IRT 1: Item Response Modelle für Dichotome Items

We are happy to share our materials openly:
The content of these Open Educational Resources by Lehrstuhl für Psychologische Methodenlehre und Diagnostik, Ludwig-Maximilians-Universität München is licensed under CC BY-SA 4.0. The CC Attribution-ShareAlike 4.0 International license means that you can reuse or transform the content of our materials for any purpose as long as you cite our original materials and share your derivatives under the same license.

Wiederholung: Testtheorie

- Psychologischer Test:
  - Wissenschaftliches Instrument zur Erfassung psychologischer Phänomene
  - Getestete Personen müssen meist mehrere Fragen beantworten oder mehrere Aufgaben lösen. Diese einzelnen Fragen/Aufgaben nennt man „Items“.

- Zwei Ziele bei der Entwicklung/Verwendung psychologischer Tests:
  - Sammeln nützlicher Informationen zur Vorhersage interessierender Kriterien (z.B. Berufserfolg, Planung eines Suizidversuchs, ...)
    -> Prädiktive Modellierung und Machine Learning, Wintersemester
  - Messung psychologischer Merkmale:
    Wie kann die Bearbeitung eines psychologischen Tests durch eine Person mathematisch so beschrieben werden, dass der Ausprägung der Person auf dem psychologischen Konstrukt eine Zahl zugeordnet werden kann?
    -> Strukturgleichungsmodelle und Item Response Theorie, Sommersemester

Wiederholung: Testtheoretisches Modell

- Manifeste vs. Latente Variablen:
  - Manifeste Variable: direkt beobachtbar (z.B. Körpergröße, Hautleitfähigkeit, Reaktionszeiten, Antworten auf Items)
  - Latente Variable: nicht direkt beobachtbar (z.B. Intelligenz, Aufmerksamkeit, Extraversion, Depressivität)

- Testtheoretisches Modell:
  - Statistisches Modell, mit dem anhand manifester Itemantworten auf latente Variablen geschlossen werden kann (z.B. τ - kongenerisches Modell)
  - Besteht aus Gleichungen, die manifeste Itemantworten mit latenten Variablen in Beziehung setzen (Erweiterungen von Regressionsmodellen auf statistische Probleme mit latenten Variablen)
    -> „psychometrisches“ Modell zur Messung psychologischer Merkmale

Klassische vs. Probabilistische Testtheorie

- Klassische Testtheorie (KTT; siehe Vorlesung im Bachelor und SEM-Teil aus FSM2)
  - “Klassisch“ in dem Sinne, dass diese Modelle zuerst entwickelt wurden und die psychologische Testtheorie für sehr lange Zeit geprägt haben
  - Testmodelle für Items mit kontinuierlichem Format (z.B. Reaktionszeiten)
  - Vollwertige Testtheorie, die in keinerlei Hinsicht minderwertig oder veraltet ist (aber: Items mit kontinuierlichem Antwortformat sind in der Praxis selten)
- Probabilistische Testtheorie (teilweise synonym: „Item Response Theorie“; IRT)
  - Testmodelle für Items mit kategorialem Antwortformat:
    - Dichotome Items (z.B. Intelligenzitem: „gelöst“ vs. „nicht gelöst“)
    - Ordinale Items (z.B. Persönlichkeitsitem: „trifft gar nicht zu“, “trifft nicht zu“, „weder noch“, „trifft zu“, „trifft voll und ganz zu“)
  - Die Bezeichnung „probabilistisch“ ist eher verwirrend und darf nicht so verstanden werden, dass die klassische Testtheorie Itemantworten als deterministisch betrachtet (beide Theorien berücksichtigen Messfehler). Wir bevorzugen eher den Begriff „Item Response Theorie“.

Überleitung SEM zu IRT

- Warum nochmal ein eigenes Kapitel zu Item Response Theorie? Testtheoretische Modelle sind doch auch als Strukturgleichungsmodell darstellbar (z.B. alle Modelle der KTT können wir als SEM schätzen).
- SEMs nehmen typischerweise an, dass alle endogenen manifesten Variablen multivariat normalverteilt sind. Dies ist für die meisten Items mit denen Konstrukte in der Psychologie gemessen werden unplausibel.
- Moderne SEM Software wie lavaan beinhaltet sehr viele Möglichkeiten, um auch Modelle mit realistischeren Annahmen für die Verteilung der Itemantworten zu schätzen. Zum Beispiel impliziert der WLSMV Schätzer ein bestimmtes IRT Modell für dichotome oder ordinale Itemantworten.
- Jedoch gibt es viele komplexere IRT Modelle, die nicht im SEM Kontext darstellbar sind. Es lohnt sich also, sich nochmal allgemein mit der Modellierung von Itemantworten zu beschäftigen. Damit geht es im IRT Kontext im Kern um eine möglichst realistische Modellierung psychologischer Daten, ohne dabei an ein bestimmtes statistisches Framework (z.B. lineare Regression, LMM, SEM, …) gebunden zu sein.

Ablauf des zweiten Teils der Vorlesung

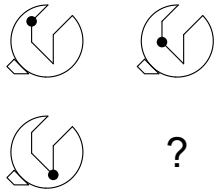
- Testmodelle für dichotome Items (Raschmodell, 2PL Modell)
- Testmodelle für ordinale Items (Partial Credit Modell, GPCM)
- Schätzung der Modellparameter, Modelltests und Modellvergleiche
- Konkrete Anwendungen im Beispiel
- (Adaptives Testen)

Eindimensionalität

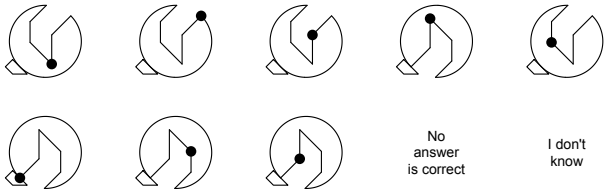
- Psychologische Testmodelle sollen das Zustandekommen der Antworten einer Reihe von Personen auf eine Reihe von Items statistisch beschreiben.
- Ein Testmodell bezeichnen wir als eindimensional, wenn es nur **eine einzige** latente Variable enthält, die die Ausprägung der Personen auf dem psychologischen Merkmal quantifiziert.
- Eindimensionale Testmodelle gehen also davon aus, dass die erwartete Antwort einer bestimmten Person auf ein bestimmtes Item ausschließlich von Parametern des Items (können mehrere sein) sowie dem Wert der Person auf der **einen** latenten Variable abhängt.
- Wir werden in dieser Vorlesung nur eindimensionale Testmodelle besprechen.
- Alle hier dargestellten Testmodelle lassen sich prinzipiell zu Modellen mit mehreren latenten Variablen erweitern (analog zum mehrdimensionalen τ – kongenerischen Modell für Items mit kontinuierlichem Antwortformat).

Dichotome Items in der Praxis

- Dichotome Items findet man meistens in Intelligenz– oder anderen Leistungstests
- Bsp.: Matrizenaufgabe in Intelligenztests



$x_{i,Person} = 1$ (richtige Antwort „c“)
$x_{i,Person} = 0$ (falsche Antwort „a, b, d, e“)



- Man spricht daher bei dichotomen Items häufig davon, dass eine Person ein Item „löst“ oder „nicht löst“, sowie von der zu messenden „Fähigkeit“ einer Person.
- Messmodelle für dichotome Items sind jedoch genauso anwendbar auf dichotome Fragebogenitems zur Messung von Persönlichkeitseigenschaften.
- Bsp.: Leistungsmotivationsitem im „Freiburger Persönlichkeitsinventar“ (FPI-R)

Ich ziehe das Handeln dem Päneschmieden vor.	stimmt nicht	stimmt
	☐	☐

$x_{i,Person} = 1$ („stimmt“ wurde angekreuzt)
$x_{i,Person} = 0$ („stimmt nicht“ wurde angekreuzt)

- Im Fragebogenkontext macht vor allem die Bezeichnung „ein Item zu lösen“ sprachlich wenig Sinn, wird jedoch oft trotzdem verwendet.

Wiederholung: τ – kongenerisches Modell

- Modellannahmen (auf der Ebene einer festen Person):

$$\tau_{i,Person} = E\left(X_{i,Person} | \xi_{Person}\right) = \lambda_{i,0} + \lambda_{i,1} \cdot \xi_{Person}$$

Der wahre Wert τ auf Item i einer konkreten (festen) Person → ein konkreter, konstanter Wert

Der Erwartungswert E() für die dahinterliegende Zufallsvariable X für das Item i einer Person

Der wahre Wert ξ (xi) der Person auf dem latenten Konstrukt

$$Cov\left(\varepsilon_{i,Person}, \varepsilon_{j,Person}\right) = 0$$

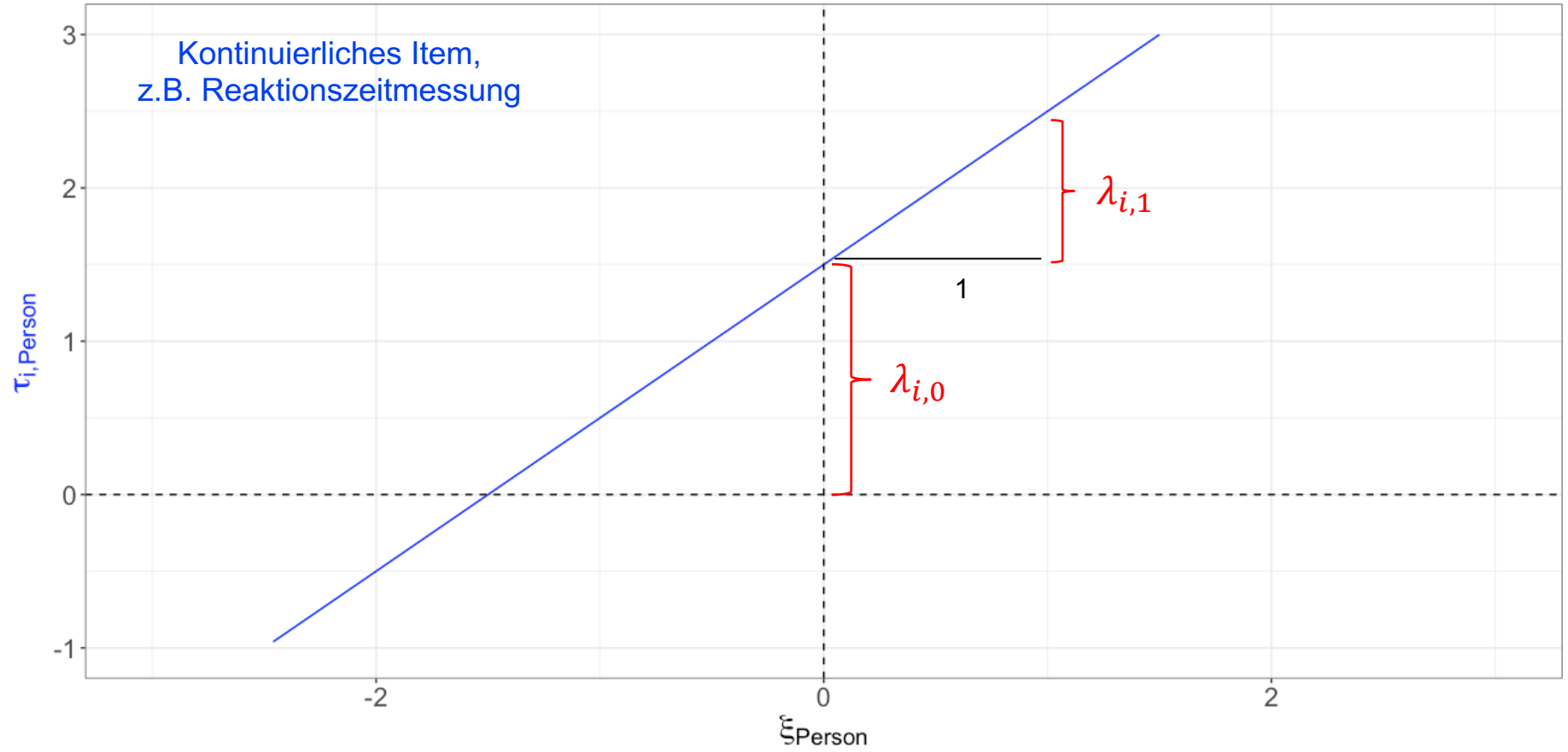
(die Fehlervarianzen der Items sind unkorreliert)

Wiederholung: τ – kongenerisches Modell

- Modellannahmen (auf der Ebene fester Person):

$$\tau_{i,Person} = E\left(X_{i,Person} | \xi_{Person}\right) = \lambda_{i,0} + \lambda_{i,1} \cdot \xi_{Person}$$

$$Cov\left(\varepsilon_{i,Person}, \varepsilon_{j,Person}\right) = 0$$

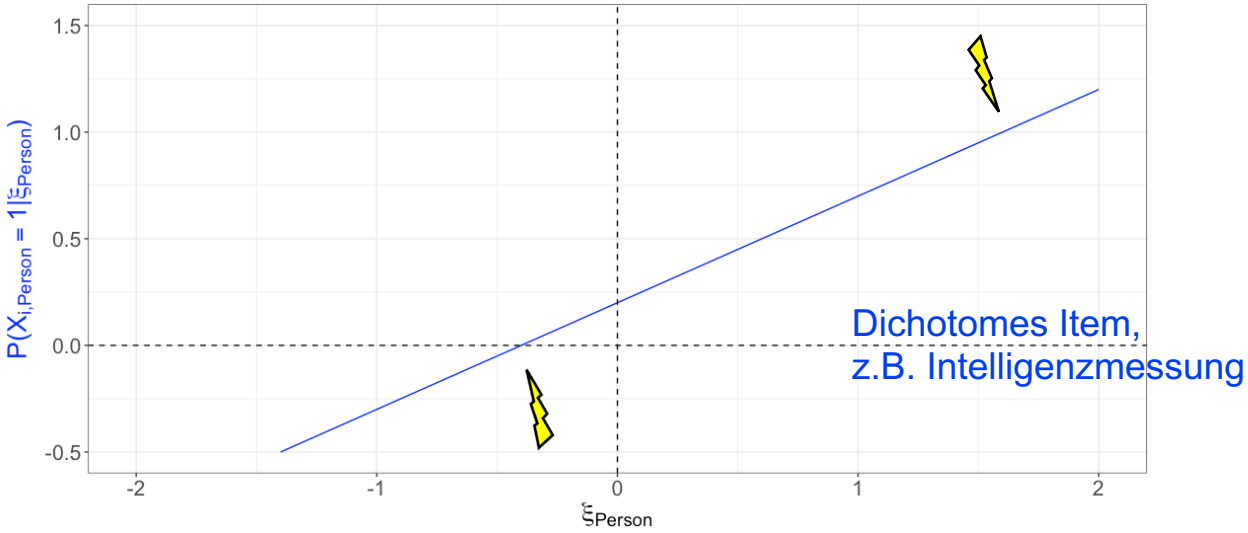


Herleitung: Testmodelle für dichotome Items

- Analog zur logistischen Regression modelliert man in Testmodellen für dichotome Items die Wahrscheinlichkeit, dass die Person ein Item i entweder löst ($x_{i,Person} = 1$) oder nicht löst ($x_{i,Person} = 0$), bei gegebenem Wert ξ_{Person} der Person auf der latenten Variable:

$$P\left(X_{i,Person} = 1 \middle| \xi_{Person}\right) = 1 - P\left(X_{i,Person} = 0 \middle| \xi_{Person}\right)$$

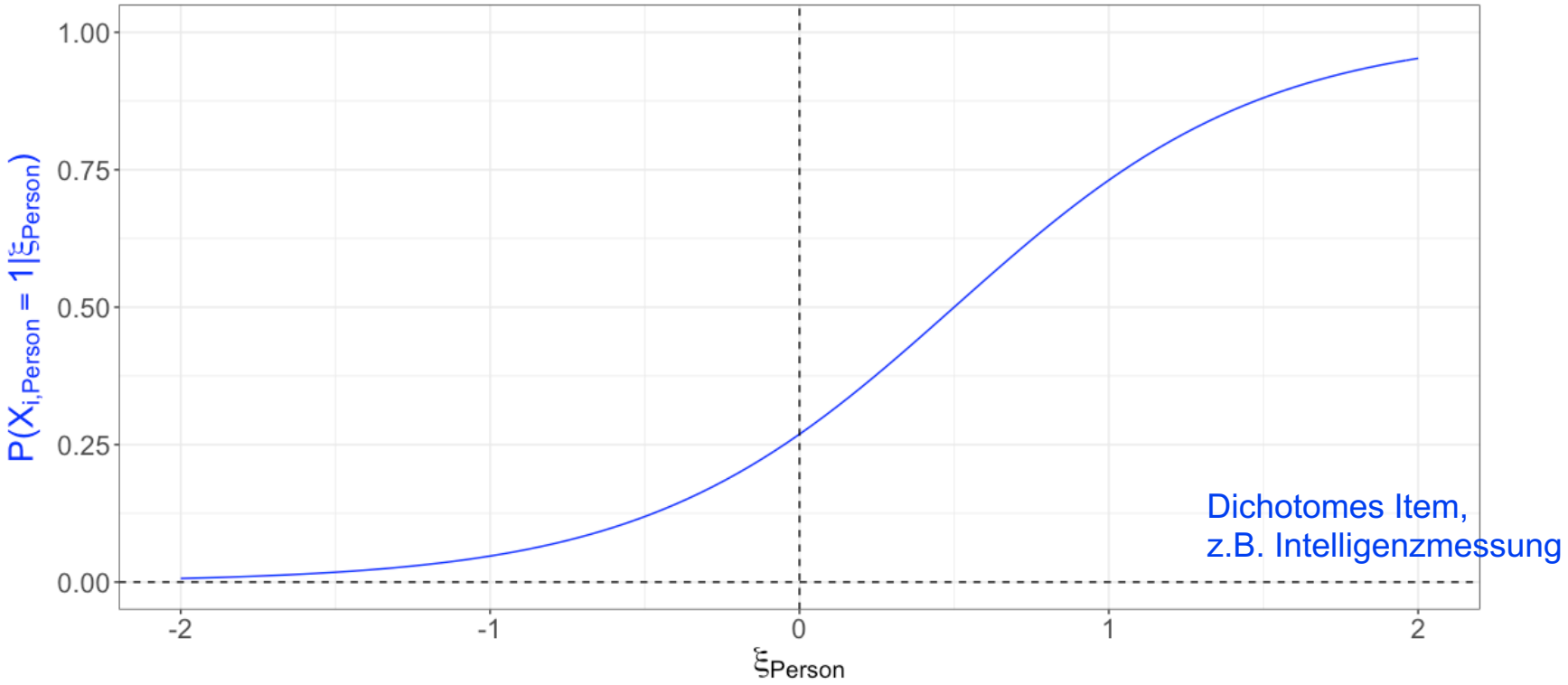
- Überträgt man das τ – kongenerische Modell direkt auf dichotome Items, ergibt sich wieder das Problem, dass eine Wahrscheinlichkeit zwischen 0 und 1, der Wertebereich des Testmodells jedoch zwischen $-\infty$ und $+\infty$ liegt.



Herleitung: Testmodelle für dichotome Items

- **Lösung wie in der logistischen Regression:**
  Transformiere den Wertebereich des τ – kongenerischen Modells von $]-\infty;+\infty[$ mithilfe der logistischen Funktion in Werte zwischen 0 und 1:

$$P(X_{i,Person}=1|\xi_{Person}) = E(X_{i,Person}|\xi_{Person}) = \frac{e^{\lambda_{i,0}+\lambda_{i,1}\cdot\xi_{Person}}}{1+e^{\lambda_{i,0}+\lambda_{i,1}\cdot\xi_{Person}}}$$



Herleitung: Dichotomes Raschmodell

- Für das einfachste Testmodell nehmen wir an, dass der Steigungsparameter $\lambda_{i,1}$ für alle Items den Wert 1 beträgt (vgl. τ – äquivalente Modelle):

$$P\left(X_{i,Person} = 1 \middle| \xi_{Person}\right) = \frac{e^{\lambda_{i,0}+\xi_{Person}}}{1 + e^{\lambda_{i,0}+\xi_{Person}}}$$

- Weiterhin stellen wir fest, dass der „Schwierigkeitsparameter“ $\lambda_{i,0}$ eigentlich ein „Leichtigkeitsparameter“ ist, da für eine Person mit Fähigkeit ξ_{Person} gilt:
  Je größer $\lambda_{i,0}$, desto höher ist die Wahrscheinlichkeit Item i zu lösen. Einen Schwierigkeitsparameter erhält man durch die Reparametrisierung: $\sigma_i = -\lambda_{i,0}$

$$P\left(X_{i,Person} = 1 \middle| \xi_{Person}\right) = \frac{e^{\xi_{Person}-\sigma_i}}{1 + e^{\xi_{Person}-\sigma_i}}$$

- Außerdem hat sich in der probabilistischen Testtheorie für die Ausprägung einer Person p auf der latenten Variable die Schreibweise θ_p etabliert. Damit erhält man die am häufigsten verwendete Form des berühmtesten Testmodells für dichotome Items, dem sogenannten dichotomen Raschmodell (Rasch, 1960).

Modellgleichung: Dichotomes Raschmodell

- Wahrscheinlichkeit im Raschmodell, dass eine feste Person p das Item i „löst“ ($x_{pi} = 1$) bzw. „nicht löst“ ($x_{pi} = 0$):

$$P(X_{pi} = 1|\theta_p) = \frac{e^{(\theta_p - \sigma_i)}}{1 + e^{(\theta_p - \sigma_i)}}$$

$$P(X_{pi} = 0|\theta_p) = 1 - P(X_{pi} = 1|\theta_p) = \underbrace{\frac{1+e^{(\theta_p-\sigma_i)}}{1+e^{(\theta_p-\sigma_i)}}}_{1} - \frac{e^{(\theta_p-\sigma_i)}}{1+e^{(\theta_p-\sigma_i)}} = \frac{1}{1+e^{(\theta_p-\sigma_i)}}$$

- Da $e^0 = 1$ und $x_{pi} \in \{0,1\}$, kann man beide Gleichungen zu einer kombinieren.
  -> Modellgleichung des dichotomen Raschmodells für feste Personen:

$$P(X_{pi} = x_{pi}|\theta_p) = \frac{e^{x_{pi}\cdot(\theta_p - \sigma_i)}}{1 + e^{(\theta_p - \sigma_i)}}$$

Durch das x_{pi} wird der Zähler 1 wenn man x_{pi}=0 einsetzt und $e^{(\theta_p-\sigma_i)}$ wenn man x_{pi}= 1 einsetzt.

- x_{pi}: beobachtete Itemantwort von Person p auf Item i ($x_{pi} \in \{0,1\}$)
- θ_p: Wert von Person p auf der latenten Variable ($\theta_p \in]-\infty; +\infty[\,, p \in \{1, \dots, P\}$)
- σ_i: Itemparameter von Item i ($\sigma_i \in]-\infty; +\infty[\,, i \in \{1, \dots, I\}$)

Beobachtete vs. Modellimplizierte Datenmatrix

- Im Gegensatz zur logistischen Regression besteht ein testtheoretisches Modell nicht nur aus einer Gleichung für jede Beobachtung, sondern aus einer Gleichung für jede Kombination aus P Personen und I Items (insgesamt $P \cdot I$ Gleichungen).
- Ein probabilistisches Testmodell wie das Raschmodell impliziert also eine gewisse Wahrscheinlichkeit für jeden Eintrag in der beobachteten Datenmatrix:

$$\frac{e^{(\theta_1-\sigma_2)}}{1+e^{(\theta_1-\sigma_2)}}$$

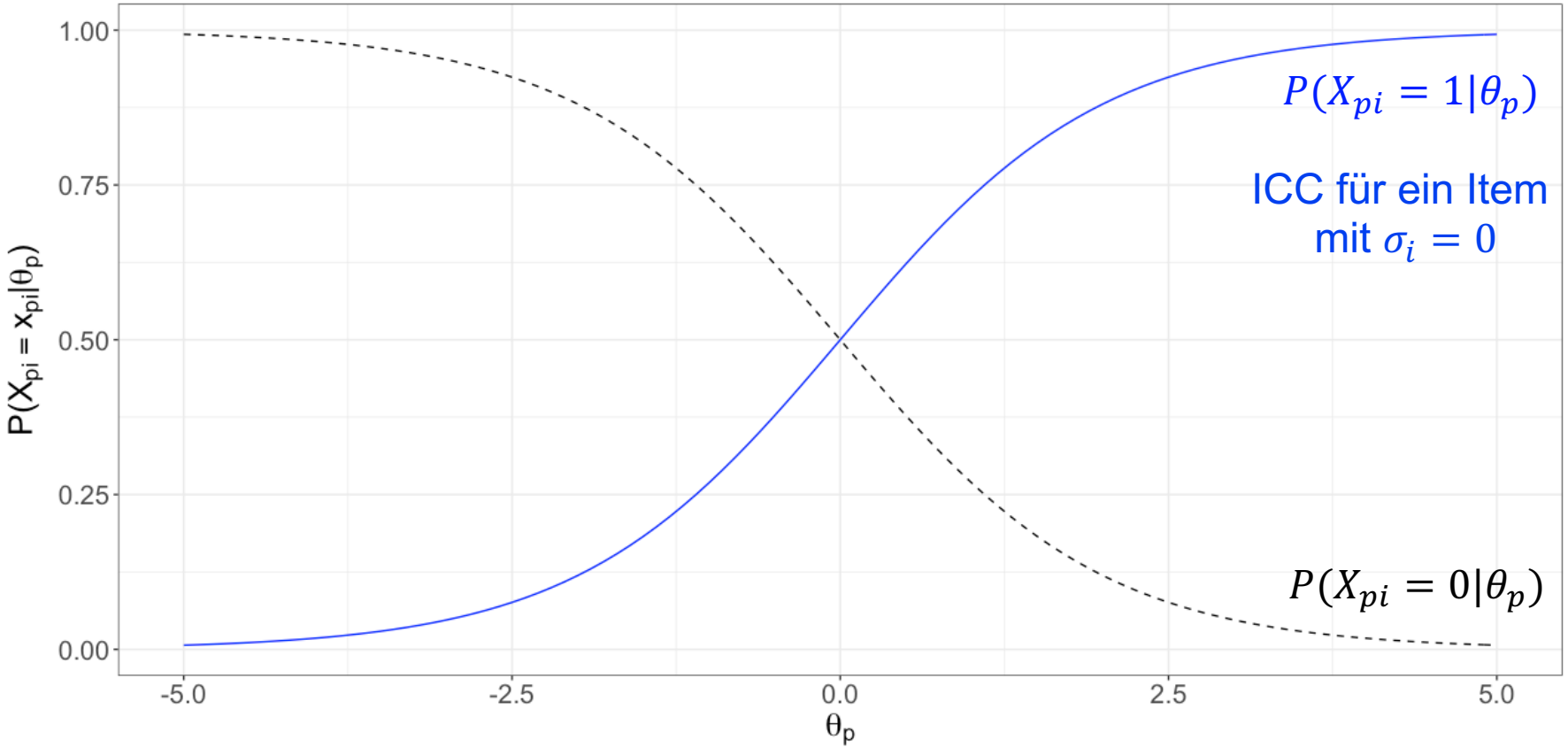
Beobachtete Datenmatrix

x_{pi}	1	2	3	…	I
1		1	0		
2					
…					
P					

Modellimplizierte Datenmatrix

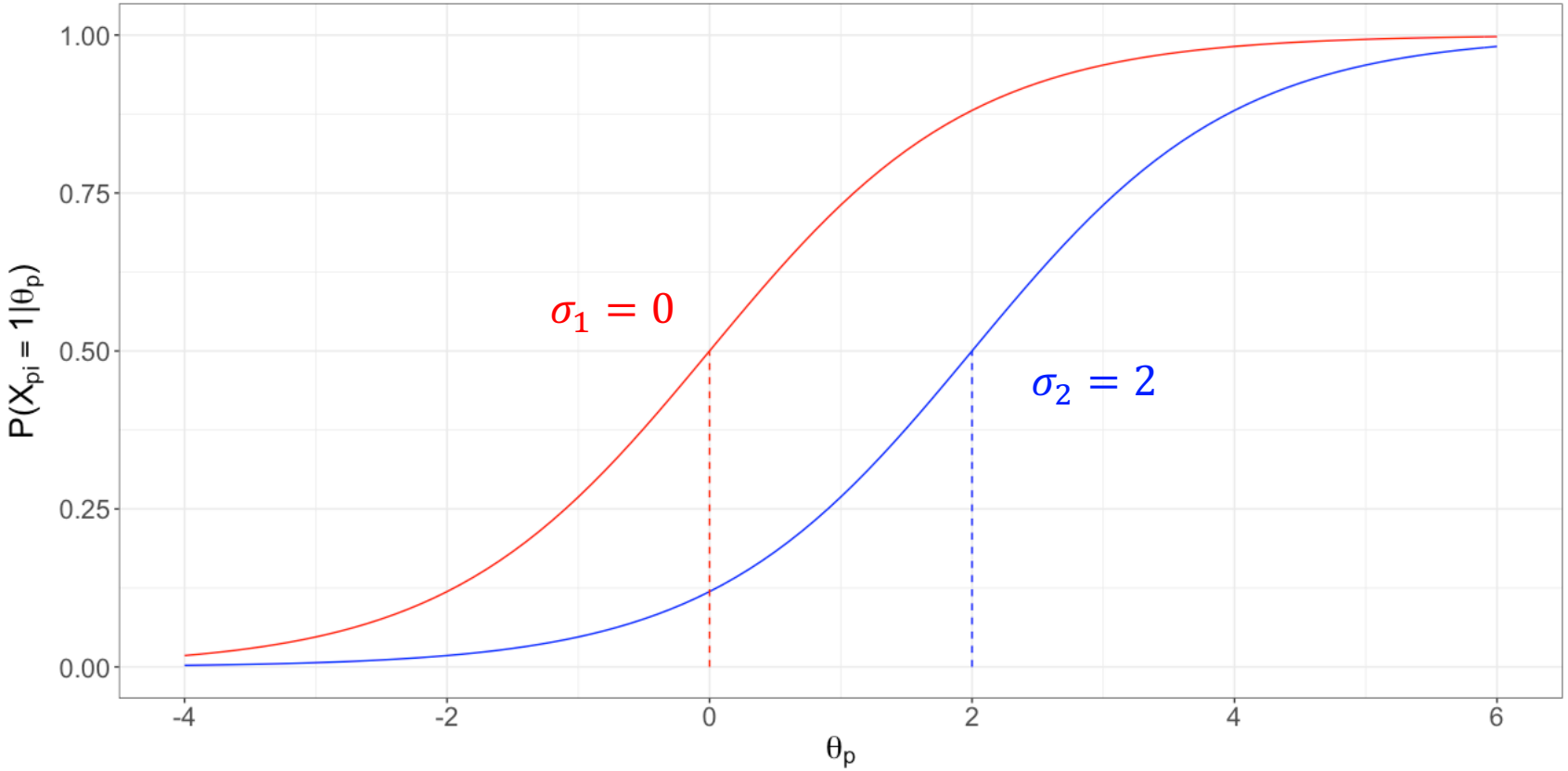
$P(X_{pi}=1\|\theta_p)$	1	2	3	…	I
1		$\frac{e^{(\theta_1-\sigma_2)}}{1+e^{(\theta_1-\sigma_2)}}$	$\frac{e^{(\theta_1-\sigma_3)}}{1+e^{(\theta_1-\sigma_3)}}$		
2					
…					
P					

Item Characteristic Curve



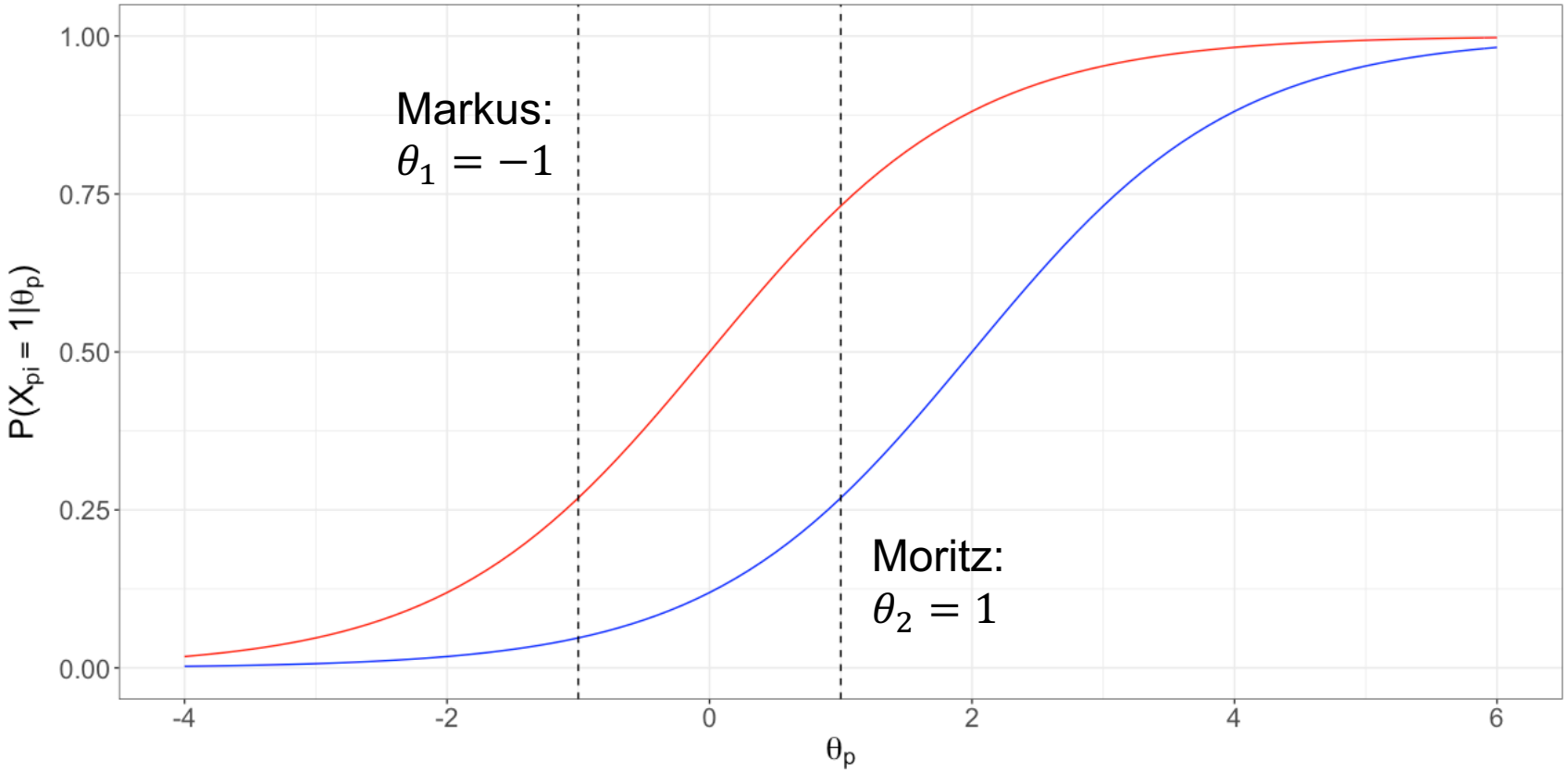
- Für jeden Wert θ_p gilt $P(X_{pi} = 1|\theta_p) + P(X_{pi} = 0|\theta_p) = 1$
- Man betrachtet daher in der Regel nur die Funktion $P(X_{pi} = 1|\theta_p)$ in θ_p und bezeichnet diese als „Item Characteristic Curve“ (ICC) von Item i.

Interpretation der Itemparameter



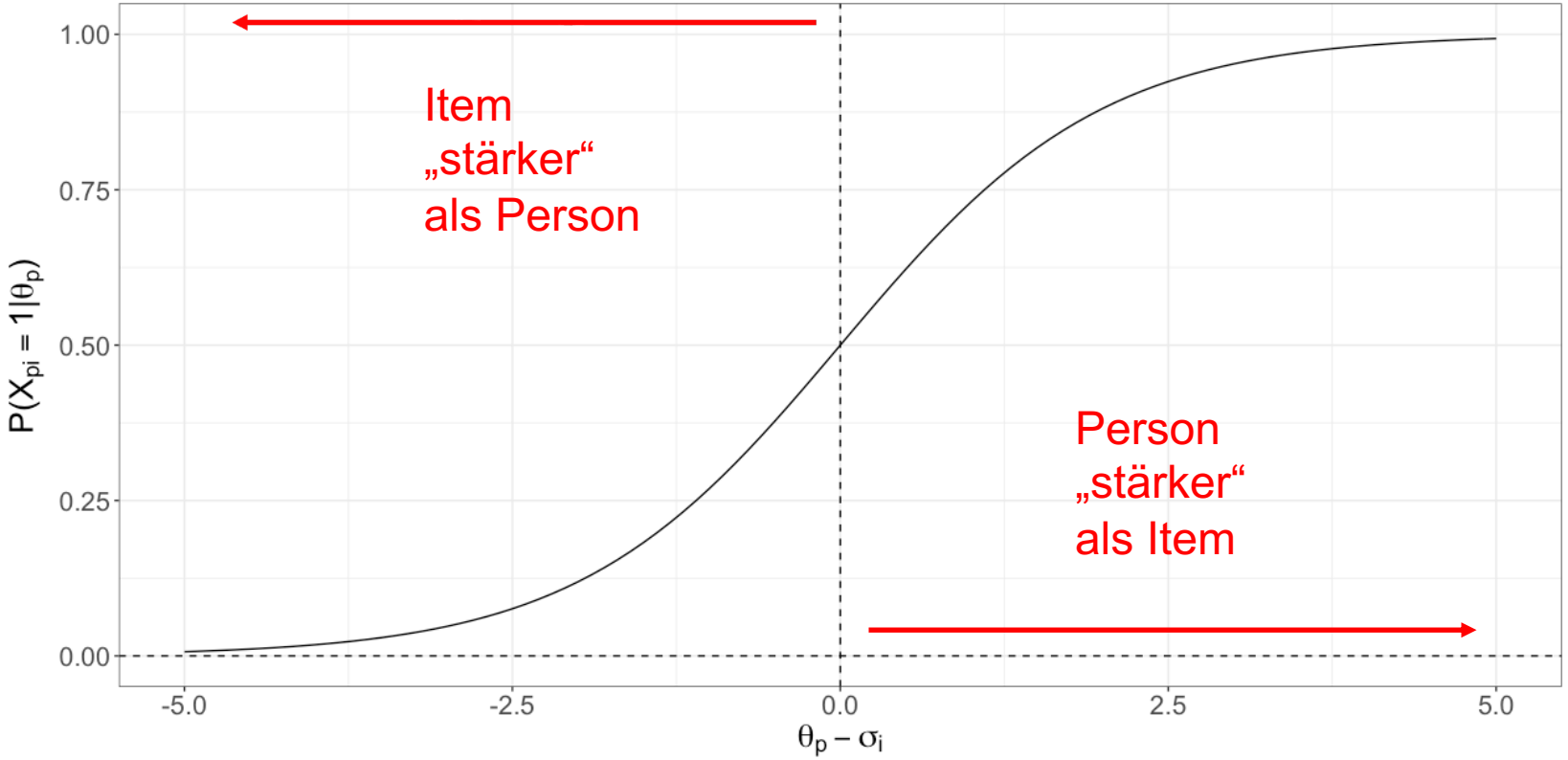
- Im dichotomen Raschmodell wird ein Item vollständig durch seinen Itemparameter σ_i beschrieben
- σ_i gibt an, für welchen Wert auf der latenten Variable die Wahrscheinlichkeit einer Person p das Item i zu lösen genau $0{,}5$ beträgt.
- Der Itemparameter σ_i wird daher als „Schwierigkeit“ von Item i interpretiert.
  -> je höher σ_i, desto schwieriger ist es das Item i zu lösen

Interpretation der Itemparameter



- Für jede Person ist es wahrscheinlicher das rote Item mit der niedrigeren Schwierigkeit zu lösen als das blaue Item: $P(X_{p1} = 1|\theta_p) > P(X_{p2} = 1|\theta_p)$
- Da Moritz eine höhere „Fähigkeit“ hat als Markus, hat Moritz für jedes Item eine höhere Lösungswahrscheinlichkeit: $P(X_{2i} = 1|\theta_2) > P(X_{1i} = 1|\theta_1)$ für alle i

Itemantwort als Wettkampf zwischen einer Person p und einem Item i



- Der Personenparameter θ_p (die Fähigkeit der Person) und der Itemparameter σ_i (die Schwierigkeit des Items) liegen auf der selben Skala
- Man kann sich die Itemantwort einer Person p auf einem Item i als „Wettkampf“ zwischen der Person und dem Item vorstellen.
- **Idee:** Wer ist „stärker“, die Person oder das Item?

Itemantwort als Wettkampf zwischen einer Person p und einem Item i

- Sind Personenfähigkeit θ_p und Itemschwierigkeit σ_i identisch, dann gilt:

  $$P\left(X_{pi} = 1 \middle| \theta_p\right) = \frac{e^{(\theta_p - \sigma_i)}}{1+e^{(\theta_p - \sigma_i)}} = \frac{e^{(0)}}{1+e^{(0)}} = \frac{1}{1+1} = 0{,}5$$

  -> „Genauso wahrscheinlich das Item zu lösen wie es nicht zu lösen“

- Ist die Personenfähigkeit größer als die Itemschwierigkeit, dann gilt:
  $P\left(X_{pi} = 1 \middle| \theta_p\right) > 0{,}5$ (Begründung: $\theta_p > \sigma_i \Leftrightarrow \theta_p - \sigma_i > 0$)
  -> „Höhere Wahrscheinlichkeit das Item zu lösen als es nicht zu lösen“

- Ist die Personenfähigkeit kleiner als die Itemschwierigkeit, dann gilt:
  $P\left(X_{pi} = 1 \middle| \theta_p\right) < 0{,}5$ (Begründung: $\theta_p < \sigma_i \Leftrightarrow \theta_p - \sigma_i < 0$)
  -> „Höhere Wahrscheinlichkeit das Item nicht zu lösen als es zu lösen“

Alternative Modellschreibweisen

- Analog zur logistischen Regression gibt es auch für die Modellgleichung des dichotomen Raschmodells zwei weitere äquivalente Schreibweisen.

  - Odds Schreibweise:

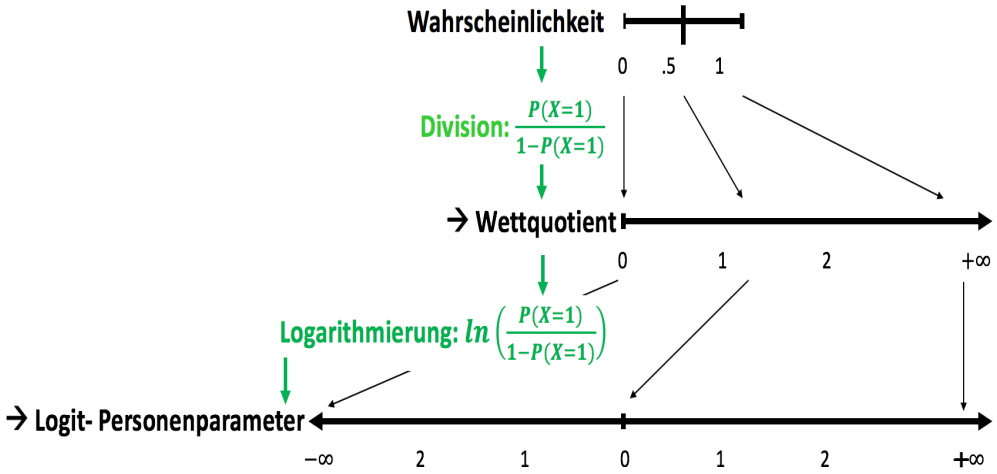
$$\frac{P(X_{pi}=1|\theta_p)}{P(X_{pi}=0|\theta_p)} = \frac{P(X_{pi}=1|\theta_p)}{1-P(X_{pi}=1|\theta_p)} = e^{\theta_p-\sigma_i}$$

  - Log – Odds bzw. Logit Schreibweise:

$$ln\left(\frac{P(X_{pi}=1|\theta_p)}{P(X_{pi}=0|\theta_p)}\right) = ln\left(\frac{P(X_{pi}=1|\theta_p)}{1-P(X_{pi}=1|\theta_p)}\right) = \theta_p - \sigma_i$$

Veranschaulichung der Logitskala

- Logits lassen sich direkt in Lösungswahrscheinlichkeiten umrechnen.
  Bsp: $\theta_p - \sigma_i = 1 \Rightarrow e^1 = 2{,}72 \Rightarrow$
  $\Rightarrow P(X_{pi} = 1|\theta_p) = \frac{2{,}72}{1 + 2{,}72} = 0{,}73$
- Logits liegen meist im Bereich zwischen – 3 und + 3 (außer für extreme Wahrscheinlichkeiten)



$P(X_{pi}=1\|\theta_p)$	$P(X_{pi}=0\|\theta_p)$	$\frac{P(X_{pi}=1\|\theta_p)}{P(X_{pi}=0\|\theta_p)}$	$ln\left(\frac{P(X_{pi}=1\|\theta_p)}{P(X_{pi}=0\|\theta_p)}\right)$
0,95	0,05	19,00	2,94
0,9	0,1	9,00	2,20
0,85	0,15	5,67	1,73
0,8	0,2	4,00	1,39
0,75	0,25	3,00	1,10
0,7	0,3	2,33	0,85
0,65	0,35	1,86	0,62
0,6	0,4	1,50	0,41
0,55	0,45	1,22	0,20
0,5	0,5	1,00	0,00
0,45	0,55	0,82	-0,20
0,4	0,6	0,67	-0,41
0,35	0,65	0,54	-0,62
0,3	0,7	0,43	-0,85
0,25	0,75	0,33	-1,10
0,2	0,8	0,25	-1,39
0,15	0,85	0,18	-1,73
0,1	0,9	0,11	-2,20
0,05	0,95	0,05	-2,94

Spezifische Objektivität: Exkurs

- Im Raschmodell ist die Differenz der Logits zweier Personen immer gleich, solange zur Messung beider Personen das gleiche Item verwendet wurde.

$$ln\left(\frac{P(X_{11}=1|\theta_1)}{P(X_{11}=0|\theta_1)}\right) - ln\left(\frac{P(X_{21}=1|\theta_2)}{P(X_{21}=0|\theta_2)}\right) = \theta_1 - \theta_2$$

-> „Ein Vergleich zweier Personen ist unabhängig vom Messinstrument“

- Gleichermaßen ist auch die Differenz der Logits zweier Items immer gleich, solange zur Beurteilung der beiden Items die gleiche Person verwendet wurde.

$$ln\left(\frac{P(X_{11}=1|\theta_1)}{P(X_{11}=0|\theta_1)}\right) - ln\left(\frac{P(X_{12}=1|\theta_1)}{P(X_{12}=0|\theta_1)}\right) = \sigma_2 - \sigma_1$$

-> „Ein Vergleich zweier Items ist unabhängig vom Messobjekt“

- Diese beiden Punkte bezeichnet man auch als „Spezifische Objektivität“

Spezifische Objektivität: Exkurs

- **„Ein Vergleich zweier Personen ist unabhängig vom Messinstrument“ – Beispiel (mit R-Code statt Formeln)**

Messung 1:

- $\theta_1 = 0$, $\theta_2 = 1$, $\sigma_1 = 0$
- $Prob_1$ = exp(0-0) / (1+ exp(0-0)) = 0.5; $Odds_1$ = 0.5/0.5 = 1;
- $Prob_2$ = exp(1-0) / (1+ exp(1-0)) = 0.73; $Odds_2$ = 0.73/(1-0.73) = 2.7
- Differenz der logits = log(1) – log(2.7) = **-1**

Messung 2:

- $\theta_1 = 0$, $\theta_2 = 1$, $\sigma_1 = 2$
- $Prob_1$ = exp(0-2) / (1+ exp(0-2)) = 0.12; $Odds_1$ = 0.12/(1-0.12) = 0.136
- $Prob_2$ = exp(1-2) / (1+ exp(1-2)) = 0.27; $Odds_2$ = 0.27/(1-0.27) = 0.37
- Differenz = log(0.136) – log(0.37) = **-1**

Fazit: Auch wenn sich die konkreten Lösungswahrscheinlichkeiten und die Odds zwischen den Messungen unterscheiden, ist die Differenz der Logits gleich

Spezifische Objektivität: Exkurs

- Ein Teil der psychometrischen Community misst der spezifischen Objektivität sehr hohe Bedeutung zu, da sie einem gewissen idealtypischen Verständnis physikalischer Messungen entspricht.
- Da komplexere Testmodelle mit mehreren Itemparametern die spezifische Objektivität nicht aufweisen, argumentiert dieser Teil der Community, dass nur von psychologischer Messung gesprochen werden sollte, wenn ein psychologischer Test den Annahmen des Raschmodells genügt.
- Ein anderer Teil vertritt hingegen die Ansicht, dass das Ziel testtheoretischer Modelle darin besteht, die Itemantworten eines psychologischen Tests möglichst präzise abzubilden. Da das Raschmodell sehr strenge Annahmen trifft, die in der empirischen Praxis meist verletzt sind, werden damit auch flexiblere Testmodelle betrachtet, für die keine spezifische Objektivität gilt.
- Wir werden uns in dieser Vorlesung der zweiten Meinung anschließen. Unserer Ansicht nach bietet die spezifische Objektivität keine entscheidenden Vorteile und schränkt die Wahl sinnvoller Testmodelle in der Praxis zu stark ein.

Mehr Flexibilität: Vom 1PL Modell zum 2PL Modell

- Im Raschmodell wird jedes Item durch genau einen Parameter charakterisiert (die Itemschwierigkeit σ_i). Man bezeichnet es daher auch als „one parameter logistic“ (1PL) Modell.
- Die Steigung β_i ist auf 1 fixiert (d.h., alle Items haben dieselbe Steigung).
- Man könnte also das Modell flexibler machen, indem Items unterschiedliche Steigungen haben dürfen. Durch Aufnahme eines zusätzlichen freien Parameters β_i für jedes Item können wir das Modell zum „2PL – Modell“ erweitern.
- Lösungswahrscheinlichkeit im 2PL Modell:

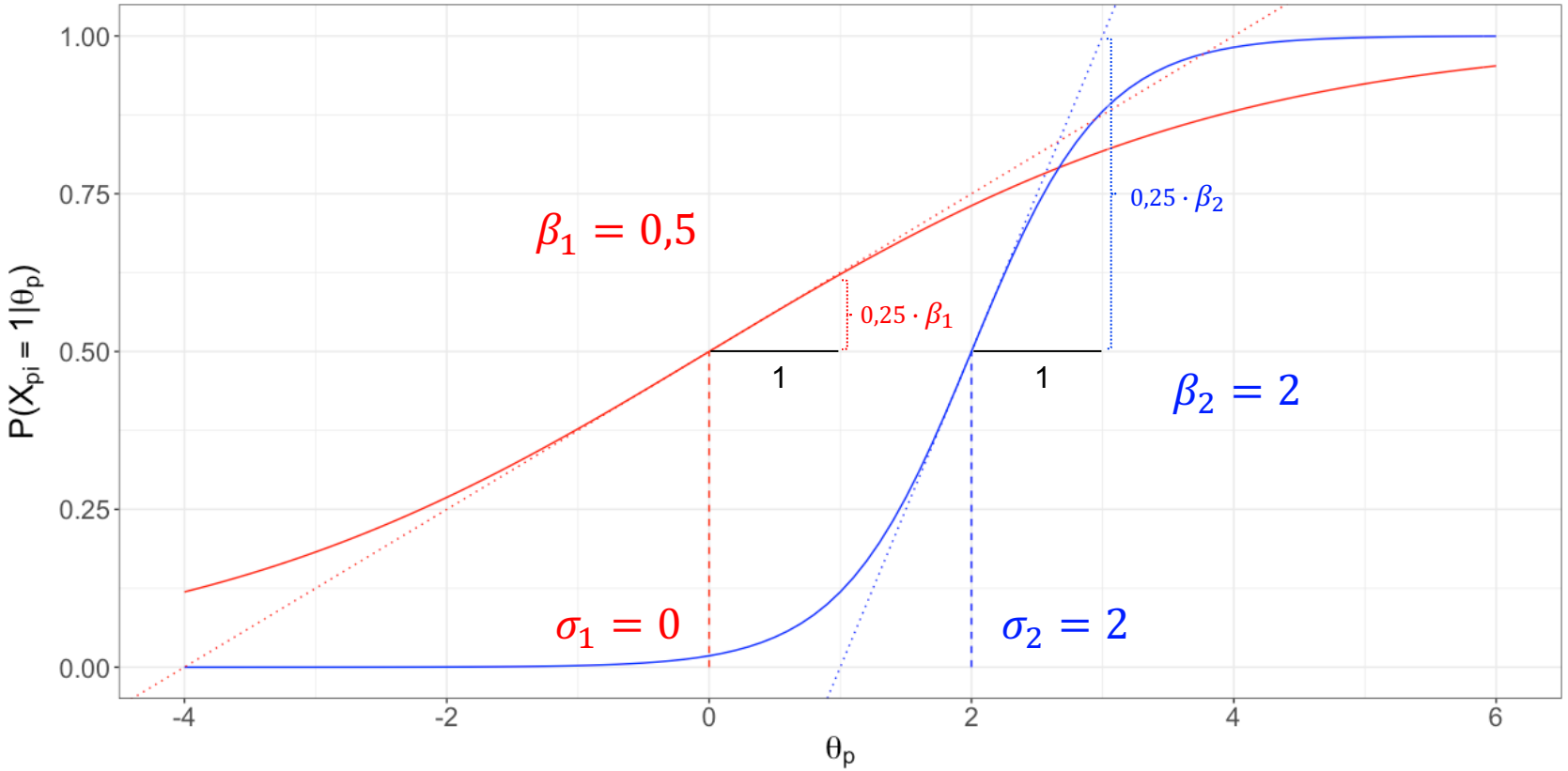
$$P(X_{pi} = 1|\theta_p) = \frac{e^{\boldsymbol{\beta_i}(\theta_p - \sigma_i)}}{1 + e^{\boldsymbol{\beta_i}(\theta_p - \sigma_i)}}$$

- Allgemeine Modellgleichung:

$$P(X_{pi} = x_{pi}|\theta_p) = \frac{e^{x_{pi} \cdot \beta_i(\theta_p - \sigma_i)}}{1 + e^{\beta_i(\theta_p - \sigma_i)}}$$

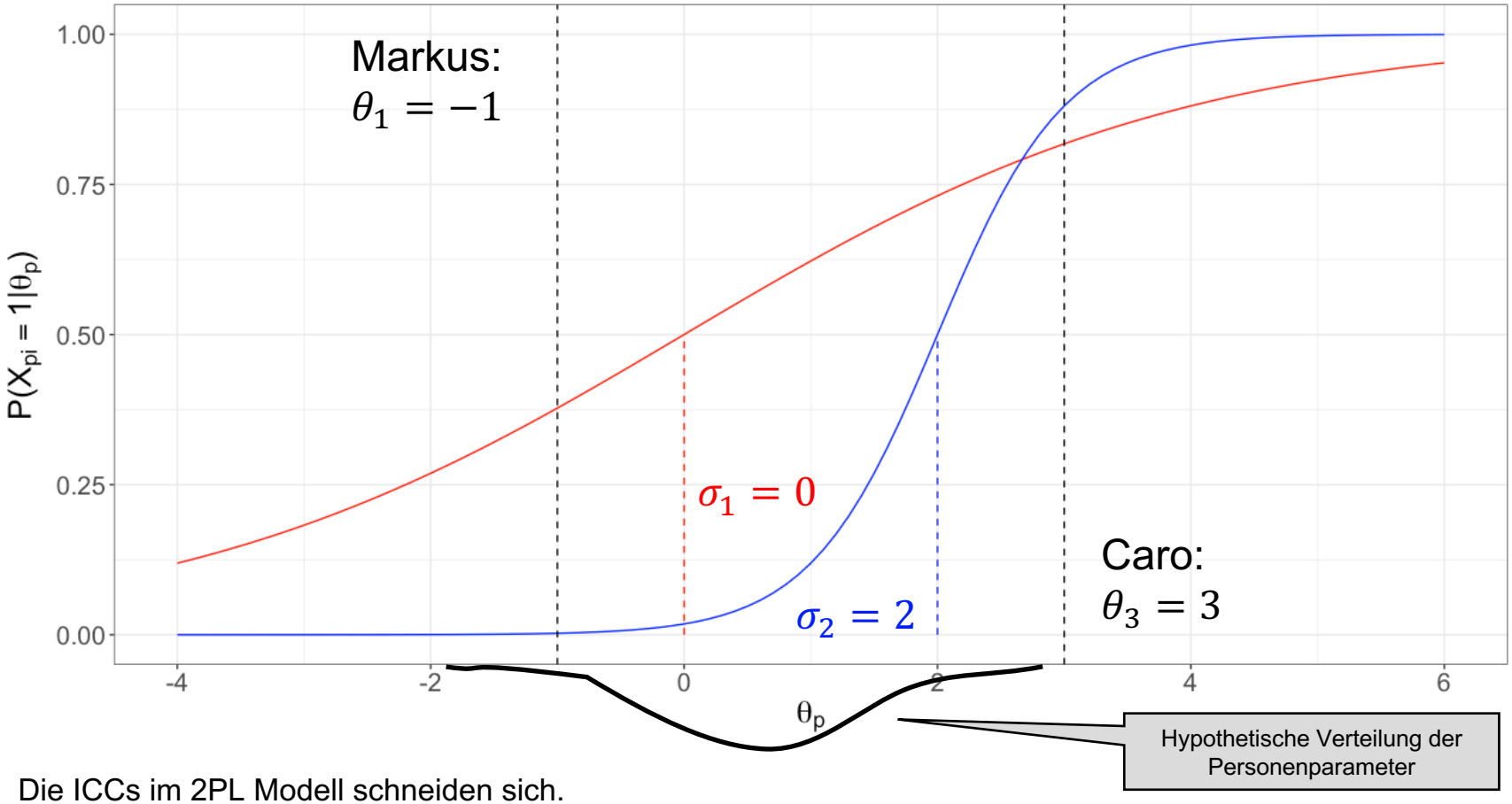
- Im Gegensatz zu σ_i und θ_p kann der Parameter β_i nur positive Werte annehmen ($\beta_i \in]0; +\infty[$)

Interpretation der Modellparameter



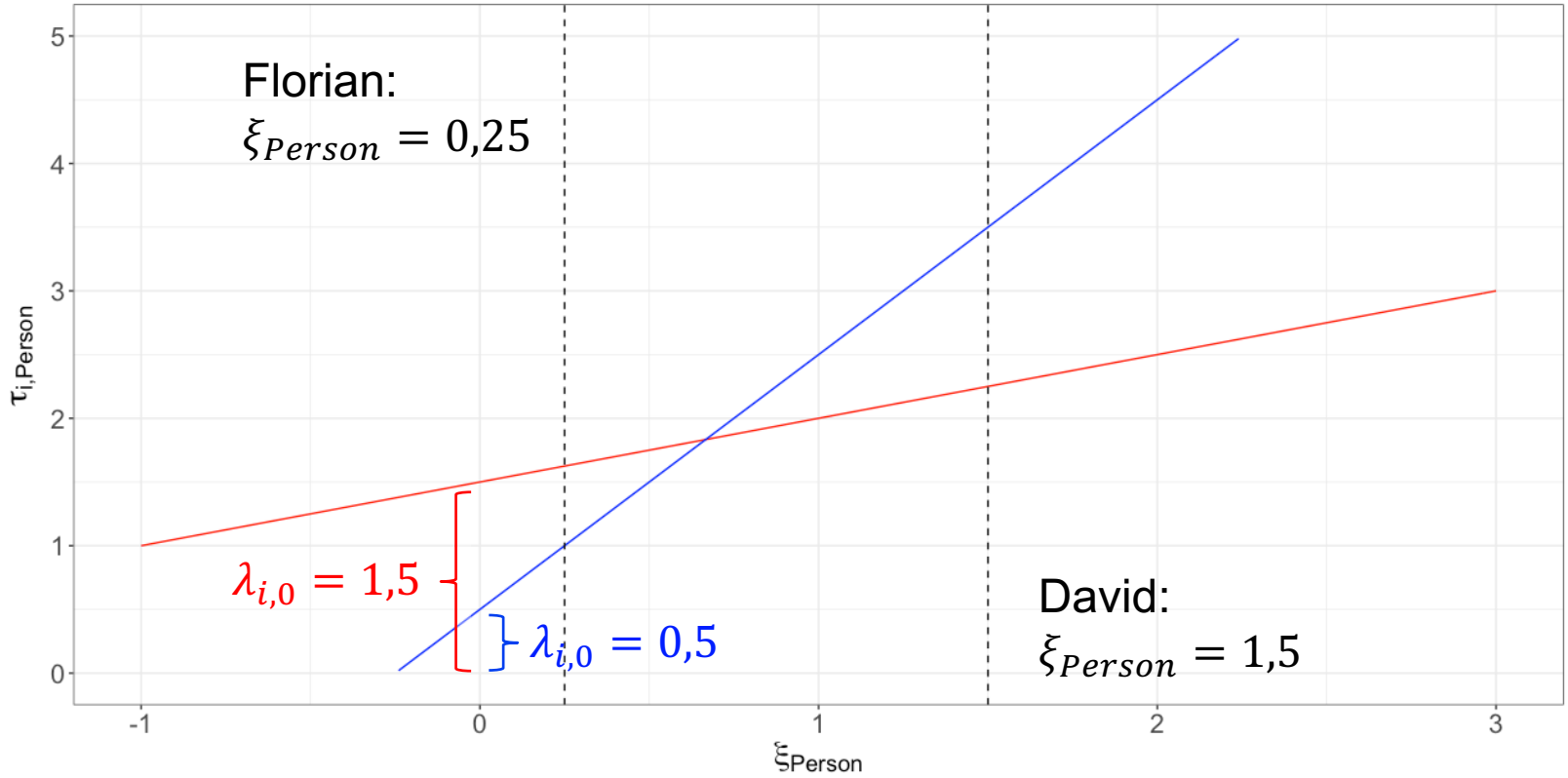
- Der Itemparameter β_i hängt mit der Steigung der ICC von Item i zusammen.
- Je größer β_i, desto stärker verändert sich die Lösungswahrscheinlichkeit für Werte von θ_p nahe bei σ_i (desto „steiler“ ist die ICC am Wendepunkt)
- Der Itemparameter β_i wird daher als „Diskrimination“ von Item i interpretiert.

Schneidende ICCs im 2PL Modell



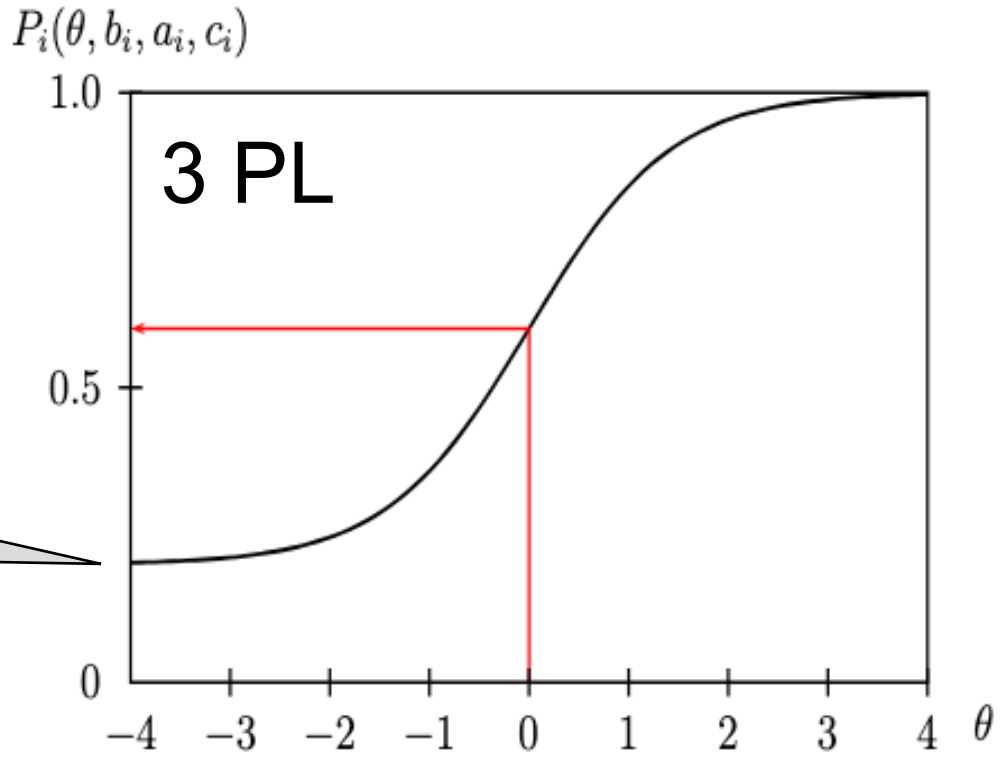
- Die ICCs im 2PL Modell schneiden sich.
- Markus hat eine höhere Wahrscheinlichkeit das rote Item zu lösen als das blaue Item. Caro hingegen hat eine höhere Wahrscheinlichkeit das blaue Item zu lösen als das rote Item. → Vereinbar mit unserer Vorstellung von „Messen“?
- Der Schnitt der Kurven könnte in Bereichen passieren, die empirisch weitestgehend irrelevant sind (siehe „Hypothetische Verteilung der Personenparameter“) – für den von den Daten abgedeckten Bereich können sich die Modelle dann aber flexibler an die Daten anpassen.

Vgl.: Schneidende Itemkurven im τ – kongenerischen Modell



- Dasselbe Phänomen tritt übrigens auch im τ – kongenerischen Modell auf.
- Für Florian ist der wahre Wert auf dem roten Item höher als auf dem blauen Item. Für David hingegen ist der wahre Wert auf dem blauen Item höher als auf dem roten Item. -> Es müssen immer beide Itemparameter berücksichtigt werden

Ausblick: 3PL und 4PL



3. Parameter: Ratewahrscheinlichkeit (startet nicht bei 0)

https://www.metheval.uni-jena.de/irt/VisualIRT.pdf
An Empirical Evaluation of the Slip Correction in the Four Parameter Logistic Models With Computerized Adaptive Testing. Applied Psychological Measurement 36(2):75-87 DOI: 10.1177/0146621611432862

Ausblick: 3PL und 4PL

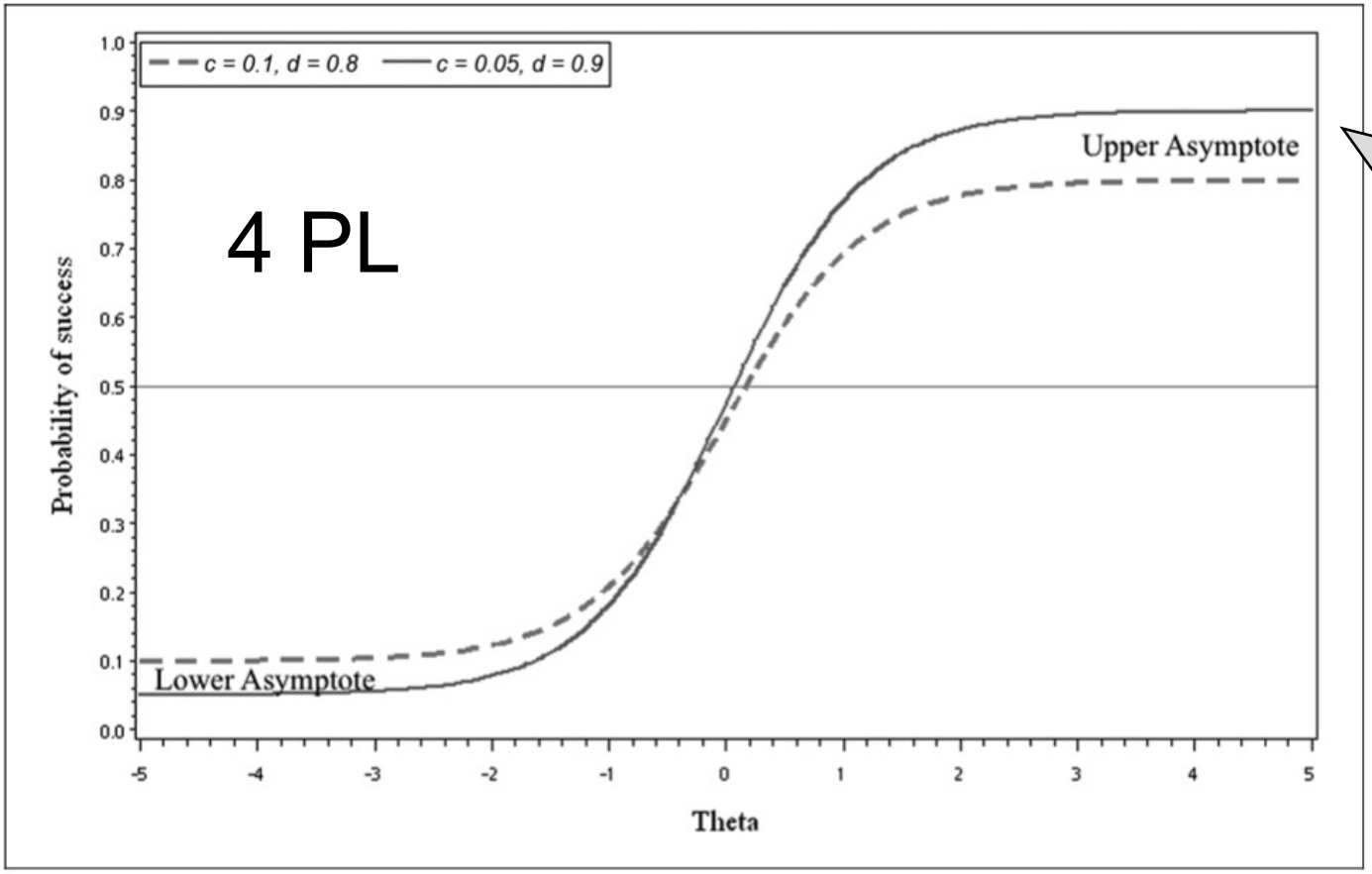


Figure 1. Two typical ICCs for the 4PL IRT model

4. Parameter: Egal wie hoch die Fähigkeit, man erreicht nie (asymptotisch) 100%.
→ „Item specific carelessness“

https://www.metheval.uni-jena.de/irt/VisualIRT.pdf
An Empirical Evaluation of the Slip Correction in the Four Parameter Logistic Models With Computerized Adaptive Testing. Applied Psychological Measurement 36(2):75-87 DOI: 10.1177/0146621611432862

Zusammenfassung: Testmodelle für dichotome Items ($x_{pi} \in \{0,1\}$)

- 1PL Modell („Raschmodell“):

$$P(X_{pi} = x_{pi}|\theta_p) = \frac{e^{x_{pi}\cdot(\theta_p - \sigma_i)}}{1 + e^{(\theta_p - \sigma_i)}}$$

- Items unterscheiden sich nur in ihrer „Schwierigkeit“
- Gegenstück zum essentiell τ – äquivalenten Modell bei Items mit kontinuierlichem Antwortformat

- 2PL Modell:

$$P(X_{pi} = x_{pi}|\theta_p) = \frac{e^{x_{pi}\cdot\beta_i(\theta_p - \sigma_i)}}{1 + e^{\beta_i(\theta_p - \sigma_i)}}$$

- Items unterschieden sich in ihrer „Schwierigkeit“ und in ihrer „Diskrimination“
- Gegenstück zum τ – kongenerischen Modell bei Items mit kontinuierlichem Antwortformat