IRT 2: Item Response Modelle für Ordinale Items

Ordinale Items in der Praxis

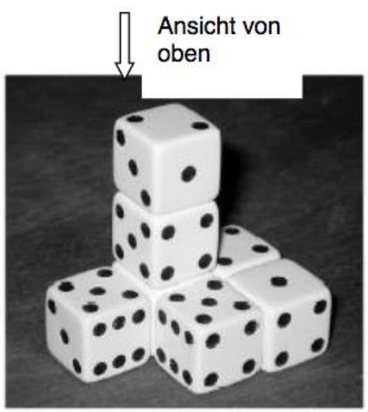
- Ordinale Items findet man meistens in Persönlichkeitstests bzw. Fragebögen
- Bsp: Ordentlichkeitsitem im „NEO Persönlichkeitsinventar“ (NEO-PI-R)

Ich halte meine Sachen ordentlich und sauber.

○	○	○	○	○
starke Ablehnung	Ablehnung	neutral	Zustimmung	starke Zustimmung

$x_{ip} \in \{0, 1, 2, 3, 4\}$

- Wir werden daher oft von der Wahrscheinlichkeit sprechen, dass eine Person bei einem Item eine bestimmte Kategorie „ankreuzt“.
- Messmodelle für ordinale Items sind jedoch auch anwendbar auf Aufgaben in Leistungstests mit Teilpunkten („Partial Credit Items“).
- Bsp: Mathematikaufgabe aus der „PISA Studie 2012“



Im Bild unten wurde ein „Bauwerk“ aus sieben identischen Würfeln gebaut, deren Flächen von 1 bis 6 nummeriert sind.

Wenn man das Bauwerk von oben betrachtet, sind nur 5 Würfel zu sehen.

Wie viele Punkte sind insgesamt sichtbar, wenn man dieses Bauwerk von oben betrachtet?

$x_{ip} \in \{0, 1, 2\}$

$x_{ip} = 2$ („17 Punkte“)
$x_{ip} = 1$ („16 Punkte“)
$x_{ip} = 0$ (andere Anzahl Punkte)

Testmodelle für ordinale Items

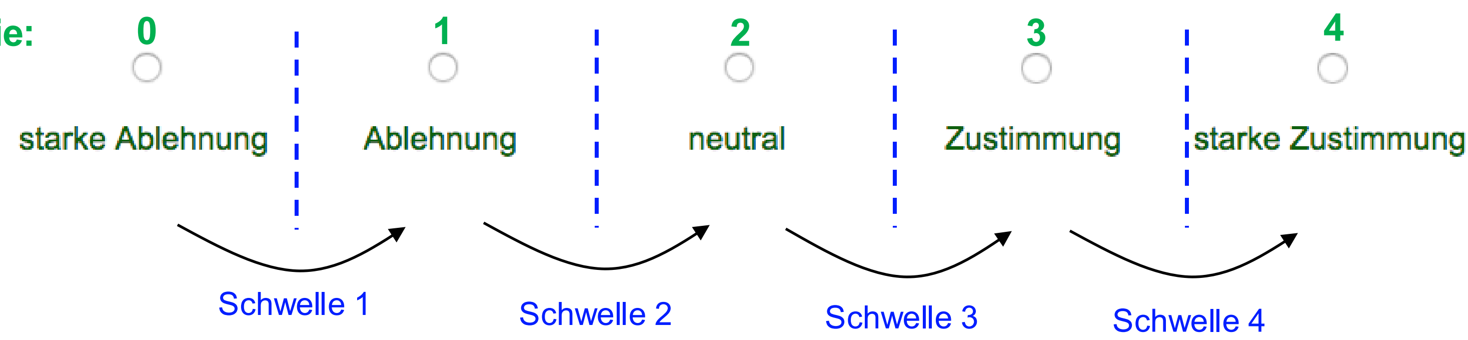
- In Testmodellen für dichotome Items werden für die Antwort einer Person p auf einem Item i die zwei Wahrscheinlichkeiten $P\left(X_{pi} = 1 \middle| \theta_p\right)$ und $P\left(X_{pi} = 0 \middle| \theta_p\right)$ modelliert.
- In Testmodellen für ordinale Items mit K Kategorien werden für die Antwort einer Person p auf einem Item i, K Wahrscheinlichkeiten betrachtet. Es gilt:
$$P\left(X_{pi} = 0 \middle| \theta_p\right) + P\left(X_{pi} = 1 \middle| \theta_p\right) + \ \ldots + P\left(X_{pi} = K - 1 \middle| \theta_p\right) = 1$$
→ Diese Wahrscheinlichkeiten nennt man „Kategorienwahrscheinlichkeiten"
- Analog zu dichotomen Items wird die niedrigste Antwortkategorie typischerweise mit 0 kodiert. Durch diese Konvention lassen sich auch für ordinale Testmodelle alle Kategorienwahrscheinlichkeiten mit einer einzigen Formel berechnen.
- Um Verwirrungen bei den Modellgleichungen zu vermeiden, muss man sich klar machen, dass bei einem Item mit K Kategorien die größtmögliche Itemausprägung $K - 1$ beträgt.
→ Bsp: Item mit fünf Antwortkategorien $(K = 5)$, $x_{ip} \in \{0, 1, 2, 3, 4\}$

Schwellen

$K = 5$

Ich halte meine Sachen ordentlich und sauber.

Antwortkategorie:



Lehrstuhl für Psychologische Methodenlehre und Diagnostik der Ludwig-Maximilians-Universität München

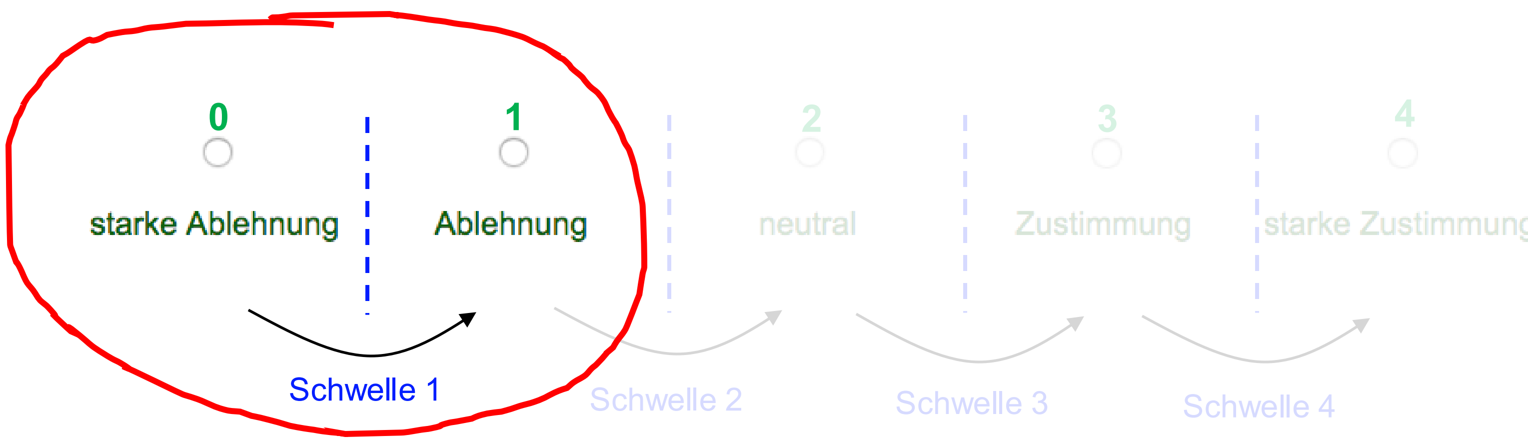
Herleitung: Testmodelle für ordinale Items

- Im dichotomen Raschmodell gilt:

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

- Im „Partial Credit Modell“ für ordinale Items wird die Form des dichotomen Raschmodells zur Modellierung von $K - 1$ dieser sogenannten „Schwellenwahrscheinlichkeiten“ übernommen:

$$\frac{P(X_{pi} = c|\theta_p)}{P(X_{pi} = c|\theta_p) + P(X_{pi} = c - 1|\theta_p)} = \frac{e^{(\theta_p - \tau_{ic})}}{1 + e^{(\theta_p - \tau_{ic})}} \; mit \; c \in \{1, \ldots, K - 1\}$$



Herleitung: Testmodelle für ordinale Items

- Die Schwellenwahrscheinlichkeiten lassen sich interpretieren, als die Wahrscheinlichkeit, eine bestimmte Kategorie c anzukreuzen, unter der Bedingung, dass entweder Kategorie c oder Kategorie $c-1$ angekreuzt wird.

- Anders als bei dichotomen Items gilt bei Items mit mehr als zwei Kategorien:

$$P\left(X_{pi} = c \middle| \theta_p\right) + P\left(X_{pi} = c-1 \middle| \theta_p\right) < 1$$

Exkurs: Herleitung des Partial Credit Modells

- Die Formel für die Schwellenwahrscheinlichkeit lässt sich nach $P(X_{pi} = c|\theta_p)$ auflösen. Damit ergibt sich die folgende rekursive Gleichung:
$$P\left(X_{pi} = c\middle|\theta_p\right) = e^{(\theta_p - \tau_{ic})} \cdot P\left(X_{pi} = c - 1\middle|\theta_p\right)$$
- Wählt man für c zum Beispiel den Wert $K - 1$, kann man die rekursive Gleichung immer wieder in sich selbst einsetzen, bis man bei $P\left(X_{pi} = 0\middle|\theta_p\right)$ angelangt ist.
- Eine weitere Gleichung zur Berechnung von $P\left(X_{pi} = 0\middle|\theta_p\right)$ erhält man durch die zusätzliche Information:
$$P\left(X_{pi} = 0\middle|\theta_p\right) + \cdots + P\left(X_{pi} = K - 1\middle|\theta_p\right) = 1$$
- Löst man diese Formel nach $P\left(X_{pi} = 0\middle|\theta_p\right)$ auf und kombiniert diese mit der oben abgebildeten rekursiven Gleichung, erhält man die allgemeine Modellgleichung des „Partial Credit Modells".

Modellgleichung: Partial Credit Modell

- Modellgleichung des „Partial Credit Modells“ (PCM) für feste Personen:

$$P\left(X_{pi} = x_{pi} \middle| \theta_p\right) = \frac{e^{\sum_{c=1}^{x_{pi}}(\theta_p - \tau_{ic})}}{1 + \sum_{s=1}^{K-1} e^{\sum_{c=1}^{s}(\theta_p - \tau_{ic})}}$$

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

Plural: Jedes Item hat mehrere Parameter, da es mehrere Kategorienübergänge hat

- Um alle Kategorienwahrscheinlichkeiten mithilfe einer einzigen Gleichung darstellen zu können, wird die folgende übliche Konvention getroffen:

$$\sum_{c=1}^{0}(\theta_p - \tau_{ic}) = 0$$

- In der Mathematik nennt man $\sum_{c=1}^{0} x$ eine leere Summe. Per Konvention hat die leere Summe immer den Wert Null, unabhängig vom Wert von x.

Kategorienwahrscheinlichkeiten im PCM ausrechnen

- Beispiel: Berechnung aller Kategorienwahrscheinlichkeiten bei einem Item i mit $K = 3$ Antwortkategorien ($\tau_{i1} = -1, \tau_{i2} = 1$) für eine Person p mit $\theta_p = 1$.

- $P\left(X_{pi} = 0 \middle| \theta_p\right) = \frac{e^{\sum_{c=1}^{0}(\theta_p - \tau_{ic})}}{1+\sum_{s=1}^{2} e^{\sum_{c=1}^{S}(\theta_p - \tau_{ic})}} = \frac{e^0}{1+e^{\sum_{c=1}^{1}(\theta_p - \tau_{ic})}+e^{\sum_{c=1}^{2}(\theta_p - \tau_{ic})}} =$
  $$= \frac{1}{1 + e^{(\theta_p - \tau_{i1})} + e^{(\theta_p - \tau_{i1})+(\theta_p - \tau_{i2})}} = \frac{1}{1 + e^{(1-(-1))} + e^{(1-(-1))+(1-1)}} = 0{,}06$$

- $P\left(X_{pi} = 1 \middle| \theta_p\right) = \frac{e^{\sum_{c=1}^{1}(\theta_p - \tau_{ic})}}{1+\sum_{s=1}^{2} e^{\sum_{c=1}^{S}(\theta_p - \tau_{ic})}} = \frac{e^{(\theta_p - \tau_{i1})}}{1+e^{\sum_{c=1}^{1}(\theta_p - \tau_{ic})}+e^{\sum_{c=1}^{2}(\theta_p - \tau_{ic})}} =$
  $$= \frac{e^{(\theta_p - \tau_{i1})}}{1 + e^{(\theta_p - \tau_{i1})} + e^{(\theta_p - \tau_{i1})+(\theta_p - \tau_{i2})}} = \frac{e^2}{1 + e^2 + e^{2+0}} = 0{,}47$$

- $P\left(X_{pi} = 2 \middle| \theta_p\right) = \frac{e^{\sum_{c=1}^{2}(\theta_p - \tau_{ic})}}{1+\sum_{s=1}^{2} e^{\sum_{c=1}^{S}(\theta_p - \tau_{ic})}} = \frac{e^{(\theta_p - \tau_{i1})+(\theta_p - \tau_{i2})}}{1+e^{\sum_{c=1}^{1}(\theta_p - \tau_{ic})}+e^{\sum_{c=1}^{2}(\theta_p - \tau_{ic})}} =$
  $$= \frac{e^{(\theta_p - \tau_{i1})+(\theta_p - \tau_{i2})}}{1 + e^{(\theta_p - \tau_{i1})} + e^{(\theta_p - \tau_{i1})+(\theta_p - \tau_{i2})}} = \frac{e^{2+0}}{1 + e^2 + e^{2+0}} = 0{,}47$$

Normierungskonstante des PCM

- Die komplizierte „Normierungskonstante“ im Nenner der Modellgleichung des PCM sorgt dafür, dass sich die Kategorienwahrscheinlichkeiten eines Items i für ein beliebiges θ_p zu 1 aufsummieren.
- Veranschaulichung am Beispiel eines Items mit drei Antwortkategorien:

$$P\left(X_{pi}=0\middle|\theta_p\right)+P\left(X_{pi}=1\middle|\theta_p\right)+P\left(X_{pi}=2\middle|\theta_p\right)=$$

$$=\frac{1}{1+e^{(\theta_p-\tau_{i1})}+e^{(\theta_p-\tau_{i1})+(\theta_p-\tau_{i2})}}+$$

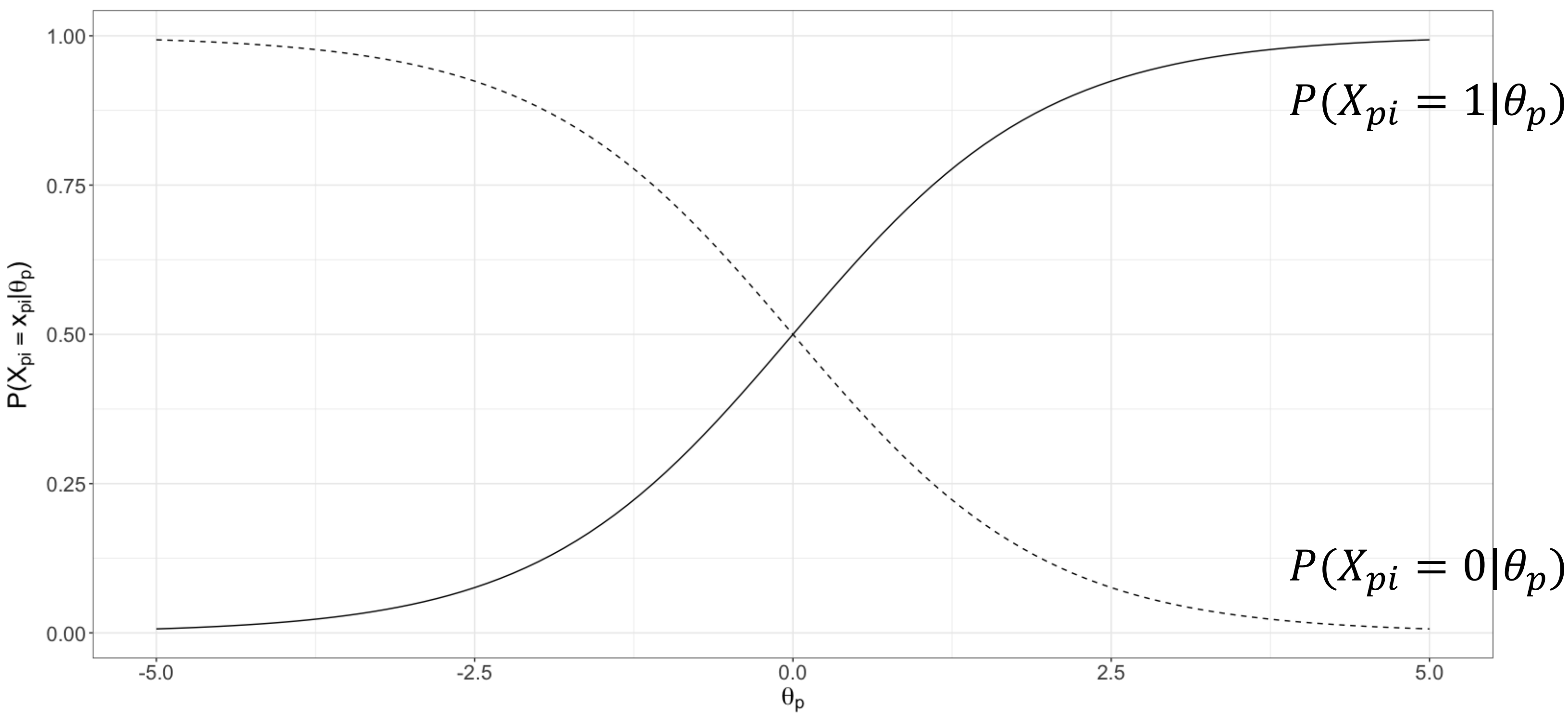
$$+\frac{e^{(\theta_p-\tau_{i1})}}{1+e^{(\theta_p-\tau_{i1})}+e^{(\theta_p-\tau_{i1})+(\theta_p-\tau_{i2})}}+$$

$$+\frac{e^{(\theta_p-\tau_{i1})+(\theta_p-\tau_{i2})}}{1+e^{(\theta_p-\tau_{i1})}+e^{(\theta_p-\tau_{i1})+(\theta_p-\tau_{i2})}}=$$

$$=\frac{1+e^{(\theta_p-\tau_{i1})}+e^{(\theta_p-\tau_{i1})+(\theta_p-\tau_{i2})}}{1+e^{(\theta_p-\tau_{i1})}+e^{(\theta_p-\tau_{i1})+(\theta_p-\tau_{i2})}}=1$$

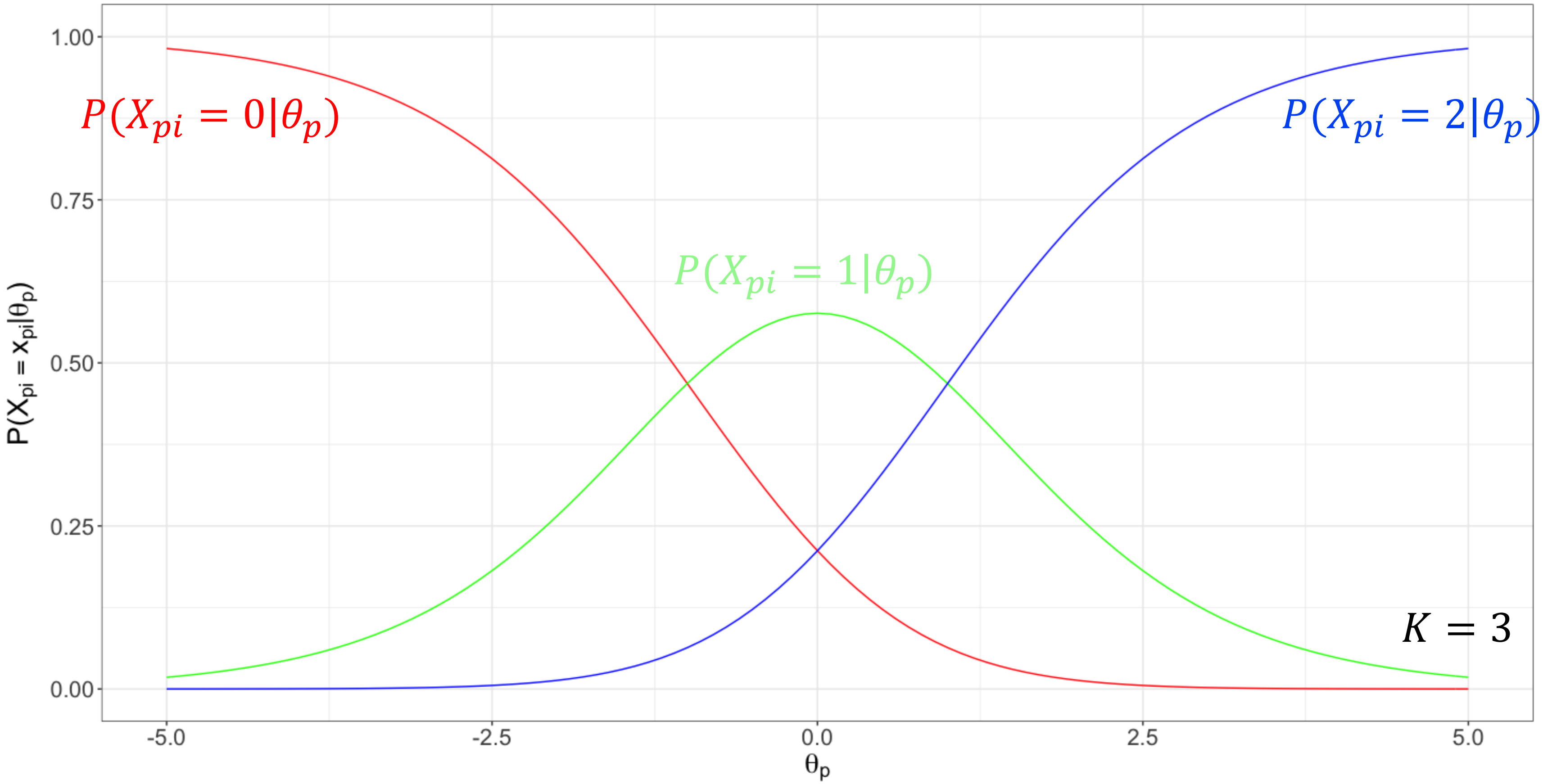
Category Characteristic Curves

- In Item Response Modellen betrachtet man zur grafischen Interpretation eines Items i häufig die Kategorienwahrscheinlichkeiten als Funktion in Abhängigkeit des Werts einer Person auf der latenten Variablen θ_p.
- Wie bereits erwähnt, betrachtet man bei dichotomen Items meist nur die Funktion $P\left(X_{pi} = 1 \middle| \theta_p\right)$ und bezeichnet diese als ICC des Items i.

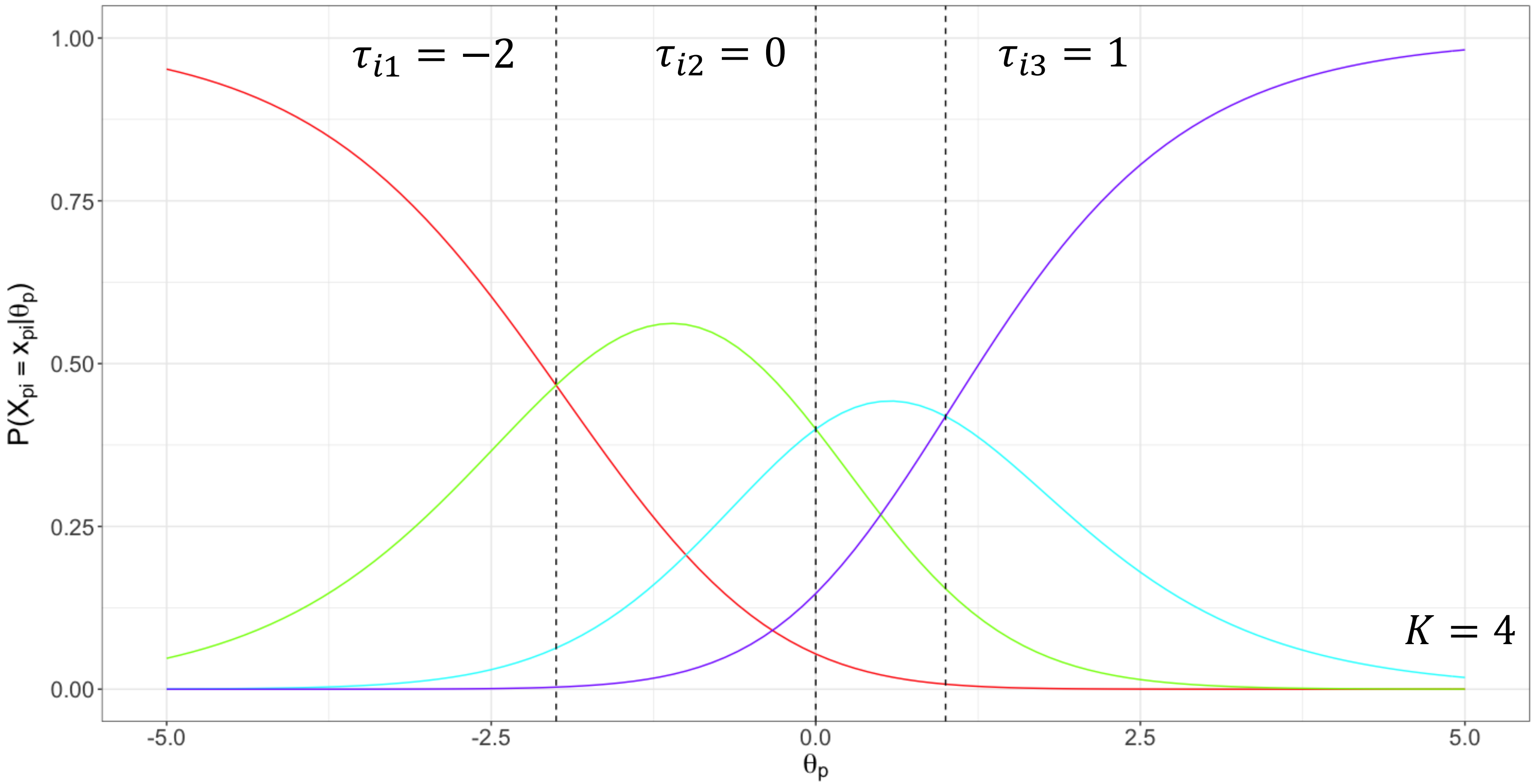


Category Characteristic Curves

- Bei ordinalen Items mit mehr als zwei Antwortkategorien werden hingegen sinnvollerweise alle K „Category Characteristic Curves" (CCCs) betrachtet.
- Für jeden Wert θ_p gilt: $P\left(X_{pi} = 0 \middle| \theta_p\right) + \cdots + P\left(X_{pi} = K - 1 \middle| \theta_p\right) = 1$



Interpretation der Itemparameter

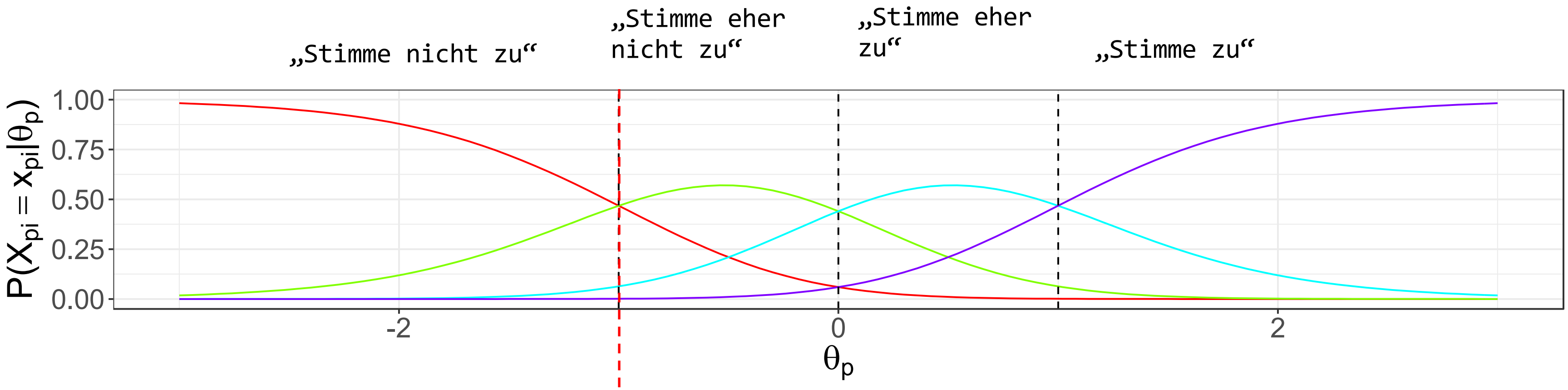


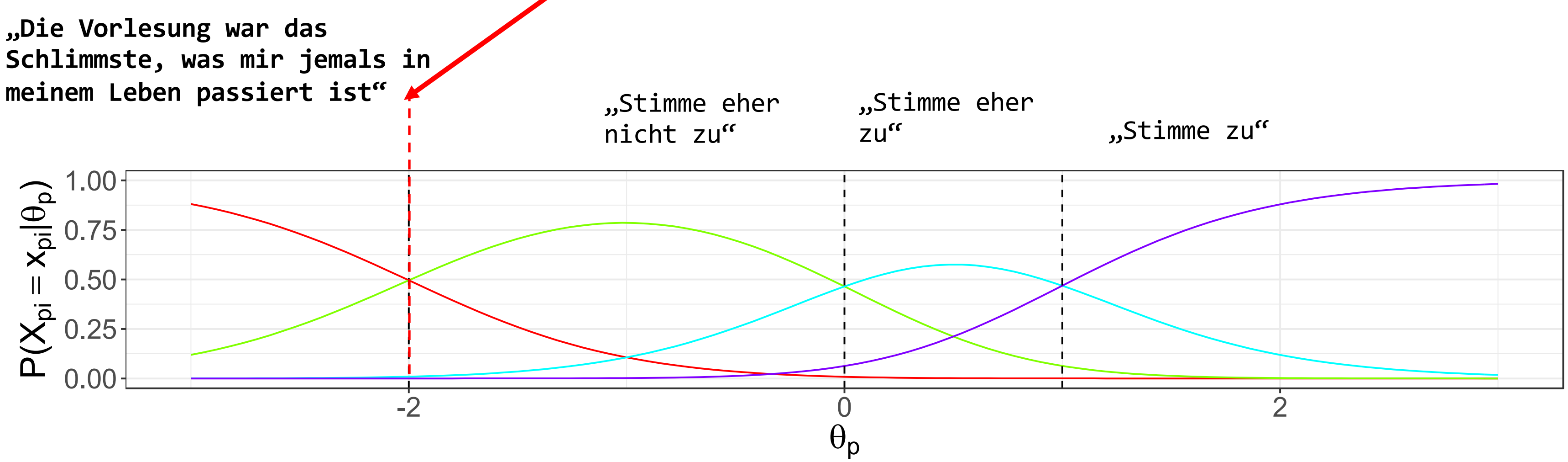
- Der „Schwellenparameter“ τ_{ic} markiert die Stelle, an denen sich die CCCs der Kategorien c und $c-1$ von Item i schneiden.
- Für eine Person mit $\theta_p = \tau_{ic}$ ist die Wahrscheinlichkeit Kategorie c anzukreuzen genauso groß wie die Wahrscheinlichkeit Kategorie $c-1$ anzukreuzen.

Lehrstuhl für Psychologische Methodenlehre und Diagnostik der Ludwig-Maximilians-Universität München

Ankerformulierung → Verschiebung der Schwellenwerte

Wie sehr stimmen Sie dieser Aussage zu? „Die Vorlesung in Item Response Theorie gefällt mir gut.“

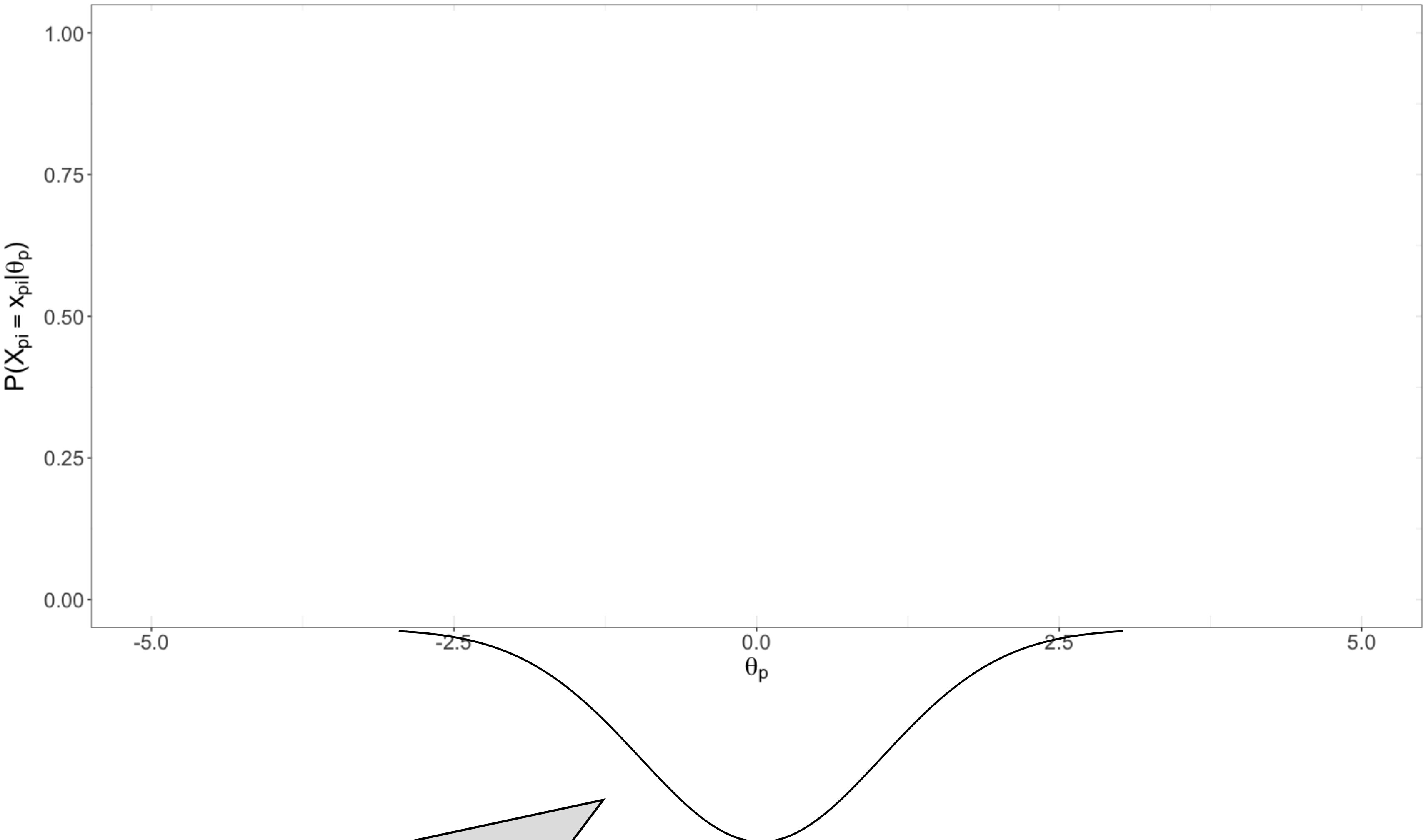




CCCs und die Verteilung der Itemantworten

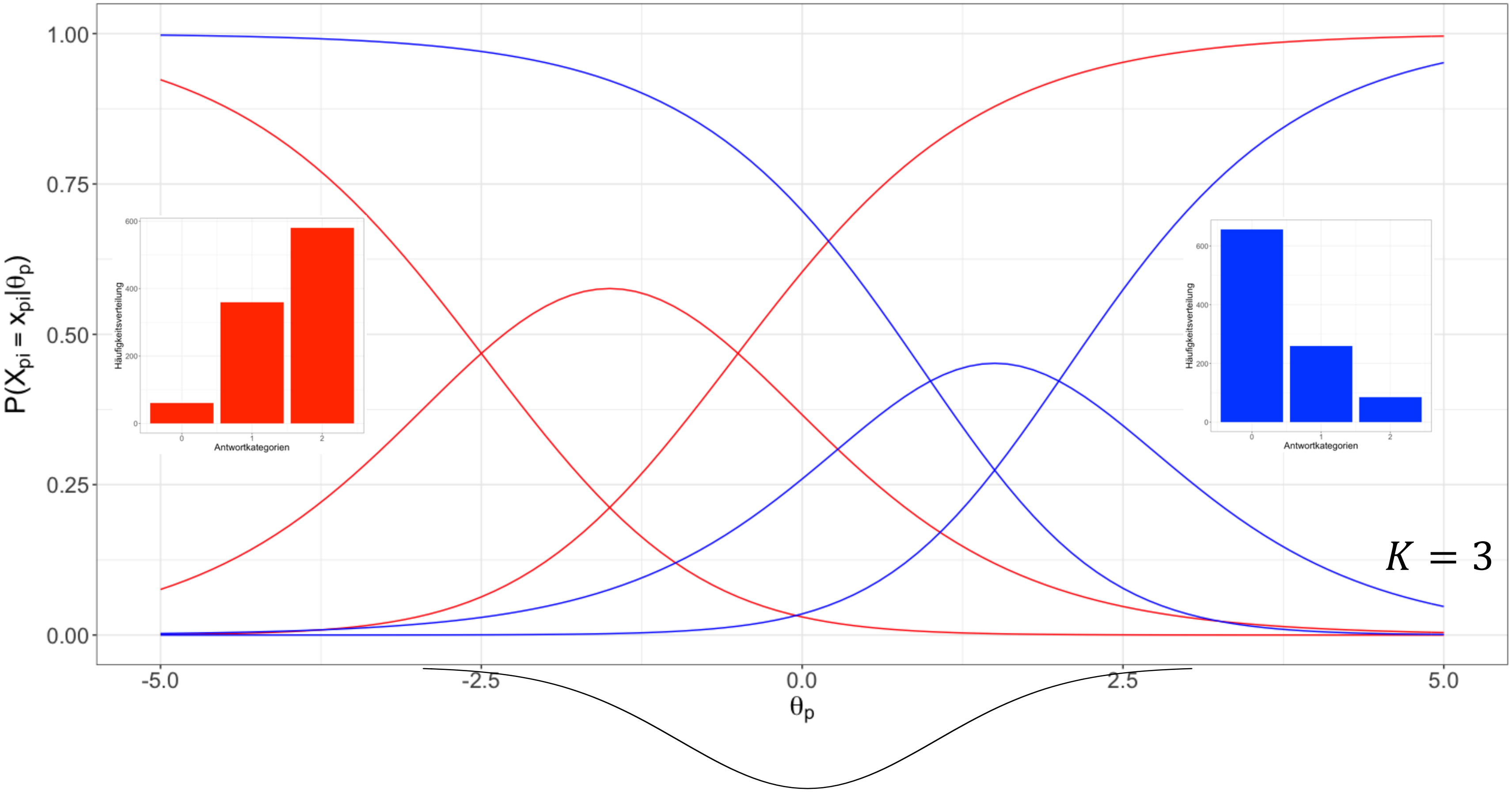
- Zur Interpretation der CCCs eines Items ist es hilfreich sich zu überlegen, welche theoretische Verteilung der Itemantworten durch die CCCs impliziert wird.
- Man beachte dabei, dass die Itemantworten nicht nur von den Itemparametern, sondern auch von der Verteilung der Personen auf der latenten Variable abhängen.
- Aus Gründen der Normierung, die im Kapitel zur Modellschätzung noch genauer behandelt werden, nimmt man für die Verteilung der Personen auf der latenten Variable in der Regel eine Standardnormalverteilung an.
- Um für ein Item mit gegebenen Schwellenwerten die Verteilung der Itemantworten grafisch in einem Balkendiagramm darstellen zu können, zieht man zunächst eine große Menge von θ_p aus der Standardnormalverteilung und simuliert dann für jedes p eine Itemantwort mithilfe der Modellgleichung.
- **Hinweis:**
  Für alle Beispiele wurden zur Darstellung der Balkendiagramme immer $P = 1000$ „Personen“ zufällig aus der Standardnormalverteilung gezogen.

1000 simulierte Personen



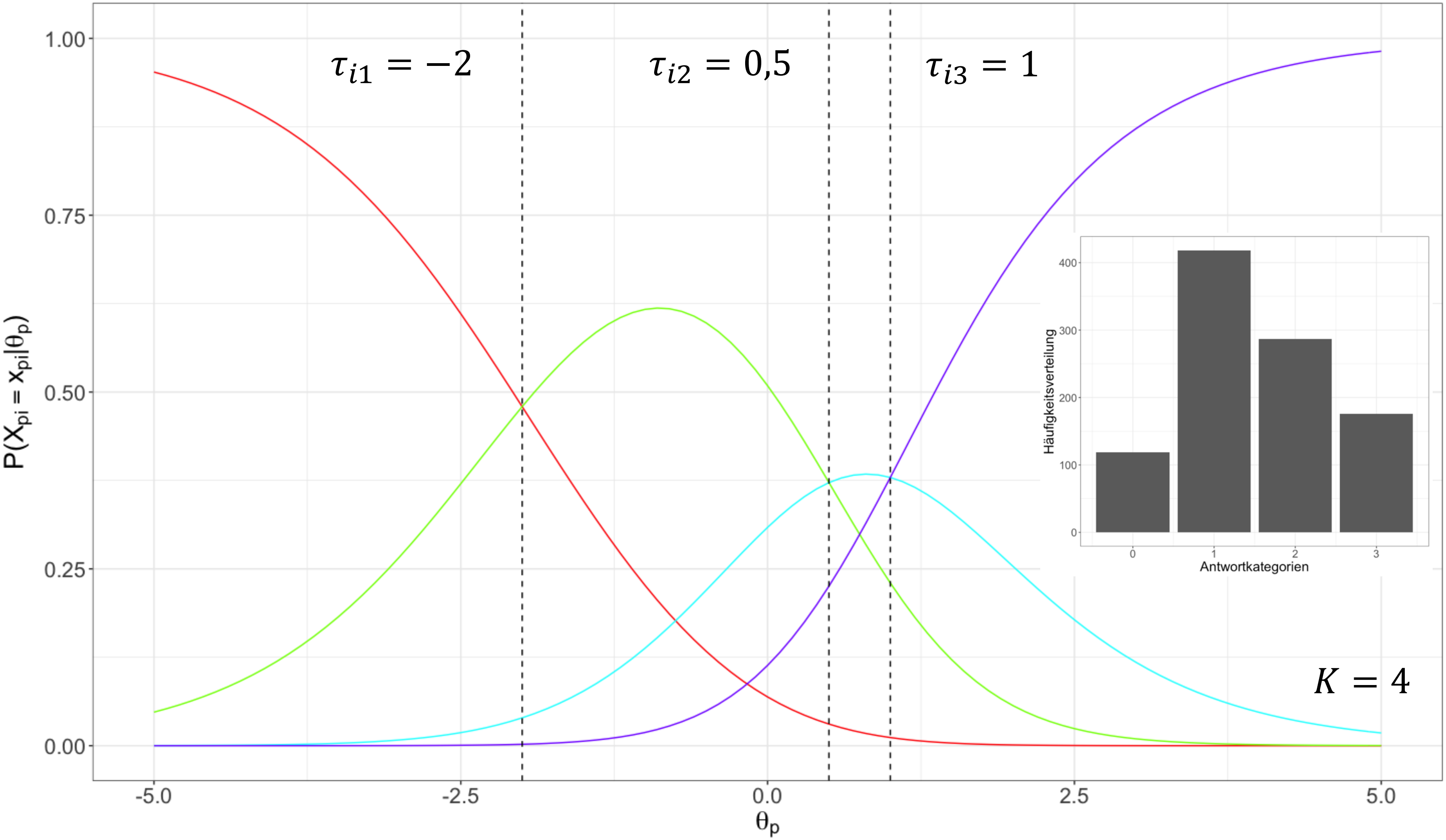
Simulierte Verteilung der Personenparameter (standardnormalverteilt)

Simulierte Itemantworten der 1000 simulierten Personen auf zwei Items

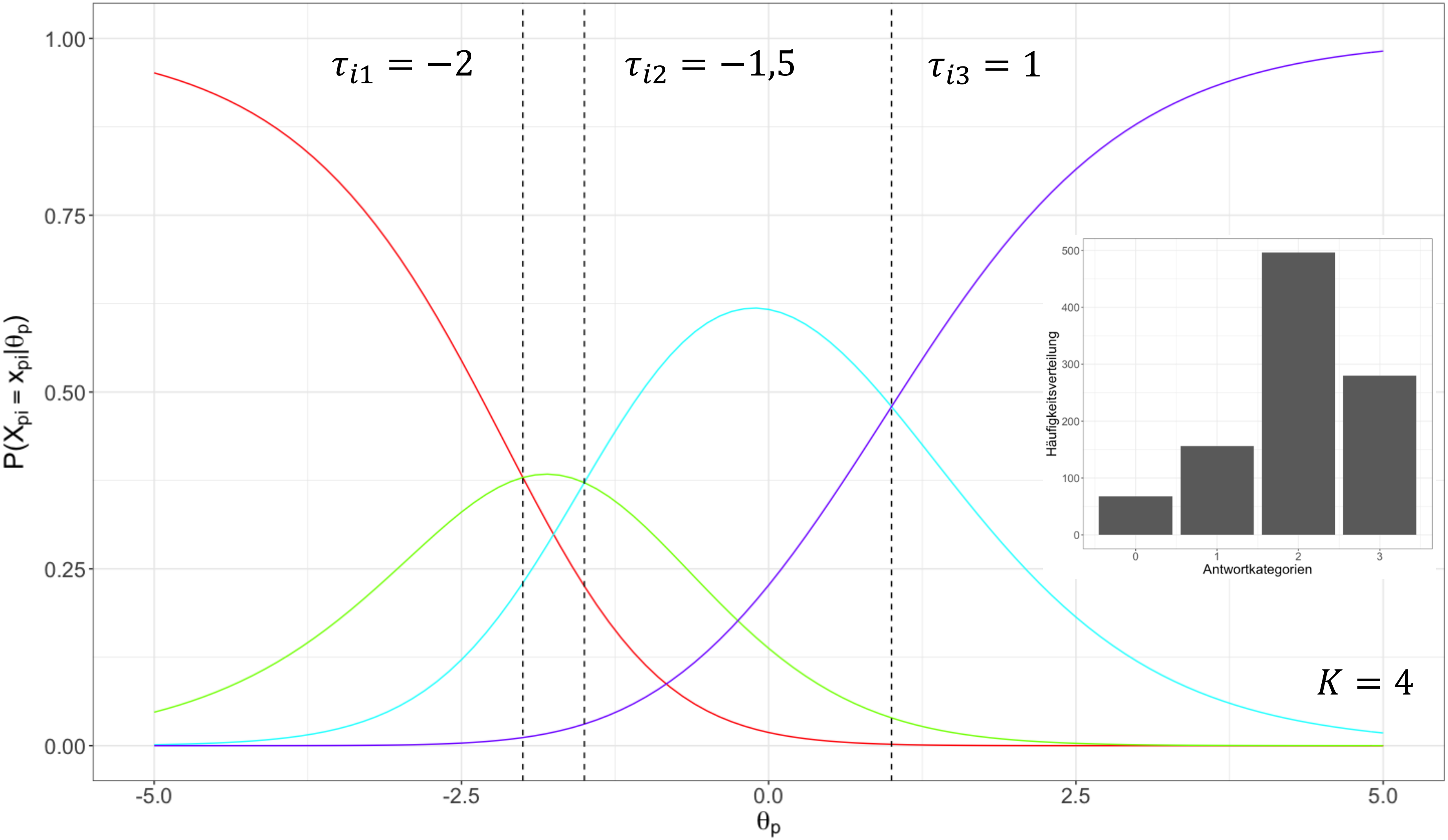


- Die zwei Histogramme für das rote und das blaue Item zeigen die simulierten Itemantworten von den 1000 simulierten Personen.
- Das rote Item ist „leichter“: Es werden eher hohe Antwortkategorien angekreuzt. Das blaue Item ist schwerer: Bei denselben Ausprägungen der Personenparameter werden eher niedrige Antwortkategorien angekreuzt.

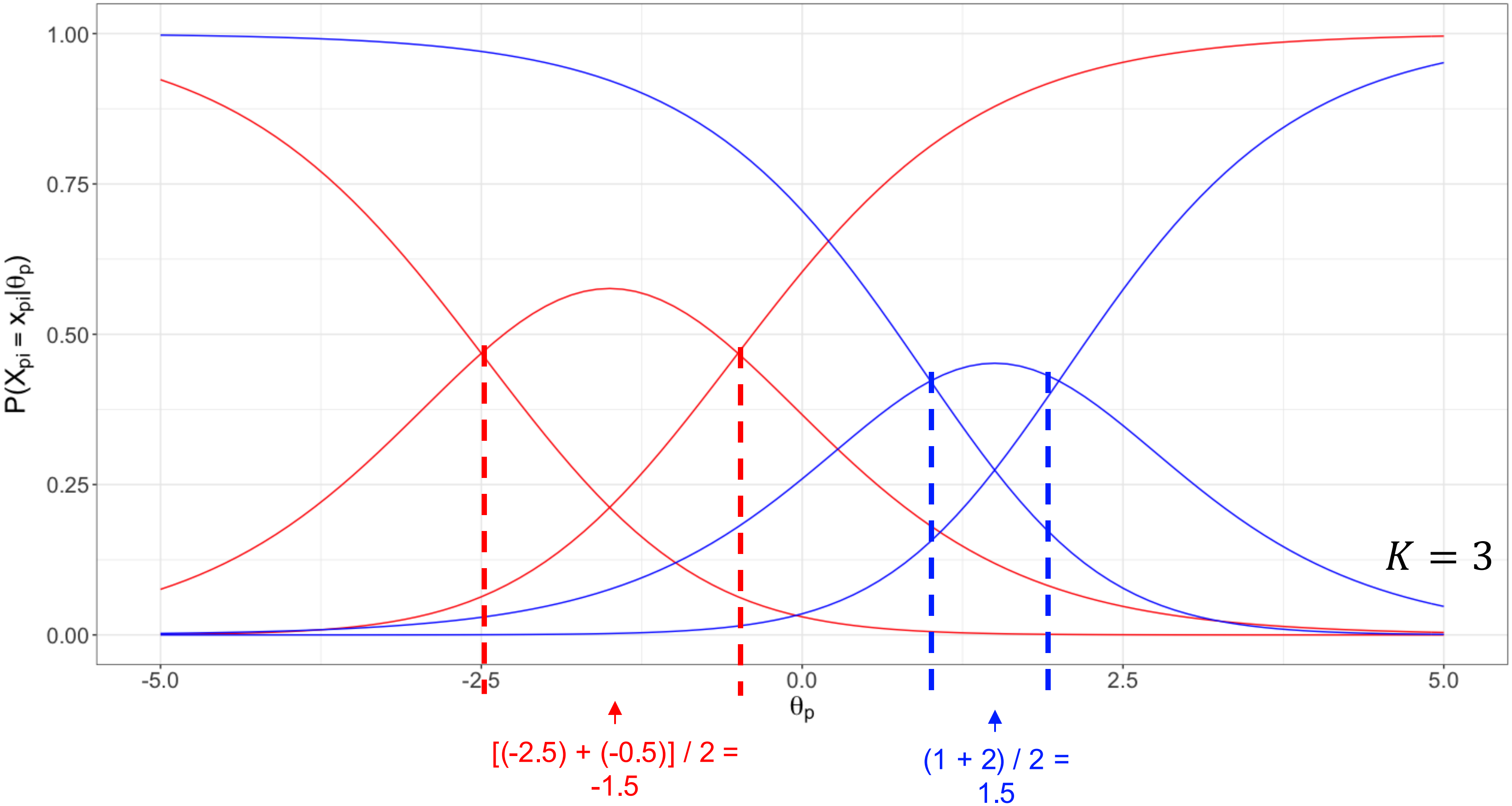
Jede CCC eines Items hängt von allen Schwellen ab



Jede CCC eines Items hängt von allen Schwellen ab



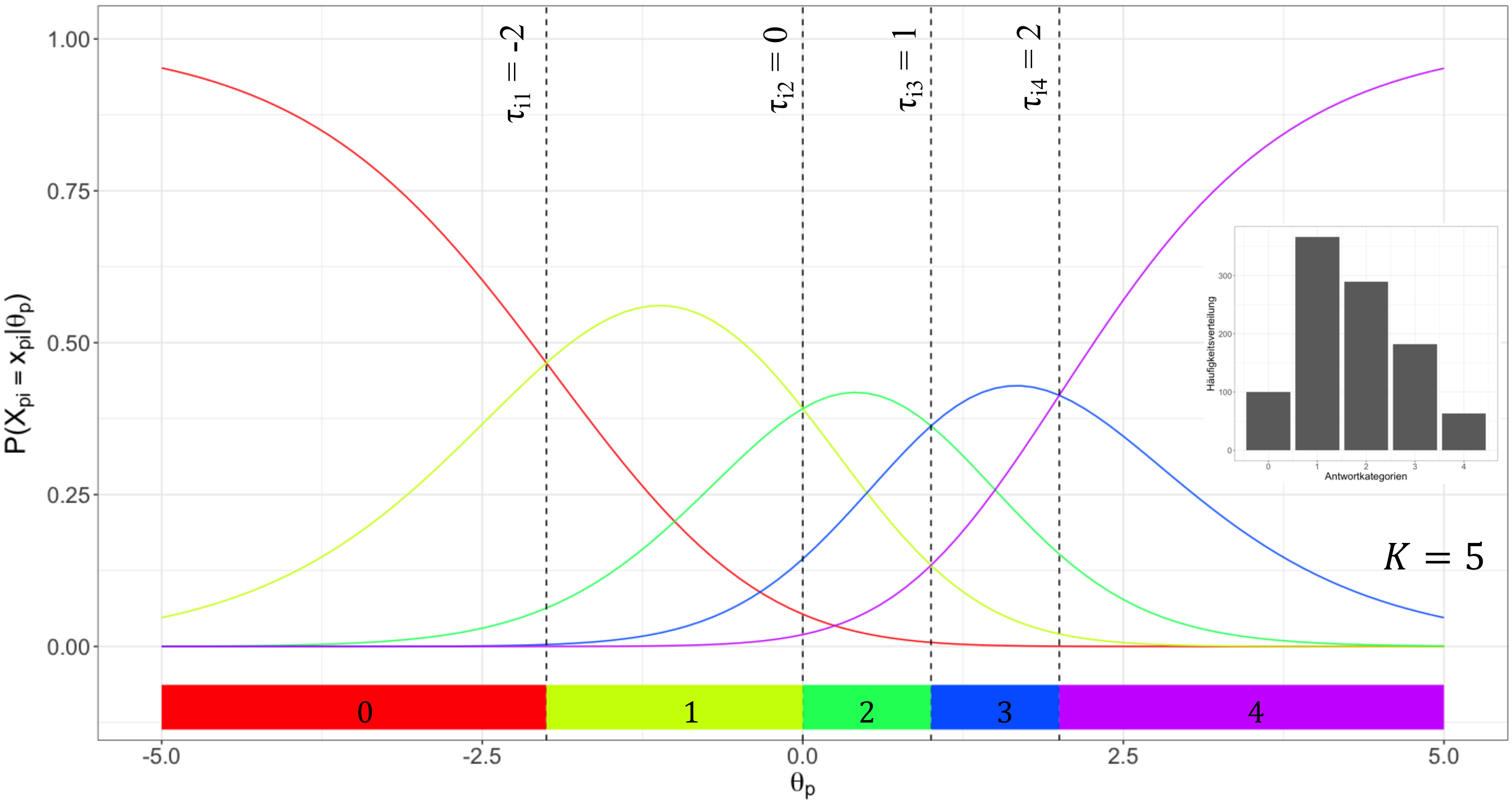
Itemschwierigkeit im PCM (Darstellung von zwei Items)



- Im PCM wird die „Schwierigkeit" eines Items nicht mehr durch einen einzelnen Itemparameter charakterisiert, sondern hängt von der Lage aller Schwellen ab.
- Eine sinnvolle Möglichkeit, „die Schwierigkeit" eines Items im PCM zu quantifizieren ist der Mittelwert der Schwellen $\frac{1}{K-1}\sum_{c=1}^{K-1}\tau_{ic}$

Regionen der latenten Dimension

- Zur Interpretation der CCCs eines Items ist es sinnvoll zu betrachten, welche Antwortkategorie an einem bestimmten Punkt θ_p am wahrscheinlichsten ist:



Item Characteristic Curve (ICC) für ordinale Items

- Als Alternative zu den Regionen mit der höchsten Wahrscheinlichkeit für eine bestimmte Antwortkategorie anhand der CCCs, kann auch die ICC für ein ordinales Item grafisch veranschaulicht werden.
- Bei dichotomen Items gibt die ICC die erwartete Itemantwort für einen bestimmten Wert auf der latenten Variable an:

$$E\left(X_{pi} \middle| \theta_p\right) = P\left(X_{pi} = 1 \middle| \theta_p\right)$$

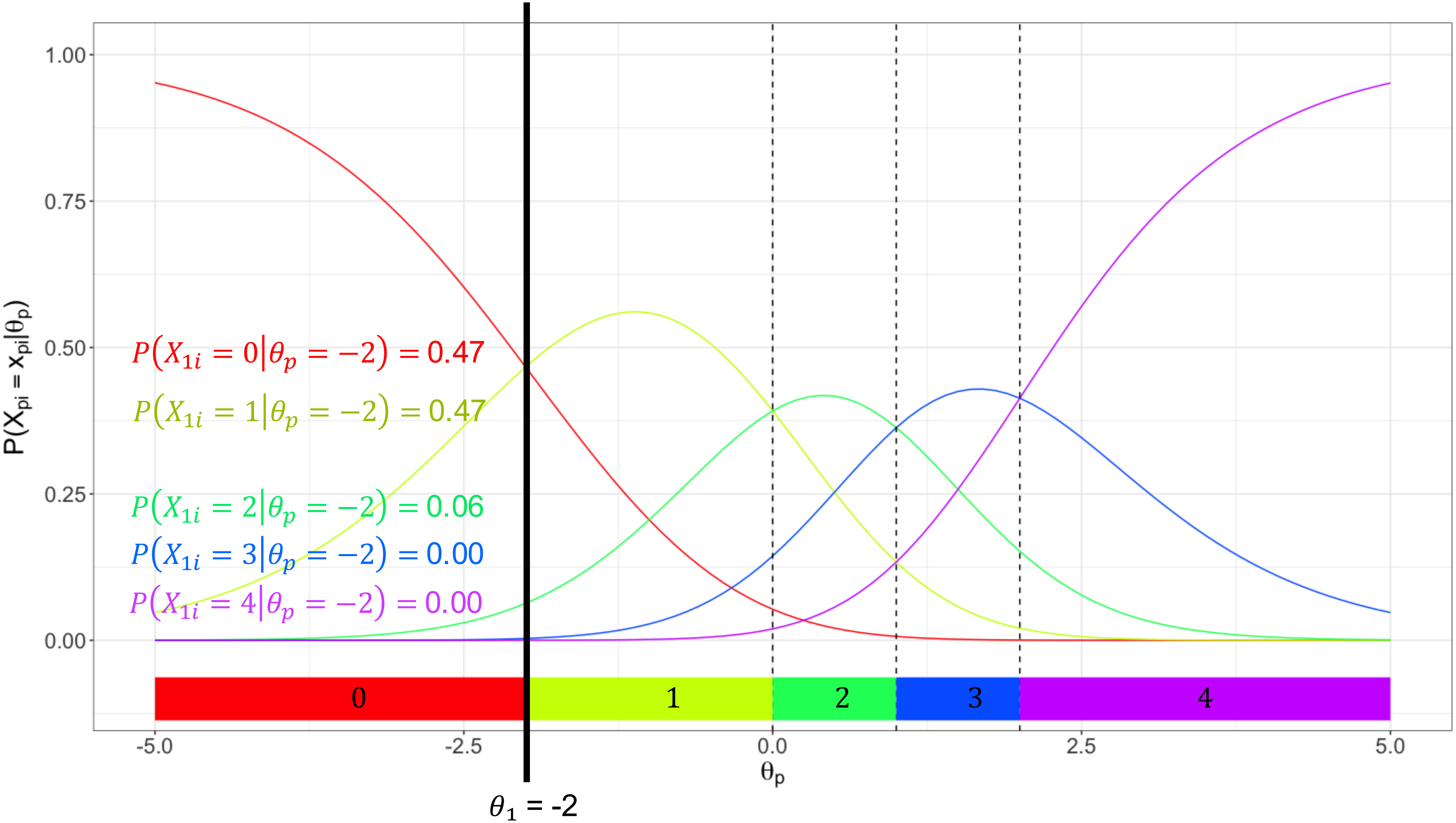
- Interpretation: Angenommen Person p bearbeitet Item i unendlich oft, dann entspricht $E\left(X_{pi} \middle| \theta_p\right)$ dem Mittelwert dieser unendlich vielen Itemantworten.
- Die erwartete Itemantwort, bedingt auf den Wert der Person auf der latenten Variable lässt sich im Rahmen des PCMs auch für ordinale Items berechnen:

$$E\left(X_{pi} \middle| \theta_p\right) = \sum_{k=0}^{K-1} k \cdot P\left(X_{pi} = k \middle| \theta_p\right)$$

Allgemeine Form des (bedingten) Erwartungswert einer diskreten Zufallsvariable

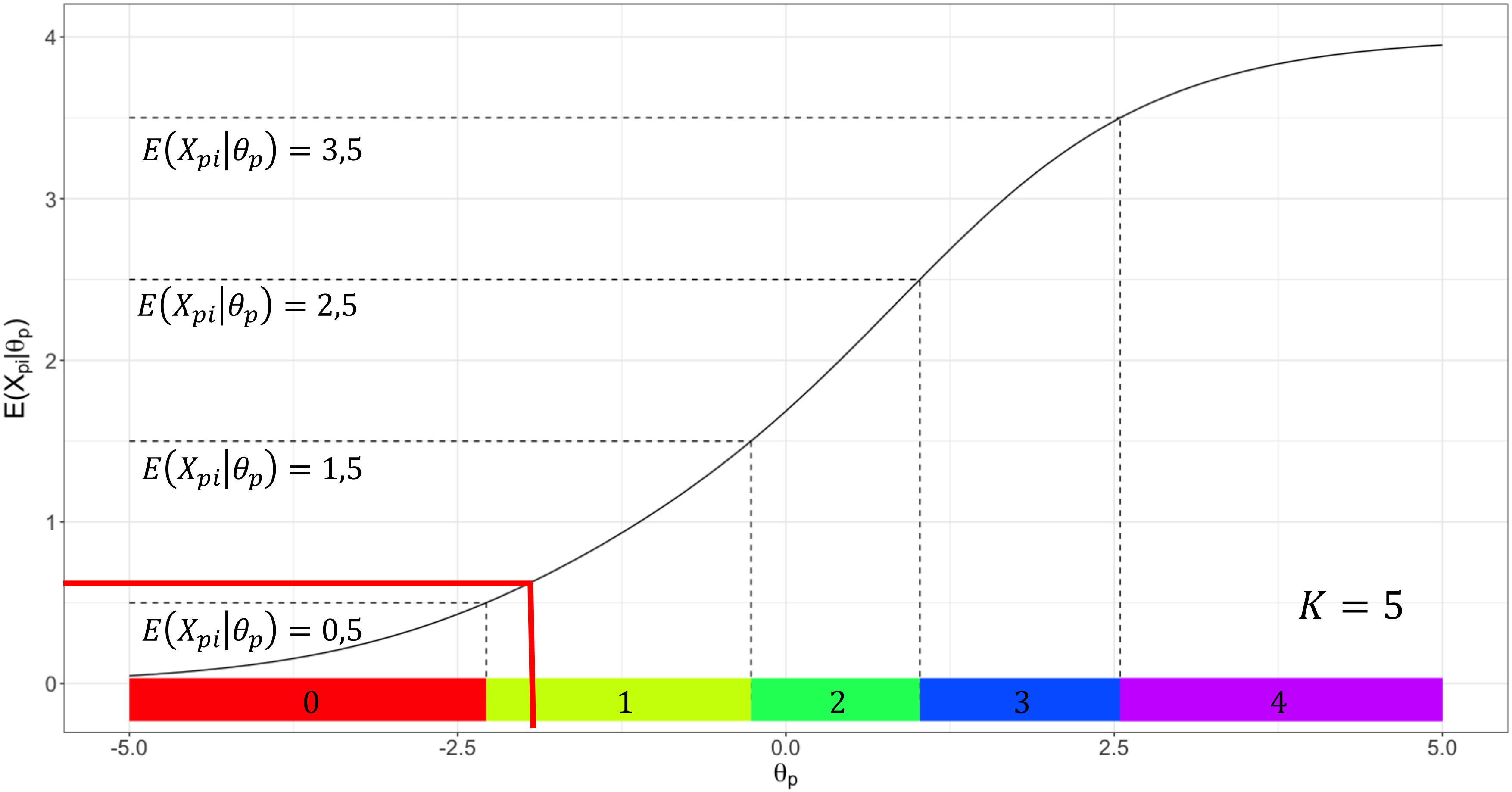
Formel für die Berechnung Kategorienwahrscheinlichkeiten aus der Modellgleichung des PCM

Erwartungswert einer Person mit $\theta_1 = -2$



$$E(X_{1i}|\theta_1 = -2) = 0.47 \cdot 0 + 0.47 \cdot 1 + 0.06 \cdot 2 + 0.00 \cdot 3 + 0.00 \cdot 4 = 0.59$$

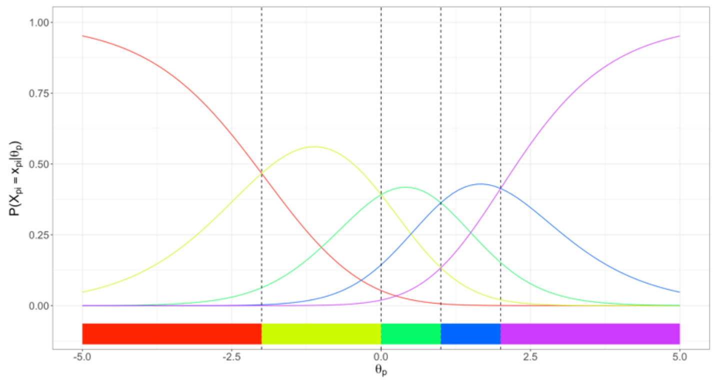
ICC (gleiches Item wie auf Folie 21)



- Die Grenzen in dieser Darstellung der ICC des ordinalen Items von Folie 21 entsprechen nicht den Schwellenwerten τ_{ic}, sondern stellen eine alternative Unterteilung der latenten Variable dar.
  → $E(X_{pi}|\theta_p)$ wird über alle Antwortkategorien berechnet und „gerundet"

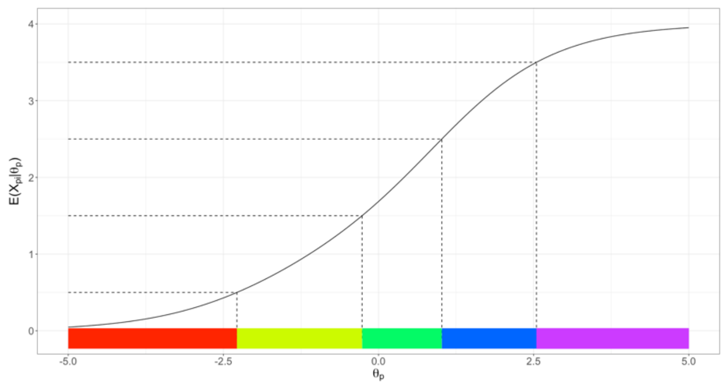
CCC vs. ICC

CCC



Was ist die wahrscheinlichste Antwortkategorie gegeben einer latenten Ausprägung?

ICC

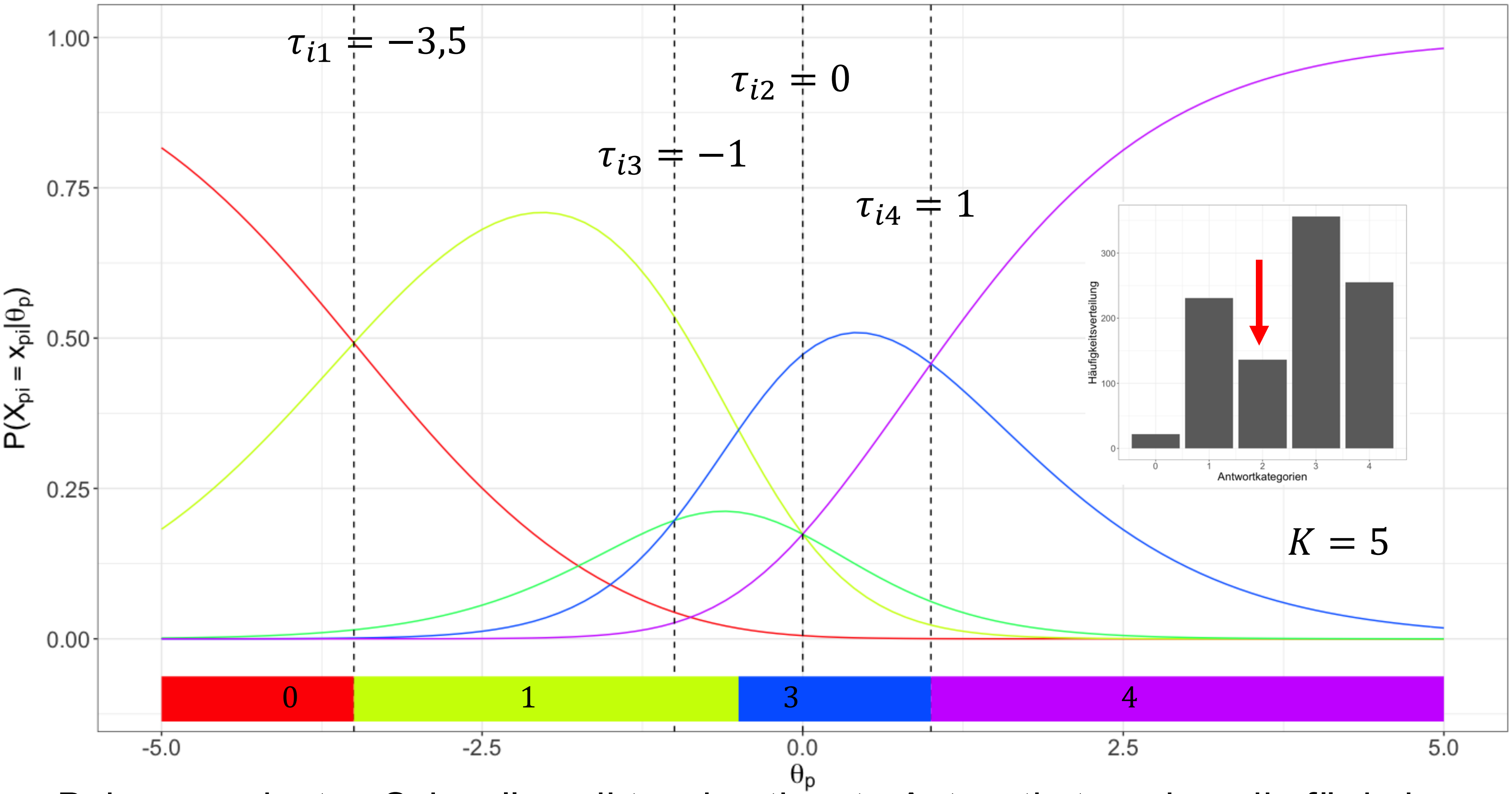


Was ist der erwartete Antwortwert gegeben einer latenten Ausprägung (d.h., gemittelt über alle Antwortkategorien)?

Ordnung der Schwellenparameter

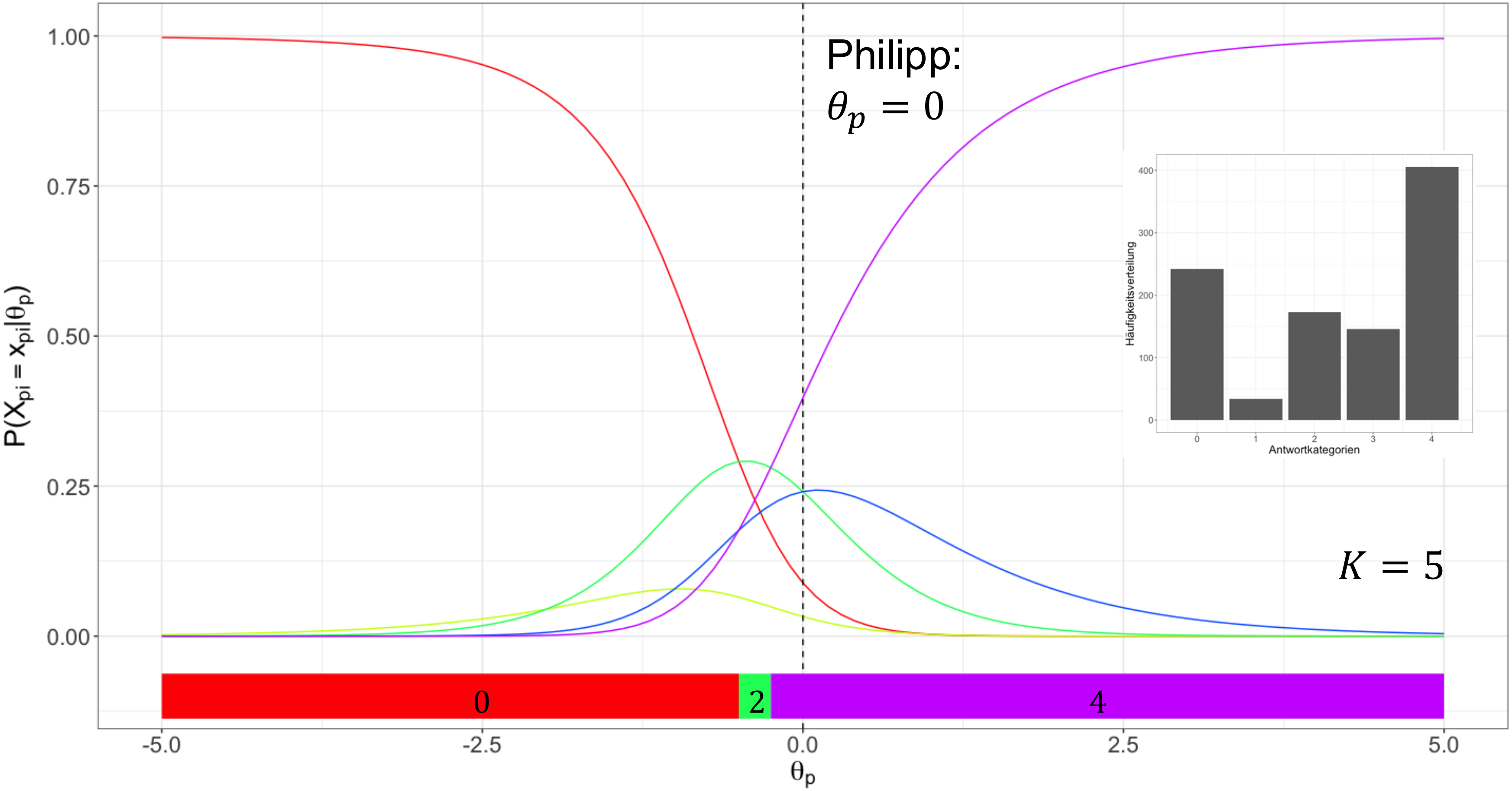
- In den bisherigen Beispielen galt implizit immer $\tau_{i1} < \tau_{i2} < \cdots < \tau_{i,K-1}$.
- Diese intuitive Ordnung der Schwellen ist jedoch durch das PCM nicht vorgeschrieben. Tatsächlich erlaubt das Modell beliebige Rangreihen der Schwellenparameter τ_{ic}.
- Damit ist das PCM flexibel genug, um auch Items mit komplizierten Verteilungen der Antwortkategorien abbilden zu können.
- Items mit ungeordneten Schwellen sind keine rein theoretische Konstruktion, sondern treten auch in der praktischen Auswertung von Fragebögen häufig auf.
- Obwohl ungeordnete Schwellen an sich keine Verletzung des Testmodells darstellen, deuten sie trotzdem häufig auf Probleme etwa bei der Item-formulierung, dem Itemverständnis oder der Wahl der Antwortkategorien hin.

CCCs mit ungeordneten Schwellen



- Bei ungeordneten Schwellen gibt es bestimmte Antwortkategorien, die für keine Ausprägung der Personen auf der latenten Variable am wahrscheinlichsten sind. Diese Items weisen häufig eine auffällige Verteilung der Itemantworten auf.

CCCs mit ungeordneten Schwellen



- Angenommen, das abgebildete Item stellt eine Klausuraufgabe dar: In dieser Aufgabe ist es für Philipp ebenso wahrscheinlich 2 Punkte zu erzielen, wie 3 Punkte. Er erhält wahrscheinlicher 0 Punkte als 1 Punkt. Am wahrscheinlichsten ist es jedoch, dass Philipp 4 Punkte erzielt.

Generalized Partial Credit Model

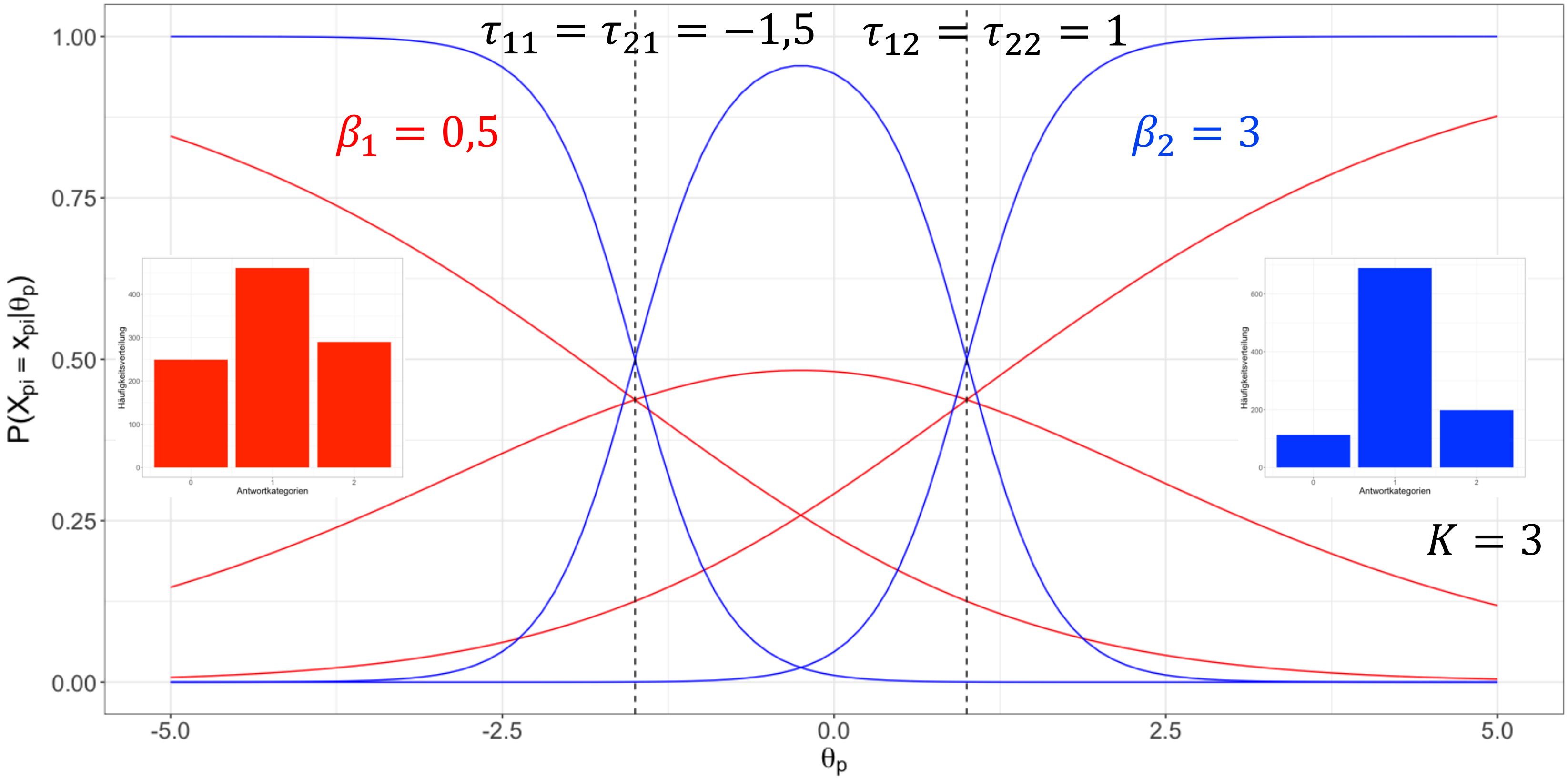
- Im PCM wird jedes Item durch $K-1$ Parameter charakterisiert.
- Analog zum 2PL Modell für dichotome Items, lässt sich auch das PCM durch Aufnahme eines zusätzlichen Parameters β_i für jedes Item zum „Generalized Partial Credit Modell" (GPCM) erweitern.
- Allgemeine Modellgleichung:

$$P(X_{pi} = x_{pi} | \theta_p) = \frac{e^{\sum_{c=1}^{x_{pi}} \beta_i(\theta_p - \tau_{ic})}}{1 + \sum_{s=1}^{K-1} e^{\sum_{c=1}^{s} \beta_i(\theta_p - \tau_{ic})}}$$

- Wie im 2PL Modell kann der Parameter β_i auch im GPCM nur positive Werte annehmen ($\beta_i \in]0; +\infty[$).
- Wie im PCM gilt auch für die Modellgleichung des GPCM die Konvention:

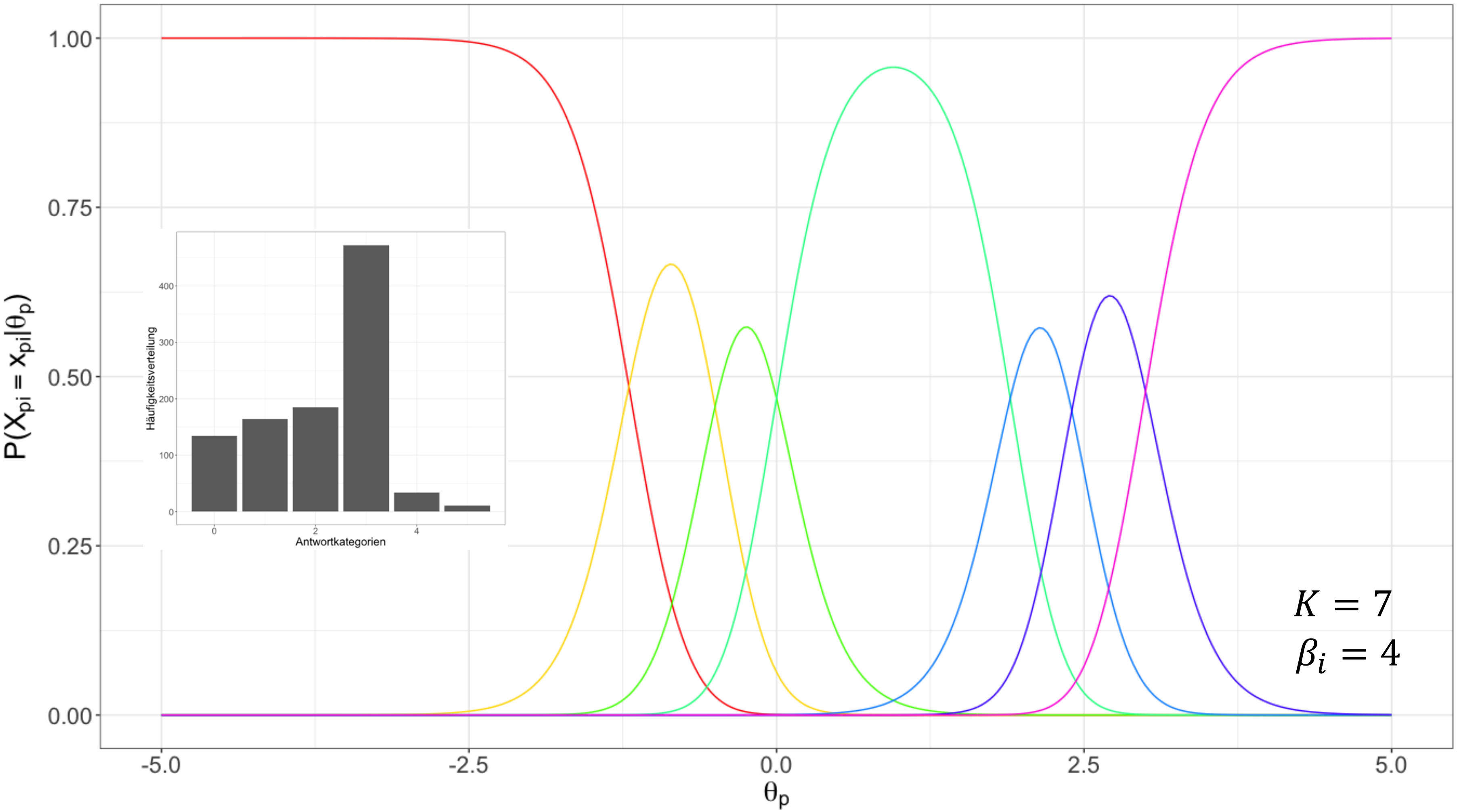
$$\sum_{c=1}^{0} (\theta_p - \tau_{ic}) = 0$$

Interpretation der Itemparameter



- Der Itemparameter β_i hängt mit der Steigung der CCCs von Item i zusammen.
- Je größer β_i, desto stärker verändern sich die Kategorienwahrscheinlichkeiten für Werte von θ_p nahe der τ_{ic} (desto „steiler“ sind die CCCs nahe der Schwellen)
- Wie im 2PL Modell wird β_i daher als „Diskrimination“ von Item i interpretiert.

Beispiel: CCCs für ein Item mit 7 Antwortkategorien



GPCM vs. PCM: Gemeinsamkeiten und Unterschiede

- Das GPCM stellt eine naheliegende Erweiterung des PCM dar. Durch die zusätzlichen Diskriminationsparameter ist das GPCM das flexiblere Testmodell und kann damit noch komplexere Zusammenhänge zwischen der latenten Variable und den manifesten Itemantworten abbilden.
  - Wie im PCM müssen auch im GPCM die Schwellen nicht geordnet sein
  - Wie im PCM können auch im GPCM für die latente Variable Regionen mit der höchsten Kategorienwahrscheinlichkeit grafisch dargestellt werden
  - Wie im PCM können auch im GPCM völlig analog ICCs berechnet werden
- Für das PCM gilt durch die Verwendung des dichotomen Raschmodells für die Modellierung der Schwellenwahrscheinlichkeiten die Eigenschaft der spezifischen Objektivität („ordinales Raschmodell“).
- Da im GPCM das 2PL Modell zur Modellierung der Schwellenwahrscheinlichkeit herangezogen wird, liegt hier keine spezifische Objektivität vor. Dies hat jedoch praktisch keine Auswirkungen und es gilt dieselbe Argumentation wie bei den Testmodellen für dichotome Items.

Zusammenfassung: Testmodelle für ordinale Items ($x_{pi} \in \{0,1,\ldots,K-1\}$)

- PCM („Ordinales Raschmodell“):

$$P\left(X_{pi} = x_{pi} \middle| \theta_p\right) = \frac{e^{\sum_{c=1}^{x_{pi}}(\theta_p - \tau_{ic})}}{1 + \sum_{s=1}^{K-1} e^{\sum_{c=1}^{s}(\theta_p - \tau_{ic})}}$$

- Items unterscheiden sich in ihrer „Schwierigkeit“
- Gegenstück zum 1PL Modell bei Items mit dichotomem Antwortformat

- GPCM:

$$P(X_{pi} = x_{pi} | \theta_p) = \frac{e^{\sum_{c=1}^{x_{pi}} \beta_i(\theta_p - \tau_{ic})}}{1 + \sum_{s=1}^{K-1} e^{\sum_{c=1}^{s} \beta_i(\theta_p - \tau_{ic})}}$$

- Items unterschieden sich in ihrer „Schwierigkeit“ und in ihrer „Diskrimination“
- Gegenstück zum 2PL Modell bei Items mit dichotomem Antwortformat

In beiden Modellen ergibt sich die „Schwierigkeit“ eines Items nicht mehr durch einen einzigen Parameter, sondern durch die Lage der $K-1$ Schwellen.