geq 0 \text{ für } s = 1, \ldots, r \text{ und } \sum_{s=1}^{r} \pi_s = 1 \right\} \tag{45}$$

bezeichnen soll. Wir erinnern zunächst an den Lagrangeansatz.

Lagrangeansatz zur Optimierung einer multivariaten Funktion unter Nebenbedingungen

Es sei

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \tag{46}$$

die Lagrangefunktion zur Zielfunktion f und den Nebenbedingungsfunktionen g_i. Dann gelten an der Stelle eines lokalen Extremums x^*, das die Nebenbedingungen erfüllt, die folgenden notwendigen Bedingungen:

(1) $\nabla_x L(x^*, \lambda^*) = 0 \Leftrightarrow \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$

(2) $\nabla_\lambda L(x^*, \lambda^*) = 0$

(3) $g_i(x^*) = 0$ für alle $i = 1, \ldots, m$

Beweis (fortgeführt)

Mit

$$\sum_{s=1}^{r} \pi_s = 1 \Leftrightarrow \sum_{s=1}^{r} \pi_s - 1 = 0 \Leftrightarrow: g_1(\pi) = 0 \tag{47}$$

ergibt sich die Lagrangefunktion von (43) als

$$L(\pi, \lambda) = \sum_{s=1}^{r} \ln(\pi_s)\nu_s^{(k-1)} + \lambda \left(\sum_{s=1}^{r} \pi_s - 1 \right) = \sum_{s=1}^{r} \ln(\pi_s)\nu_s^{(k-1)} + \lambda \sum_{s=1}^{r} \pi_s - \lambda. \tag{48}$$

Für die $s = 1, .., r$ Komponenten des Gradienten $\nabla_\pi L(\pi, \lambda)$ ergibt sich

$$\begin{aligned}
\frac{\partial}{\partial \pi_s} L(\pi, \lambda) &= \frac{\partial}{\partial \pi_s} \left(\sum_{s'=1}^{r} \ln(\pi_{s'})\nu_{s'}^{(k-1)} + \lambda \sum_{s'=1}^{r} \pi_{s'} - \lambda \right) \\
&= \frac{\partial}{\partial \pi_s} \sum_{s'=1}^{r} \ln(\pi_{s'})\nu_{s'}^{(k-1)} + \lambda \frac{\partial}{\partial \pi_s} \sum_{s'=1}^{r} \pi_{s'} \\
&= \frac{\partial}{\partial \pi_s} \ln(\pi_s)\nu_s^{(k-1)} + \lambda \frac{\partial}{\partial \pi_s} \pi_s \\
&= \frac{\nu_s^{(k-1)}}{\pi_s} + \lambda
\end{aligned} \tag{49}$$

Beweis (fortgeführt)

Für die eine Komponente des Gradienten $\nabla_\lambda L(\pi, \lambda)$ ergibt sich

$$\begin{aligned}\frac{\partial}{\partial\lambda} L(\pi,\lambda) &= \frac{\partial}{\partial\lambda}\left(\sum_{s=1}^{r} \ln(\pi_s)\nu_s^{(k-1)} + \lambda\sum_{s=1}^{r}\pi_s - \lambda\right) \\ &= \frac{\partial}{\partial\lambda}\sum_{s=1}^{r} \ln(\pi_s)\nu_s^{(k-1)} + \frac{\partial}{\partial\lambda}\lambda\sum_{s=1}^{r}\pi_s - \frac{\partial}{\partial\lambda}\lambda \\ &= \sum_{s=1}^{r}\pi_s - 1\end{aligned} \tag{50}$$

Mit der notwendigen Bedingung (1) gilt dann

$$\nabla_\pi L(\pi^*,\lambda^*) = 0 \Leftrightarrow \frac{\nu_s^{(k-1)}}{\pi_s^*} + \lambda^* = 0 \text{ für } s = 1,\ldots,r \Leftrightarrow \pi_s^* = -\frac{\nu_s^{(k-1)}}{\lambda^*} \text{ für } s = 1,\ldots,r \tag{51}$$

Mit der notwendigen Bedingung (2) (sowie auch mit der notwendigen Bedingung (3)) gilt

$$L_\lambda(\pi^*,\lambda^*) = 0 \Leftrightarrow \sum_{s=1}^{r}\pi_s^* - 1 = 0 \Leftrightarrow \sum_{s=1}^{r}\pi_s^* = 1. \tag{52}$$

Einsetzen von (1) in (2) erlaubt dann das Auflösen nach dem Lagrangemultiplikator

$$\sum_{s=1}^{r}\pi_s^* = 1 \Leftrightarrow -\sum_{s=1}^{r}\frac{\nu_s^{(k-1)}}{\lambda^*} = 1 \Leftrightarrow -\frac{1}{\lambda^*}\sum_{s=1}^{r}\nu_s^{(k-1)} = 1 \Leftrightarrow \lambda^* = -\sum_{s=1}^{r}\nu_s^{(k-1)} \tag{53}$$

Beweis (fortgeführt)

Einsetzen des Ergebnisses für λ^* in die Umformulierung von (1) ergibt dann die Lösung

$$\pi_s^* = \frac{\nu_s^{(k-1)}}{\sum_{s=1}^{r} \nu_s^{(k-1)}} \text{ für } s = 1, \ldots, r \tag{54}$$

und entsprechend die Parameterupdategleichung

$$\pi_s^{(k)} := \frac{\nu_s^{(k-1)}}{\sum_{s=1}^{r} \nu_s^{(k-1)}} \text{ für } s = 1, \ldots, r \tag{55}$$

Beweis (fortgeführt)

(4) M-Schritt Optimierung von $g(\kappa)$

Wir sind auf das restringierte Optimierungsproblem

$$\max g(\kappa) \text{ unter den Nebenbedingungen } \kappa_j \in \mathbb{R}_{>0} \times \mathbb{R} \times [0,1] \text{ für } j = 1, \ldots, m \tag{56}$$

mit

$$g : \{\mathbb{R}_{>0} \times \mathbb{R} \times [0,1]\}^m \to \mathbb{R}, \kappa \mapsto g(\kappa) := \sum_{s=1}^{r} \sum_{j=1}^{m} \left(\ln\left(\mu_{sj}^{\kappa}\right) \rho_{js}^{(k-1)} + \ln\left(1 - \mu_{sj}^{\kappa}\right) \left(\nu_s^{(k-1)} - \rho_{js}^{(k-1)} \right) \right) \tag{57}$$

mit

$$\nu_s^{(k-1)} := \sum_{i=1}^{n} p_{\theta^{(k-1)}}(x_i = s|y_i) \text{ und } \rho_{js}^{(k-1)} := \sum_{i=1}^{n} y_{ij} p_{\theta^{(k-1)}}(x_i = s|y_i) \tag{58}$$

sowie

$$p_{\theta^{(k-1)}}(x_i = s|y_i) = \frac{\pi_s^{(k-1)} p_{\kappa^{(k-1)}}(y_i|x_i = s)}{\sum_{\tilde{s}=1}^{r} \pi_{\tilde{s}}^{(k-1)} p_{\kappa^{(k-1)}}(y_i|x_i = \tilde{s})} \tag{59}$$

geführt. Mit

$$p_{\kappa}(y_{ij} = 1|x_i = s) = \mu_{sj}^{\kappa} \text{ und } p_{\kappa}(y_{ij} = 0|x_i = s) = 1 - \mu_{sj}^{\kappa} \tag{60}$$

ergibt sich dann die behauptete Form des M-Schritts. □

Grundlagen

Birnbaum-Modell

Rasch-Modell

Selbstkontrollfragen

Überblick

Geschichte

- Entwickelt in den 1950er Jahren in der dänischen Bildungsforschung durch Georg Rasch (1901-1980)
- Für Anwender:innen popularisiert durch z.B. Wright & Panchapakesan (1969) und Wright (1977)

Grundlegende Motivationsrhethorik

- "In the present work a new approach to test-psychology is attempted" (Rasch (1960))
- "Spezifisch objektive" Messung psychologischer Eigenschaften von Einzelpersonen
- Verzicht auf Schätzung von "Populationsparametern" im Sinne der klassischen Testtheorie

Rhethorische messtheoretische Anbindung an physikalische Messungen

- Vergleiche von Objekteigenschaften (Masse, Temperatur) sollten unabhängig von Messinstrumenten sein
- Vergleiche von Messinstrumenten (Maßband, Thermometer) sollten unabhängig von Objekten sein

Invariante Vergleichseigenschaften bei psychologischen Messungen

- Personenvergleiche sollten unabhängig von den verwendeten Items sein.
- Itemvergleiche sollten unabhängig von den verwendeten Personen sein.

Aus dem Vorwort von B.D. Wright in Rasch (1960): "The psychometric research done by Rasch between 1951 and 1959, which he explains and illustrates in this book, marks the point at which psychometrics moved from being purely descriptive to become a science of objective measurement."

Überblick

Zum Begriff der Spezifischen Objektivität des Rasch-Modells

- Rasch bezeichnete die invarianten Vergleichseigenschaften seines Modells als "Spezifische Objektivität"
- Rhetorisches Ziel war die Abgrenzung von naiver Messobjektivität als auch von klassischen Testtheorien
- Grundidee der spezifischen Objektivität sind obige invariante Vergleichseigenschaften
- Dies gelten allerdings für das Modell, nicht notwendigerweise für die Realität

Die Objektivität ist hier "spezifisch" in Bezug

... auf die betrachteten Vergleiche

- Objektivität gilt immer für einzelne konkrete Vergleiche, wie Person 1 vs. Person 2 und Item 1 vs. Item 2

... auf das Modell, d.h. sie gilt nur wenn

- das Rasch-Modell korrekt spezifiziert ist und die Annahmen des Modells erfüllt sind

... auf die Datenbasis

- Personenparameter sind objektiv gegeben die Items, Itemparameter sind objektiv gegeben die Personen

"Messung" als Konstruktion von Bedingungen, unter denen Vergleiche sinnvoll und invariant sind.

Das Rasch-Modell hat Bezüge zur additiven-konjunkten Messung der Repräsentationstheorie (Luce & Tukey (1964))

Das Rasch-Modell ist mit der additiven-konjunkten Messung aber nicht identisch, vgl. Kyngdon (2008)

Definition (Rasch-Modell)

Für $i = 1, \ldots, n$ Personen und $j = 1, \ldots, m$ dichotome Items seien

- $\xi_i > 0$ die *Personeneigenschaft* von Person i,
- $\beta_j > 0$ der *Schwierigkeitsparameter* von Item j,
- $\lambda_{ij} := \frac{\xi_i}{\beta_j}$ der *Situationsparameter* von Person i und Item j
- $\lambda_i := (\lambda_{ij})_{j=1,\ldots,m}$ der *Situationsparametervektor* von Person i,
- y_{ij} die Zufallsvariable zur Modellierung der $j$ten Itemantwort von Person i mit Wertebereich $\{0, 1\}$,
- $y_i := (y_{i1}, \ldots, y_{im})^T$ der Zufallsvektor der Itemantworten von Person i,

Dann heißt die Item-Response-Verteilung

$$p_{\lambda_i}(y_i) := \prod_{j=1}^{m} p_{\lambda_{ij}}(y_{ij}) \text{ mit } p_{\lambda_{ij}}(y_{ij}) := \left(\frac{\lambda_{ij}}{1+\lambda_{ij}}\right)^{y_{ij}} \left(\frac{1}{1+\lambda_{ij}}\right)^{1-y_{ij}} \tag{61}$$

Rasch-Modell.

Bemerkungen

- Im Gegensatz zum (zeitgenössischen) Birnbaum-Modell sind die Personeneigenschaften keine Zufallsvariablen
- Notationell schreiben wir $p(y|\theta)$, wenn θ zufällig ist, und $p_\theta(y)$, wenn es dies nicht ist
- $\prod_{j=1}^{m} p_{\lambda_{ij}}(y_{ij})$ modelliert die personeneigenschaftsbedingte Unabhängigkeit der Itemantworten

Bemerkungen (fortgeführt)

- Offenbar gelten für alle $i = 1, \ldots, n$ und $j = 1, \ldots, m$

$$p_{\lambda_{ij}}(y_{ij} = 1) = \frac{\lambda_{ij}}{1 + \lambda_{ij}} \tag{62}$$

sowie

$$p_{\lambda_{ij}}(y_{ij} = 0) = \frac{1}{1 + \lambda_{ij}} = \frac{1 + \lambda_{ij}}{1 + \lambda_{ij}} - \frac{\lambda_{ij}}{1 + \lambda_{ij}} = 1 - \frac{\lambda_{ij}}{1 + \lambda_{ij}} \tag{63}$$

Invariante Vergleichseigenschaften auf Parameterebene

- Der Vergleich der Personeneigenschaft zweier Personen $i = 1, 2$ ist unabhängig von dem betrachteten Item j:
- Für jedes Item $j = 1, \ldots, m$ gilt

$$\frac{\lambda_{1j}}{\lambda_{2j}} = \frac{\frac{\xi_1}{\beta_j}}{\frac{\xi_2}{\beta_j}} = \frac{\xi_1}{\beta_j} \cdot \frac{\beta_j}{\xi_2} = \frac{\xi_1}{\xi_2} \tag{64}$$

- Der Vergleich der Itemschwierigkeiten zweier Items $j = 1, 2$ ist unabhängig von der betrachteten Person i:
- Für jede Person $i = 1, \ldots, n$ gilt

$$\frac{\lambda_{i1}}{\lambda_{i2}} = \frac{\frac{\xi_i}{\beta_1}}{\frac{\xi_i}{\beta_2}} = \frac{\xi_i}{\beta_1} \cdot \frac{\beta_2}{\xi_i} = \frac{\beta_2}{\beta_1} \tag{65}$$

Theorem (Rasch-Modell und 1PL-Birnbaum-Modell)

Gegeben sei das Rasch-Modell. Seien ferner

$$x_i := \ln(\xi_i) \text{ und } b_j := \ln(\beta_j) \tag{66}$$

und x_i als Zufallsvariable konzipiert. Dann sind die Item-Response-Verteilungen des Rasch-Modells und des 1PL-Birnbaum-Modells identisch, es gilt also

$$p_{\lambda_i}(y_i) = p_\kappa(y_i|x_i). \tag{67}$$

Bemerkungen

- Im zeitgenössischen Birnbaum-Modell sind die Personeneigenschaften Zufallsvariablen, im Rasch-Modell nicht.
- Für eine Person i und ein Item j gelten nach dem Rasch-Modell also wie nach dem 1PL-Birnbaum-Modell

$$p_{x_i,b_j}\left(y_{ij}=1\right) = \frac{1}{1+\exp(-(x_i-b_j))} \text{ und } p_{x_i,b_j}\left(y_{ij}=0\right) = 1 - \frac{1}{1+\exp(-(x_i-b_j))} \tag{68}$$

Beweis

Wir zeigen die Äquivalenz für $y_{ij} = 1$, der Fall $y_{ij} = 0$ folgt analog:

$$
\begin{aligned}
p_{\lambda_{ij}}(y_{ij} = 1) &= \frac{\lambda_{ij}}{1 + \lambda_{ij}} \\
&= \frac{\frac{\xi_i}{\beta_j}}{1 + \frac{\xi_i}{\beta_j}} \\
&= \frac{\frac{\exp(\ln(\xi_i))}{\exp(\ln(\beta_j))}}{1 + \frac{\exp(\ln(\xi_i))}{\exp(\ln(\beta_j))}} \\
&= \frac{\exp(\ln(\xi_i) - \ln(\beta_j))}{1 + \exp(\ln(\xi_i) - \ln(\beta_j))} \\
&= \frac{\exp(x_i - b_j)}{1 + \exp(x_i - b_j)} \\
&= \frac{1}{1 + \exp(-(x_i - b_j))} \\
&= p_{\kappa_j}(y_{ij} = 1 | x_i) \text{ mit } \kappa_j = (1, b_j, 0)
\end{aligned}
\tag{69}
$$

Im letzten Schritt haben wir die Personeneigenschaft dabei als Zufallsvariable konzipiert. Die lokale Unkorreliertheit von Birnbaum- und Rasch-Modell impliziert dann die im Theorem behauptete Äquivalenz.

Rasch-Modell

Bemerkungen (fortgeführt)

- Für eine Person i und ein Item j gilt nach dem Rasch-Modell gilt also wie nach dem 1PL-Birnbaum-Modell

$$p_{x_i,b_j}\left(y_{ij}=1\right)=\frac{\exp(x_i-b_j)}{1+\exp(x_i-b_j)}=\frac{\exp(1\cdot(x_i-b_j))}{1+\exp(x_i-b_j)}=\frac{\exp(y_{ij}(x_i-b_j))}{1+\exp(x_i-b_j)} \tag{70}$$

  sowie

$$\begin{aligned}
p_{x_i,b_j}\left(y_{ij}=0\right)&=1-\frac{\exp(x_i-b_j)}{1+\exp(x_i-b_j)}\\
&=\frac{1+\exp(x_i-b_j)}{1+\exp(x_i-b_j)}-\frac{\exp(x_i-b_j)}{1+\exp(x_i-b_j)}\\
&=\frac{1}{1+\exp(x_i-b_j)}\\
&=\frac{\exp(0)}{1+\exp(x_i-b_j)}\\
&=\frac{\exp(0\cdot(x_i-b_j))}{1+\exp(x_i-b_j)}\\
&=\frac{\exp(y_{ij}(x_i-b_j))}{1+\exp(x_i-b_j)}
\end{aligned} \tag{71}$$

- Man schreibt deshalb

$$p_{x_i,b_j}\left(y_{ij}\right)=\frac{\exp(y_{ij}(x_i-b_j))}{1+\exp(x_i-b_j)} \tag{72}$$

Rasch-Modell

Bemerkungen (fortgeführt)

Invariante Vergleichseigenschaften auf der Ebene von Odds-Ratios

- Bei dichotomen Ereignissen $y = 0$ oder $y = 1$ wird häufig von *Odds* (Chancen) gesprochen
- Odds bezeichnen das Verhältnis der Eintrittschance zur Nichteintrittschance,

$$\text{Odds} := \frac{p(y=1)}{p(y=0)} \tag{73}$$

- Zum Beispiel gilt

$$p(y=1) = \frac{1}{4} \text{ und } p(y=0) = \frac{3}{4} \Rightarrow \text{ Odds } = \frac{1}{4} \cdot \frac{4}{3} = \frac{1}{3} \text{ ("Chance von 1 zu 3").} \tag{74}$$

- *Odds-Ratio (OR)* vergleicht die Odds zwischen zwei Dingen

$$\text{Odds-Ratio } := \frac{\text{Odds}_1}{\text{Odds}_2}. = \frac{\frac{p_1(y=1)}{p_1(y=0)}}{\frac{p_2(y=1)}{p_2(y=0)}} \tag{75}$$

  Zum Beispiel gilt bei Odds von 1:4 und 1:10 unter zwei Bedingungen

$$\text{Odds-Ratio } := \frac{\text{Odds}_1}{\text{Odds}_2} = \frac{\frac{1}{4}}{\frac{1}{10}} = 2.5 \tag{76}$$

- Die erste Bedingung (1:4) weist 2,5-fach höhere Odds für das Ereignis auf als die zweite Bedingung (1:10).
- Das Odds Ratio wird häufig in klinischen Fall-Kontroll-Studien verwendet
- Popularität von Odds-Ratios als Belege für Assoziation von Rauchen und Lungenkrebs (vgl. Doll & Hill (1950))

Rasch-Modell

Bemerkungen (fortgeführt)

Invariante Vergleichseigenschaften auf der Ebene von Odds-Ratios

- Die Odds für eine richtige Antwort von Person i in Item j sind nach dem Rasch-Modell

$$\text{Odds}_{ij} = \frac{p_{x_i,b_j}\left(y_{ij}=1\right)}{p_{x_i,b_j}\left(y_{ij}=0\right)} = \frac{\frac{\exp(1\cdot(x_i-b_j))}{1+\exp(x_i-b_j)}}{\frac{\exp(0\cdot(x_i-b_j))}{1+\exp(x_i-b_j)}} = \frac{\exp(x_i-b_j)}{1+\exp(x_i-b_j)} \cdot \left(1+\exp(x_i-b_j)\right) = \exp(x_i-b_j)$$

- Die Odds-Ratio für eine richtige Antwort von Person i und Person i' in Item j ist nach dem Rasch-Modell

$$\text{Odds-Ratio}_{ii'} = \frac{\exp(x_i-b_j)}{\exp(x_{i'}-b_j)} = \exp\left(x_i-b_j-x_{i'}+b_j\right) = \exp(x_i-x_{i'}) \tag{77}$$

- Die Log-Odds-Ratio für eine richtige Antwort von Person i und Person i' in Item j ist nach dem Rasch-Modell

$$\text{Log-Odds-Ratio}_{ii'} = x_i - x_{i'} \tag{78}$$

  und hängt damit nur von den Eigenschaften von Personen i und i' aber nicht von dem gewählten Item ab.
- Die Odds-Ratio für eine richtige Antwort von Person i in Item j und Item j' ist nach dem Rasch-Modell

$$\text{Odds-Ratio}_{jj'} = \frac{\exp(x_i-b_j)}{\exp(x_i-b_{j'})} = \exp\left(x_i-b_j-x_i+b_{j'}\right) = \exp(b_{j'}-b_j) \tag{79}$$

- Die Log-Odds-Ratio für eine richtige Antwort von Person i in Item j und Item j' ist nach dem Rasch-Modell

$$\text{Log-Odds-Ratio}_{jj'} = b_{j'} - b_j \tag{80}$$

  und hängt damit nur von den Eigenschaften der Items j und j', aber nicht von der betrachteten Person ab.

Überblick zur Parameterschätzung

Conditional-Maximum-Likelihood-Schätzung (CML)

- Personeneigenschaften werden durch eine suffiziente Statistik aus der Likelihood "herauskonditioniert"
- Die Itemparameter werden dann durch Maximierung dieser bedingten Likelihood geschätzt
- Itemparameterschätzungen sind konsistent, ohne dass Annahmen über Eigenschaftsverteilungen nötig sind
- Ausschließliche Anwendung beim Rasch-Modell, auf 2PL oder 3PL nicht anwendbar (keine suffizienten Statistken)
- Klare Trennung zwischen "Itemkalibrierung" und "Personenmessung"

Suffiziente Statistik

- Eine Statistik heißt suffizient, wenn sie alle in Daten enthaltene Parameterinformation vollständig erfasst
- Die bedingte Verteilung der Daten gegeben eine suffiziente Statistik hängt nicht vom Parameter ab
- Der Parameter von Interesse sind hier die Personeneigenschaften $x_1, ..., x_n$
- Im Rasch-Modell ist der Summenscore (Anzahl gelöster Items) eine suffiziente Statistik für $x_1, ..., x_n$
- Man kann die Likelihood nur in Abhängigkeit der Itemparameter schreiben

Überblick zur Parameterschätzung

Wie im Rahmen der MML-Schätzung wird das Rasch-Modell auch im Rahmen der Conditional-Maximum-Likelihood-Schätzung als parameterisiertes Modell konzipiert, wobei die Personeneigenschaften $x_i, i = 1, \ldots, n$ allerdings als nicht zufällige Parameter einnehmen. In diesem Kontext werden sie z.T. auch als "nuisance parameter", also als Nebeneffekts- oder Störparameter bezeichnet. Seien formal also wieder

- y die Gesamtheit aller nm beobachteten Datenpunkte,
- x die Gesamtheit aller n Personeneigenschaften,
- b der Vektor aller Itemschwierigkeitsparameter.

Das Rasch-Modell entspricht dann der durch x und b parameterisierten Wahrscheinlichkeitsverteilung $p_{b,x}(y)$. Grundlegende Idee der CML-Schätzung ist es, durch Konditionierung auf einer suffizienten Statistik s für b die Personeneigenschaftsparameter aus der Likelihoodfunktion herauszurechnen und dann die Itemschwierigkeitsparameter ohne die Notwendigkeit einer Schätzung der Personeneigenschaften schätzen zu können. Betrachtet wird also die bedingte Likelihoodfunktion

$$L : b \to \mathbb{R}_{>0}, b \mapsto L(b) := p_b(y|s) \tag{81}$$

Statistisch abgesichert im Sinne sich ergebener konsistenter und asymptotisch normalverteilter Schätzer für die Itemparameter b ist dieses Verfahren durch Andersen (1970). Wir führen im Folgenden kurz den Begriff der suffizienten Statistik ein und betrachten dann das Rasch-Modell in diesem Lichte.

Theorem (Suffiziente Statistik)

Es sei $y := (y_1, ..., y_n)$ eine Stichprobe mit durch θ parameterisierter gemeinsamer Verteilung $p_\theta(y)$ und $s = s(y)$ sei eine Statistik. s wird *suffiziente Statistik* für den Parameter θ genannt, wenn eine der folgenden äquivalenten Bedingungen gilt:

(1) Die bedingte Verteilung von y gegeben s hängt nicht von θ ab, also

$$p_\theta(y|s) = p(y|s). \tag{82}$$

(2) Es existieren Funktionen $g_\theta(s)$ und $h(y)$, sodass

$$p_\theta(y) = g_\theta(s)h(y) \text{ für alle } y \text{ und alle } \theta. \tag{83}$$

Bemerkungen

- Es handelt sich um eine mathematisch sehr unpräzise Darstellung der Suffizienz.
- Eine Einführung zum Prinzip der Suffizienz findet sich z.B. in Casella & Berger (2002), Abschnitt 6.
- Würde man θ als Zufallsvariable betrachten, so wäre (1) äquivalent zu der bedingten Unabhängigkeit

$$p(y|s, \theta) = p(y|s). \tag{84}$$

- Aussage (2) wird nach (oder trotz) Halmos & Savage (1949) *Neyman-Fisher-Faktorisierung* genannt.

Beweis

Wir zeigen zunächst $(2) \Rightarrow (1)$. Es gelte also $p_\theta(y) = g_\theta(s)h(y)$ und wir betrachten die gemeinsame Verteilung von y und s. Setzt man die Kongruenz von s und y voraus, nimmt also an, dass

$$p_\theta(s|y) = 1 \text{ für } s = s(y) \text{ und } p_\theta(s|y) = 0 \text{ für } s \neq s(y), \tag{85}$$

so gilt für die gemeinsame Verteilung von y und s, dass

$$p_\theta(y, s) = p_\theta(s|y)p_\theta(y) = 1 \cdot p_\theta(y) = p_\theta(y) \text{ für alle } y \text{ und } s \text{ mit } s = s(y). \tag{86}$$

Betrachtet man dann die bedingte Verteilung von y gegeben s, so ergibt sich

$$\begin{aligned}
p_\theta(y|s) &= \frac{p_\theta(y, s)}{\int_{y'} p_\theta(y', s)} \\
&= \frac{p_\theta(y)}{\int_{y':s(y')=s} p_\theta(y', s)\, dy'} \\
&= \frac{p_\theta(y)}{\int_{y':s(y')=s} p_\theta(y')\, dy'} \\
&= \frac{g_\theta(s)h(y)}{\int_{y':s(y')=s} g_\theta(s)h(y')\, dy'} \\
&= \frac{g_\theta(s)h(y)}{g_\theta(s) \int_{y':s(y')=s} h(y')\, dy'} \\
&= \frac{h(y)}{\int_{y':s(y')=s} h(y')\, dy'}
\end{aligned} \tag{87}$$

und die rechte Seite hängt nicht mehr von θ ab und kann als $p(y|s)$ geschrieben werden.

Beweis (fortgeführt)

Wir zeigen nun $(1) \Rightarrow (2)$. s sei also suffizient, dass heißt,

$$p_\theta(y|s) = p(y|s) \tag{88}$$

Zunächst ergibt sich, wiederum unter Annahme der Kongruenz von s und y, dass

$$p_\theta(s|y) = 1 \text{ für } s = s(y) \text{ und } p_\theta(s|y) = 0 \text{ für } s \neq s(y) \tag{89}$$

und damit

$$p_\theta(y, s) = p_\theta(s|y)p_\theta(y) = 1 \cdot p_\theta(y) = p_\theta(y) \text{ für alle } y \text{ und } s \text{ mit } s = s(y). \tag{90}$$

Dann aber ergibt sich für $p_\theta(y)$, dass

$$p_\theta(y) = p_\theta(y, s) = p_\theta(s)p_\theta(y|s) = p_\theta(s)p(y|s) =: g_\theta(s)h(y). \tag{91}$$

Definition (Verteilungsform des Rasch-Modells)

Für $i = 1, \dots, n$ Personen und $j = 1, \dots, m$ dichotome Items seien

- $x_i \in \mathbb{R}$ die Personeneigenschaft von Person i
- $x := (x_1, \dots, x_n)$ die Eigenschaften aller n Personen,
- $b_j \in \mathbb{R}$ der Schwierigkeitsparameter des $j$ten Items,
- $b := (b_1, \dots, b_m)$ der Itemparametervektor,
- y_{ij} die Zufallsvariable zur Modellierung der j-ten Itemantwort von Person i mit Wertebereich $\{0, 1\}$,
- $y_i := (y_{i1}, \dots, y_{im})$ der Zufallsvektor der Itemantworten von Person i,
- $y := (y_1, \dots, y_n)$ die Zufallsmatrix der Itemantworten aller Personen r.

Dann hat die das Rasch-Modell die Verteilungsform

$$p_{x,b}(y) := \prod_{i=1}^{n} p_{x_i,b}(y_i) := \prod_{i=1}^{n} \prod_{j=1}^{m} p_{x_i,b_j}(y_{ij}), \text{ mit } p_{x_i,b_j}\left(y_{ij}\right) := \frac{\exp\left(y_{ij}(x_i - b_j)\right)}{1 + \exp(x_i - b_j)}. \tag{92}$$

Bemerkungen

- $\prod_{i=1}^{n} p_{x_i,b}(y_i)$ enkodiert die Unabhängigkeitsannahme über Personen.
- $\prod_{j=1}^{m} p_{x_i,b_j}(y_{ij})$ enkodiert die bedingte Unabhängigkeit der Itemantworten von Person i gegeben x_i.

Theorem (Summenscores-bedingte Verteilung des Rasch-Modells)

Gegeben sei die Verteilungsform des Rasch-Modells und für $i = 1, ..., n$ sei

$$s_i := \sum_{j=1}^{m} y_{ij} \tag{93}$$

der Summenscore von Person i. Sei weiterhin für alle $s_i \in \{0, ..., m\}$

$$Y_{s_i} := \left\{ y_i \in \{0,1\}^m \middle| \sum_{j=1}^{m} y_{ij} = s_i \right\} \tag{94}$$

die Menge aller zu einem Summenscore s_i kongruenten Itemantwortvektoren einer Person. Sei schließlich

$$s := (s_1, ..., s_n) \tag{95}$$

der Vektor der Summenscores aller n Personen. Dann gilt für die Verteilungsform des Rasch-Modells

$$p_{x,b}(y|s) = p_b(y|s) \tag{96}$$

und speziell

$$p_b(y|s) = \prod_{i=1}^{n} p_b(y_i|s_i) \text{ mit } p_b(y_i|s_i) = \frac{\exp\left(-\sum_{j=1}^{m} y_{ij} b_j\right)}{\sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^{m} y_{ij} b_j\right)}. \tag{97}$$

Bemerkung

- Die Aussage des Theorems wird auch als “Parameterseparierbarkeit” bezeichnet.

Beweis

Wir zeigen zunächst

$$p_{x_i,b}(y_i|s_i) = p_b(y_i|s_i) \text{ mit } p_b(y_i|s_i) = \frac{\exp\left(-\sum_{j=1}^{m} y_{ij} b_j\right)}{\sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^{m} y_{ij} b_j\right)}. \tag{98}$$

Nach Definition der bedingten Wahrscheinlichkeit gilt

$$p_{x_i,b}(y_i|s_i) = \frac{p_{x_i,b}(y_i, s_i)}{\sum_{y_i} p_{x_i,b}(y_i, s_i)} = \frac{p_{x_i,b}(y_i) p_{x_i,b}(s_i|y_i)}{\sum_{y_i} p_{x_i,b}(y_i) p_{x_i,b}(s_i|y_i)} \tag{99}$$

Da definitionsgemäß gilt, dass

$$s_i := \sum_{j=1}^{m} y_{ij}, \tag{100}$$

dass s_i also deterministisch von y_i abhängt, ist die bedingte Verteilung des Summenscores s_i gegeben die Itemantworten y_{ij} für $j = 1, \ldots, m$ der $i$ten Person degeneriert, d.h. sie nimmt nur die Werte 0 oder 1 an. Speziell gilt

$$p_{x_i,b}(s_i|y_i) = \left[\!\left[s_i = \sum_{j=1}^{m} y_{ij} \right]\!\right] = \begin{cases} 1 & \text{für } s_i = \sum_{j=1}^{m} y_{ij} \\ 0 & \text{für } s_i \neq \sum_{j=1}^{m} y_{ij} \end{cases} \tag{101}$$

Beweis (fortgeführt)

Dabei stellen wir zum einen fest, dass die bedingte Verteilung des Summenscores gegeben die Itemantworten der $i$ten Person weder von der Personeneigenschaft x_i noch vom Itemparametervektor b abhängen, also dass gilt

$$p_{x_i,b}(s_i|y_i) = p(s_i|y_i). \tag{102}$$

Zum anderen stellen wir fest, dass (99) nur im Falle der Kongruenz von s_i und y_i von null verschieden ist. Wir schränken unsere Betrachtungen deshalb im Folgenden auf den kongruenten Fall $s_i = \sum_{j=1}^{m} y_{ij}$ und damit $p(s_i|y_i) = 1$ ein. Es ergibt sich mit dieser Einschränkung also

$$p_{x_i,b}(y_i|s_i) = \frac{p_{x_i,b}(y_i)p(s_i|y_i)}{\sum_{y_i} p_{x_i,b}(y_i)p(s_i|y_i)} = \frac{p_{x_i,b}(y_i)}{\sum_{y_i \in Y_{s_i}} p_{x_i,b}(y_i)} \tag{103}$$

Wir betrachten nun zunächst den Zähler von (103). Es gilt

Beweis (fortgeführt)

$$
\begin{aligned}
p_{x_i,b}(y_i) &= \prod_{j=1}^{m} p_{x_i,b_j}(y_{ij}) \\
&= \prod_{j=1}^{m} \frac{\exp\left(y_{ij}(x_i - b_j)\right)}{1 + \exp(x_i - b_j)} \\
&= \frac{\prod_{j=1}^{m} \exp\left(y_{ij}(x_i - b_j)\right)}{\prod_{j=1}^{m} \left(1 + \exp(x_i - b_j)\right)} \\
&= \frac{\exp\left(\sum_{j=1}^{m} y_{ij}(x_i - b_j)\right)}{\prod_{j=1}^{m} \left(1 + \exp(x_i - b_j)\right)} \\
&= \frac{\exp\left(\sum_{j=1}^{m} y_{ij}x_i - \sum_{j=1}^{m} y_{ij}b_j\right)}{\prod_{j=1}^{m} \left(1 + \exp(x_i - b_j)\right)} \\
&= \frac{\exp\left(x_i \sum_{j=1}^{m} y_{ij} - \sum_{j=1}^{m} y_{ij}b_j\right)}{\prod_{j=1}^{m} \left(1 + \exp(x_i - b_j)\right)} \\
&= \frac{\exp\left(x_i s_i - \sum_{j=1}^{m} y_{ij}b_j\right)}{\prod_{j=1}^{m} \left(1 + \exp(x_i - b_j)\right)} \\
&= \frac{\exp\left(x_i s_i\right) \exp\left(-\sum_{j=1}^{m} y_{ij}b_j\right)}{\prod_{j=1}^{m} \left(1 + \exp(x_i - b_j)\right)}
\end{aligned}
\tag{104}
$$

Beweis (fortgeführt)

Für den Nenner von (103) gilt damit

$$\begin{aligned}\sum_{y_i \in Y_{s_i}} p_{x_i,b}(y_i) &= \sum_{y_i \in Y_{s_i}} \frac{\exp(x_i s_i) \exp\left(-\sum_{j=1}^m y_{ij} b_j\right)}{\prod_{j=1}^m \left(1 + \exp(x_i - b_j)\right)} \\ &= \left(\prod_{j=1}^m \left(1 + \exp(x_i - b_j)\right)\right)^{-1} \exp(x_i s_i) \sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^m y_{ij} b_j\right) \qquad (105) \\ &= \frac{\exp(x_i s_i) \sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^m y_{ij} b_j\right)}{\prod_{j=1}^m \left(1 + \exp(x_i - b_j)\right)}.\end{aligned}$$

Es ergibt sich also

$$\begin{aligned}p_{x_i,b}(y_i|s_i) &= \frac{\exp(x_i s_i) \exp\left(-\sum_{j=1}^m y_{ij} b_j\right)}{\prod_{j=1}^m \left(1 + \exp(x_i - b_j)\right)} \cdot \frac{\prod_{j=1}^m \left(1 + \exp(x_i - b_j)\right)}{\exp(x_i s_i) \sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^m y_{ij} b_j\right)} \\ &= \frac{\exp\left(-\sum_{j=1}^m y_{ij} b_j\right)}{\sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^m y_{ij} b_j\right)}.\end{aligned} \qquad (106)$$

Damit hängt $p_{x_i,b}(y_i|s_i)$ aber nicht mehr von x_i ab und es gilt

$$p_{x_i,b}(y_i|s_i) = p_b(y_i|s_i). \qquad (107)$$

Beweis (fortgeführt)

Wir betrachten nun die gemeinsame bedingte Verteilung $p_{x,b}(y|s)$. Dazu halten wir zunächst mit der Definition der bedingten Wahrscheinlichkeit fest, dass

$$p_{x,b}(y|s) = \frac{p_{x,b}(y,s)}{p_{x,b}(s)} = \frac{p_{x,b}(y,s)}{\sum_y p_{x,b}(y,s)} = \frac{p_{x,b}(y)p_{x,b}(s|y)}{\sum_y p_{x,b}(y)p_{x,b}(s|y)}\,. \tag{108}$$

Wir betrachten nun die bedingte Verteilung des Summenscorevektors s gegeben die Itemantwortmatrix aller Personen y. Mit der Definition von $s_i := \sum_{i=1}^m y_{ij}$ für alle $i = 1, \ldots, n$ gilt zunächst, dass auch diese Verteilung unabhängig von x und b ist. Weiterhin ist auch diese Verteilung degeneriert, nimmt also nur die Werte 0 und 1 an. Speziell gilt

$$p(s|y) = \left[\!\!\left[s_i = \sum_{j=1}^m y_{ij} \text{ für alle } i = 1, \ldots, n \right]\!\!\right]. \tag{109}$$

Damit gilt dann aber auch

$$p(s|y) = \prod_{i=1}^n p(s_i|y_i) \text{ mit } p(s_i|y_i) = \left[\!\!\left[s_i = \sum_{j=1}^m y_{ij} \right]\!\!\right]. \tag{110}$$

Mit der Verteilungsform des Raschmodells, der Tatsache, dass die Summation über y äquivalent zur sequentiellen Summation über $y_1, \ldots y_n$ ist, unter Einschränkung auf den kongruenten Fall $s_i = \sum_{j=1}^m y_{ij}$ (und damit $p(s_i|y_i) = 1$) für alle $i = 1, \ldots, n$ und mit obigem Ergebnis zu $p_{x_i,b}(y_i|s_i)$ können wir (108) also schreiben als

Beweis (fortgeführt)

$$
\begin{aligned}
p_{x,b}(y|s) &= \frac{\prod_{i=1}^{n} p_{x_i,b}(y_i) \prod_{i=1}^{n} p(s_i|y_i)}{\sum_{y_1} \cdots \sum_{y_n} \prod_{i=1}^{n} p_{x_i,b}(y_i) \prod_{i=1}^{n} p(s_i|y_i)} \\
&= \frac{\prod_{i=1}^{n} p_{x_i,b}(y_i) p(s_i|y_i)}{\sum_{y_1} \cdots \sum_{y_n} \prod_{i=1}^{n} p_{x_i,b}(y_i) p(s_i|y_i)} \\
&= \frac{\prod_{i=1}^{n} p_{x_i,b}(y_i) p(s_i|y_i)}{\prod_{i=1}^{n} \sum_{y_i} p_{x_i,b}(y_i) p(s_i|y_i)} \\
&= \frac{\prod_{i=1}^{n} p_{x_i,b}(y_i)}{\prod_{i=1}^{n} \sum_{y_i \in Y_{s_i}} p_{x_i,b}(y_i)} \\
&= \prod_{i=1}^{n} \frac{p_{x_i,b}(y_i)}{\sum_{y_i \in Y_{s_i}} p_{x_i,b}(y_i)} \\
&= \prod_{i=1}^{n} p_{x_i,b}(y_i|s_i) \\
&= \prod_{i=1}^{n} p_b(y_i|s_i).
\end{aligned}
\tag{111}
$$

Beweis (fortgeführt)

Wir können $p_{x,b}(y|s)$ also schreiben als

$$p_{x,b}(y|s) = p_b(y|s) = \prod_{i=1}^{n} p_b(y_i|s_i) = \prod_{i=1}^{n} \frac{\exp\left(-\sum_{j=1}^{m} y_{ij} b_j\right)}{\sum_{y_i \in Y_{s_i}} \exp\left(-\sum_{j=1}^{m} y_{ij} b_j\right)} \tag{112}$$

und damit ist alles gezeigt. □

Definition (Rasch-Modell Conditional-Maximum-Likelihood-Schätzung)

Gegeben sei die Verteilungsform des Rasch-Modells und es sei $p_b(y|s)$ die Summenscores-bedingte Verteilung des Rasch-Modells. Dann heißt die Funktion

$$L : \mathbb{R}^m \to \mathbb{R}_{>0}, b \mapsto L(b) := p_b(y|s) \tag{113}$$

die *Conditional-Likelihood-Funktion* des Rasch-Modells. Ein

$$\hat{b} \in \arg\max_{b \in \mathbb{R}^m} L(b) \tag{114}$$

heißt *Conditional-Maximum-Likelihood-Schätzer* von b.

Bemerkungen

- Die Schätzung der Itemparameter b kommt ohne die Bestimmung der Personeneigenschaften $x_1, ..., x_n$ aus.
- Als konsistentes Schätzverfahren für b ist dieses Vorgehen durch Andersen (1970) begründet worden.
- Die Anwendung der CML-Schätzung für das Rasch-Modell wurde durch Andersen (1977) geprägt.

Bemerkungen (fortgeführt)

- Für $m > 2$ muss $\hat{b}$ durch numerische Maximierung von L bestimmt werden.
- Für $m > 15$ ist eine explizite Summation über Y_{s_i} nicht möglich, es gilt

$$|Y_{s_i}| = \binom{m}{r_i} \text{ mit z.B. } \binom{30}{15} = 155.117.520 \tag{115}$$

- Es gibt rekursive Ansätze, um diese explizite Summation zu umgehen (Gustafsson (1977), Gustafsson (1980)).
- Diese Ansätze heißen *elementary symmetric functions* (Fischer (1995), Baker & Harwell (1996)).
- Der Ansatz ist nicht symmetrisch, Eigenschaften können nicht ohne Itemparameter bestimmt werden.

Generell sind CML Ansätze möglich, wenn

- die Datenverteilung zur Exponentialfamilie gehört und
- suffiziente Statistiken für Nebeneffekt-Parameter wie die Personeneigenschaften existieren.
- Beispielmodelle sind das Partial-Credit-Modell oder das Cox-Proportional-Hazards-Modell der Survival-Analyse

Grundlagen

Birnbaum-Modell

Rasch-Modell

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition der Logistischen Funktion wieder.
2. Geben Sie das Theorem zu den Eigenschaften der Logistischen Funktion wieder.
3. Geben Sie die Definition des Birnbaum-Modells wieder.
4. Erläutern Sie die Bedeutung des Itemschwierigkeitsparameters des Birnbaum-Modells.
5. Erläutern Sie die Bedeutung des Itemdiskriminationsparameters des Birnbaum-Modells.
6. Erläutern Sie die Bedeutung des Itemrateparameters des Birnbaum-Modells.
7. Erläutern Sie den Begriff der Spezifischen Objektivität im Kontext des Rasch-Modells.
8. Geben Sie die Definition des Rasch-Modells wieder.
9. Geben Sie das Theorem zum Rasch-Modell und 1PL-Birnbaum-Modell wieder.
10. Erläutern Sie die invarianten Vergleichseigenschaften des Rasch-Modells auf Parameterebene.
11. Erläutern Sie die invarianten Vergleichseigenschaften des Rasch-Modells auf Odds-Ratio-Ebene.

Beispielklausurfrage

Welche Aussage zur Logistischen Funktion

$$f : \mathbb{R} \to]0,1[, x \mapsto f(x) := \frac{1}{1+\exp(-a(x-b))} \text{ mit } a, b \in \mathbb{R}$$

für $a > 0$ trifft zu?

a) $f(b) = 0.00$

b) $f(b) = 0.25$

c) $f(b) = 0.50$

d) $f(b) = 1.00$

Ergänzungen

Theorem (Jensensche Ungleichung)

ξ sei eine Zufallsvariable und $g : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion, d.h.

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2) \tag{116}$$

für alle $x_1, x_2 \in \mathbb{R}, \lambda \in [0,1]$. Dann gilt

$$\mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)). \tag{117}$$

Analog sei $g : \mathbb{R} \to \mathbb{R}$ eine konkave Funktion, d.h.

$$g(\lambda x_1 + (1-\lambda)x_2) \geq \lambda g(x_1) + (1-\lambda)g(x_2) \tag{118}$$

für alle $x_1, x_2 \in \mathbb{R}, \lambda \in [0,1]$. Dann gilt

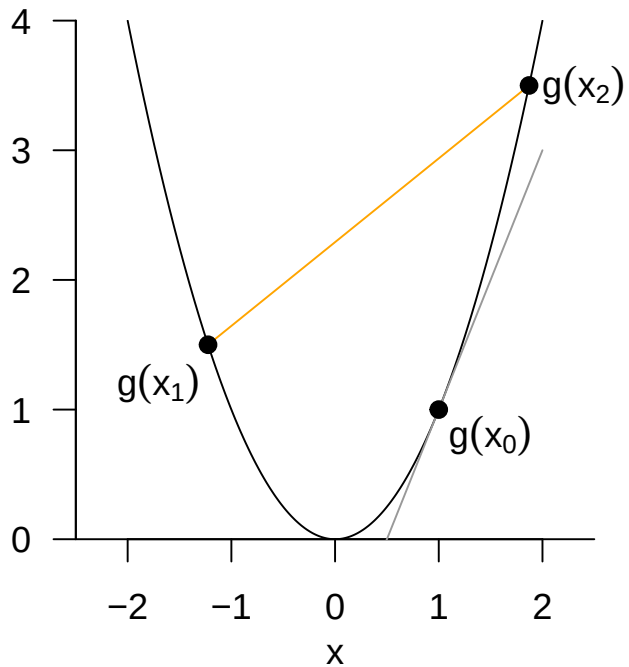
$$\mathbb{E}(g(\xi)) \leq g(\mathbb{E}(\xi)). \tag{119}$$

Bemerkungen

- Bei konvexem g liegt der Funktionsgraph unter der Geraden von $g(x_1)$ zu $g(x_2)$.
- Bei konkavem g liegt der Funktionsgraph über der Geraden von $g(x_1)$ zu $g(x_2)$.
- Der Logarithmus ist eine konkave Funktion, also gilt $\mathbb{E}(\ln \xi) \leq \ln \mathbb{E}(\xi)$.

Visualisierung einer konvexen Funktion

— $g(x) := x^2$



— $\lambda g(x_1) + (1 - \lambda) g(x_2)$ für $x_1 := -\sqrt{1.5}, x_2 := \sqrt{3.5}, \lambda \in [0, 1]$

— $t(x) := g(x_0) + g'(x_0)(x - x_0)$ für $x_0 := 1$

Beweis

Es sei g eine konvexe Funktion. Dann gilt für die Tangente t von g in $x_0 \in \mathbb{R}$ für alle $x \in \mathbb{R}$, dass

$$g(x) \geq t(x) := g(x_0) + g'(x_0)(x - x_0) \tag{120}$$

Wir setzen nun $x := \xi$ und $x_0 := \mathbb{E}(\xi)$. Dann gilt mit obiger Ungleichung, dass

$$g(\xi) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))(\xi - \mathbb{E}(\xi)) \tag{121}$$

Erwartungswertbildung ergibt dann

$$\begin{aligned} & \mathbb{E}(g(\xi)) \geq \mathbb{E}(g(\mathbb{E}(\xi))) + \mathbb{E}(g'(\mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))) \\ \Leftrightarrow\ & \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))\mathbb{E}((\xi - \mathbb{E}(\xi))) \\ \Leftrightarrow\ & \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))(\mathbb{E}(\xi) - \mathbb{E}(\xi)) \\ \Leftrightarrow\ & \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)). \end{aligned} \tag{122}$$

Sei nun g eine konkave Funktion. Dann ist $-g$ eine konvexe Funktion. Mit der Jensenschen Ungleichung für konvexe Funktionen folgt dann die Jensensche Ungleichung für konkave Funktionen aus

$$\begin{aligned} & \mathbb{E}(-g(\xi)) \geq -g(\mathbb{E}(\xi)) \\ \Leftrightarrow\ & -\mathbb{E}(g(\xi)) \geq -g(\mathbb{E}(\xi)) \\ \Leftrightarrow\ & \mathbb{E}(g(\xi)) \leq g(\mathbb{E}(\xi)). \end{aligned} \tag{123}$$

□ –>

Referenzen I

Andersen, E. B. (1970). Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, *32*(2), 283–301.

Andersen, E. B. (1977). Sufficient Statistics and Latent Trait Models. *Psychometrika*, *42*(1), 69–81. https://doi.org/10.1007/BF02293746

Andrich, D. (1978). Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers. *Applied Psychological Measurement*, *2*(4), 581–594. https://doi.org/10.1177/014662167800200413

Baker, F. B., & Harwell, M. R. (1996). Computing Elementary Symmetric Functions and Their Derivatives: A Didactic. *Applied Psychological Measurement*, *20*(2), 169–192. https://doi.org/10.1177/014662169602000206

Binet, A., & Simon, T. (1904). *Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux*. https://doi.org/10.3406/psy.1904.3675

Birnbaum, A. (1962). On the Foundations of Statistical Inference. *Journal of the American Statistical Association*, *57*(298), 269–306. https://doi.org/10.1080/01621459.1962.10480660

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689–8692. https://doi.org/10.1073/pnas.1702076114

Bock, R. D., & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, *46*(4), 443–459. https://doi.org/10.1007/BF02293801

Casella, G., & Berger, R. (2002). *Statistical Inference* (2nd ed.). Thomson Learning.

Doll, R., & Hill, A. B. (1950). Smoking and Carcinoma of the Lung. *BMJ*, *2*(4682), 739–748. https://doi.org/10.1136/bmj.2.4682.739

Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Psychology Press.

Referenzen II

Finney, D. J. (1952). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve.* (2nd ed.). Cambridge University Press.

Fischer, G. H. (1995). *Rasch Models: Foundations, Recent Developments, and Applications* (1st ed). Springer New York.

Gustafsson, J.-E. (1977). *The Rasch Model for Dichotomous Items: Theory, Applications and a Computer Program* (63). Gothenburg University, Institute of Education.

Gustafsson, J.-E. (1980). A Solution of the Conditional Estimation Problem for Long Tests in the Rasch Model for Dichotomous Items. *Educational and Psychological Measurement*, *40*(2), 377–385. https://doi.org/10.1177/001316448004000214

Haley, D. C. (1952). *Estimation of the dosage mortality relationship when the dose is subject to error.*

Halmos, P. R., & Savage, L. J. (1949). Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics. *The Annals of Mathematical Statistics*, *20*(2), 225–241. https://doi.org/10.1214/aoms/1177730032

Kaiser, F. G. (1998). A General Measure of Ecological Behavior[1]. *Journal of Applied Social Psychology*, *28*(5), 395–422. https://doi.org/10.1111/j.1559-1816.1998.tb01712.x

Kyngdon, A. (2008). The Rasch Model from the Perspective of the Representational Theory of Measurement. *Theory & Psychology*, *18*(1), 89–109. https://doi.org/10.1177/0959354307086924

Lawley, D. N. (1943). On Problems connected with Item Selection and Test Construction. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, *61*(3), 273–287. https://doi.org/10.1017/S0080454100006282

Linden, W. J. van der (Ed.). (2016). *Handbook of Item Response Theory*. CRC Press. https://doi.org/10.1201/9781315374512

Referenzen III

Lord, F. M. (1951). A theory of test scores. *ETS Research Bulletin Series*, *1951*(1). https://doi.org/10.1002/j.2333-8504.1951.tb00922.x

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores* (Nachdr. der Ausg. Reading, Mass. [u.a.], 1968). Information Age Publ.

Luce, D., & Tukey, J. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*(1), 1–27. https://doi.org/10.1016/0022-2496(64)90015-X

Mosier, C. I. (1940). Psychophysics and mental test theory: Fundamental postulates and elementary theorems. *Psychological Review*, *47*(4), 355–366. https://doi.org/10.1037/h0059934

Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. The MIT Press.

Ostwald, D., Kirilina, E., Starke, L., & Blankenburg, F. (2014). A tutorial on variational Bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology*, *60*, 1–19. https://doi.org/10.1016/j.jmp.2014.04.003

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. The MIT Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Expanded ed). University of Chicago Press.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, *19*(1), 49–57. https://doi.org/10.1111/j.2044-8317.1966.tb00354.x

Richardson, M. W. (1936). The Relation between the Difficulty and the Differential Validity of a Test. *Psychometrika*, *1*(2), 33–49. https://doi.org/10.1007/BF02288003

Rigdon, S. E., & Tsutakawa, R. K. (1983). Parameter Estimation in Latent Trait Models. *Psychometrika*, *48*(4), 567–574. https://doi.org/10.1007/BF02293880

Referenzen IV

Starke, L., & Ostwald, D. (2017). Variational Bayesian Parameter Estimation Techniques for the General Linear Model. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00504

Thissen, D. (1982). Marginal Maximum Likelihood Estimation for the One-Parameter Logistic Model. *Psychometrika*, *47*(2), 175–186. https://doi.org/10.1007/BF02296273

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1994). *Statistical analysis of finite mixture distributions* (3ème reprint). J. Wiley & sons.

Verhulst, P. F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, *18*, 14–54.

Woodruff, D. J. (1996). Estimation of Item Response Models Using the EM Algorithm for Finite Mixtures. *ACT Research Report Series*.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97–116. https://doi.org/10.1111/j.1745-3984.1977.tb00031.x

Wright, B., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychlogical Measurement*, *29*, 23–48.

Wu, M., Davis, R. L., Domingue, B. W., Piech, C., & Goodman, N. (2020). *Variational Item Response Theory: Fast, Accurate, and Expressive* (arXiv:2002.00276). arXiv. https://doi.org/10.48550/arXiv.2002.00276