zu durchschauen sind, dass sie von vornherein von den Personen ausgeschlossen werden.
In diesem Fall könnte es sein, dass die echte Ratewahrscheinlichkeit für eine Aufgabe sogar höher ist als $\frac{1}{3}$. Deshalb ist es oft interessanter, die Ratewahrscheinlichkeit nicht vorzugeben, oder für alle Aufgaben mit der gleichen Anzahl von Antwortalternativen dieselbe Ratewahrscheinlichkeit anzunehmen, sondern die Ratewahrscheinlichkeit γ_j für jede Aufgabe einzeln zu schätzen. Dies ist allerdings nur möglich, wenn genug Personen die Aufgaben beantwortet haben.
Genauso wie das 2PL-Modell ein Spezialfall des 3PL-Modells ist, bei dem die Ratewahrscheinlichkeit 0 ist, kann man auch ein zwei-parametriges Modell als Spezialfall des 3PL-Modells konstruieren, das eine positive Ratewahrscheinlichkeit zulässt, bei dem aber wieder alle Aufgaben dieselbe Trennschärfe haben, wie beim Rasch-Modell – also ein Rasch-Modell mit Rate-Parameter(n).

3PL- und 2PL-Modelle können mithilfe des marginalen Maximum-Likelihood-Ansatzes z.B. mit dem R-Paket `ltm` (Rizopoulos, 2011) geschätzt werden.

5.3 Modelle mit mehrstufigen Antwortkategorien

Bisher haben wir immer den Fall betrachtet, dass Personen eine Aufgabe lösen (kodiert mit 1) oder nicht lösen (kodiert mit 0), bzw. einer Aussage zustimmen (1) oder nicht zustimmen (0). Die folgenden Modelle stellen eine Erweiterung dieser möglichen Antwortkategorien dar: Nun soll es möglich sein, bei jeder Aufgabe z.B. zwischen 0 und 3 Punkte zu erhalten.
In einem Leistungstest kann man sich z.B. vorstellen, dass eine Person null Punkte erhält, wenn sie die Aufgabe gar nicht oder völlig falsch bearbeitet hat, einen Punkt, wenn sie einen Teil der Lösung richtig hat usw.; drei Punkte erhält die Person nur, wenn sie die Lösung komplett richtig ausgeführt hat.
Auch bei Einstellungstests ist es üblich, die Personen auf einer ordinalen Skala eintragen zu lassen, inwieweit sie einer Aussage zustimmen (z.B. von „stimme überhaupt nicht zu“ bis „stimme voll zu“).
Die folgenden Modelle sind alle zur Auswertung von Tests mit solchen mehrstufigen Antwortkategorien konzipiert.

5.3.1 Das Partial-Credit-Modell

Beispielhaft soll hier das Partial-Credit-Modell von Masters (1982) ausführlicher vorgestellt werden, aber alle Modelle für mehrstufige Antwortkategorien folgen in ihrer Formulierung einem ähnlichen Prinzip: Anstatt einer gemeinsamen Modellgleichung, die die Lösungswahrscheinlichkeit für eine Person i und eine Aufgabe j beschreibt, braucht man jetzt, wo jede Aufgabe j mit einer von mehreren Antwortkategorien beantwortet werden kann, eine eigene Gleichung für jede Antwortkategorie c jeder Aufgabe j.

In unterschiedlichen Büchern und Artikeln wird das Partial-Credit-Modell oft mit unterschiedlichen Formeln dargestellt, die sich auf den ersten Blick nicht sonderlich ähnlich sehen. Deshalb werden im Folgenden die unterschiedlichen Darstellungsweisen zueinander in Beziehung gesetzt und erklärt. (Ausserdem scheint das Partial-Credit-Modell Tippfehler geradezu magisch anzuziehen. In fast allen Büchern und Artikeln dazu habe ich mindestens einen Fehler in einer Formel gefunden – und natürlich umso mehr versucht, selber keine zu machen. Aber Menschen die Bücher schreiben sind eben auch nur Menschen, d.h. wenn Sie in unterschiedlichen Büchern und Artikeln Darstellungen des Partial-Credit-Modells entdecken, die sich nicht genau ineinander überführen lassen, liegt es vermutlich nicht an Ihnen...)
Die Herleitung des Partial-Credit-Modells nach Masters (1982) folgt der Idee, dass die Wahrscheinlichkeit des Übergangs von einer Antwortkategorie $c-1$ zur nächsten Antwortkategorie c mithilfe des Rasch-Modells beschrieben wird. Genauer gesagt beschreibt das Partial-Credit-Modell die bedingte Wahrscheinlichkeit dafür, dass die Antwort in die höhere Kategorie c fällt, unter der Bedingung, dass die Antwort in eine der beiden Kategorien $c-1$ oder c fällt, durch ein Rasch-Modell:
Für eine Aufgabe j mit den Antwortkategorien $0, \ldots, m_j$ (bei der z.B. 0, 1, 2 oder 3 Punkte erzielt werden können, so dass es insgesamt $m_j + 1 = 4$ Kategorien gibt) ist die bedingte Wahrscheinlichkeit, dass die Antwort von Person i bei Aufgabe j in die Kategorie c fällt (unter der Bedingung, dass sie in eine der beiden Kategorien $c-1$ oder c fällt):

$$P(u_{ij} = c \,|u_{ij} \in \{c-1, c\}, \theta_i, \delta_{jc}) = \frac{e^{\,\theta_i - \delta_{jc}}}{1 + e^{\,\theta_i - \delta_{jc}}}$$

In dieser Formel gibt es einen eigenen Parameter δ_{jc} für jede Kategorie c der Aufgabe j. Ansonsten entspricht die Form der des Rasch-Modells aus den früheren Kapiteln.
Aus dieser bedingten Wahrscheinlichkeit leitete Masters (1982) die unbedingte Wahrscheinlichkeit dafür ab, dass die Antwort in die Kategorie c fällt:

$$P(u_{ij} = c \,|\theta_i, \delta_{j1}, \ldots, \delta_{jm_j}) = \frac{e^{\,\sum_{k=0}^{c}(\theta_i - \delta_{jk})}}{\sum_{l=0}^{m_j} e^{\,\sum_{k=0}^{l}(\theta_i - \delta_{jk})}}$$

wobei zur Berechnung festgelegt wird, dass $\sum_{k=0}^{0} (\theta_i - \delta_{jk}) = 0$ und entsprechend $\sum_{k=0}^{l} (\theta_i - \delta_{jk}) = \sum_{k=1}^{l} (\theta_i - \delta_{jk})$.
Diese Modellgleichung ist komplizierter als die für das einfache Rasch-Modell, weil über die verschiedenen Kategorien der Aufgabe Summen gebildet werden müssen, um die gesuchte Wahrscheinlichkeit zu berechnen. Ausserdem gilt die Gleichung nicht für eine ganze Aufgabe, wie im Rasch-Modell, sondern nur für eine einzelne Kategorie einer Aufgabe.
Entsprechend gibt es im Partial-Credit-Modell auch nicht nur eine ICC pro Aufgabe, sondern es gibt so viele wie die Aufgabe Kategorien hat; diese Kurven nennt man daher auch „Category Characteristic Curves“ oder CCCs. Die

CCCs des Partial-Credit-Modells für eine Aufgabe mit vier Antwortkategorien und den entsprechenden Parametern δ_{jk} sind in Abbildung 5.3 dargestellt.
Die einzelnen Kurven entsprechen dabei den Wahrscheinlichkeiten für die einzelnen Antwortkategorien: Von links nach rechts betrachtet ist die Wahrscheinlichkeit für die erste Antwortkategorie (0 Punkte) für niedrige Werte der Fähigkeit hoch und sinkt für höhere Werte der Fähigkeit ab.
Im Gegensatz dazu sind die Wahrscheinlichkeiten für die mittleren Kategorien (1 und 2 Punkte) zunächst niedrig, bei dem der Kategorie entsprechenden Fähigkeitsniveau hoch und darüber hinaus wieder niedrig. Die Wahrscheinlichkeit für die letzte Kategorie (3 Punkte) hingegen ist zunächst niedrig und steigt für höhere Werte der Fähigkeit an.

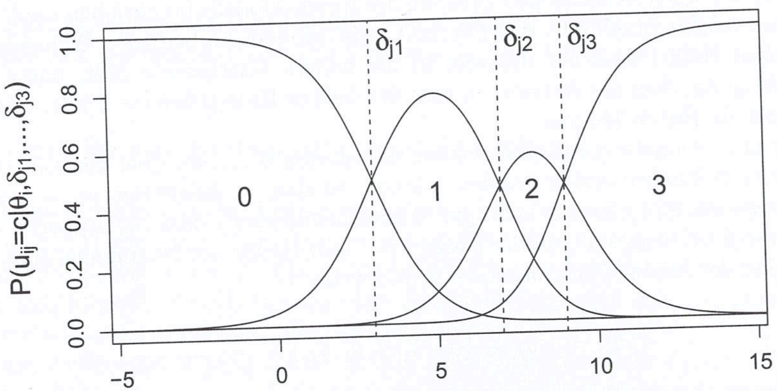

Abb. 5.3. CCCs für das Partial-Credit-Modell mit eingezeichneten δ-Parametern.

Für eine Person mit der Fähigkeit 5 kann man z.B. in Abbildung 5.3 ablesen, dass sie mit der höchsten Wahrscheinlichkeit in die zweite Kategorie fällt (d.h. 1 Punkt erzielt), während die Wahrscheinlichkeiten für die benachbarten Kategorien (0 und 2 Punkte) deutlich geringer sind und die Wahrscheinlichkeit für die höchste Kategorie (3 Punkte) fast bei 0 liegt.
Die Wahrscheinlichkeiten für alle vier Kategorien zusammen addieren sich dabei für jedes Fähigkeitsniveau zum Wert 1 auf.
Neben den Wahrscheinlichkeiten für die einzelnen Antwortkategorien sind in Abbildung 5.3 auch die Schwellenwerte auf der Fähigkeits-Skala (als gestrichelte Linien) eingezeichnet, die die Schnittpunkte zwischen den Kurven für die einzelnen Kategorien markieren: Eine Person mit der Fähigkeit 5 liegt z.B. zwischen den beiden Schwellenwerten, die die zweite Kategorie von der ersten und dritten Kategorie trennen. Für alle Personen zwischen diesen beiden

Schwellenwerten hat die zweite Kategorie die höchste Wahrscheinlichkeit für eine einzelne Kategorie.
Die Lage der Schwellenwerte entspricht genau den Parametern δ_{jk} aus der obigen Modellgleichung. Sie werden deshalb auch als Schwellenparameter bezeichnet. Dabei liegt z.B. δ_{j1} am Schnittpunkt zwischen Kategorie 0 und der nächst höheren Kategorie 1, δ_{j2} am Schnittpunkt zwischen Kategorie 1 und der nächst höheren Kategorie 2 etc., d.h. die Anzahl der Schwellenparameter ist um 1 kleiner als die Anzahl der Kategorien.
Die Schwellenparameter können in zwei andere Arten von Parametern umgerechnet werden, die dieselbe Information enthalten, aber anders zu interpretieren sind. (Die Umrechnung von einer Art von Parametern in eine andere nennt man Reparametrisierung. Dabei müssen die Parameterwerte nicht neu aus den Daten geschätzt werden, sondern es gibt eine eindeutige Rechenregel, mit der man zwischen den unterschiedlichen Parametrisierungen hin- und her-rechnen kann. Dadurch ändert sich aber auch das Aussehen der Modellgleichung, so dass man das Modell anhand der neuen Formel auf den ersten Blick manchmal nur schwer wiedererkennen kann.)
Eine häufig verwendete Parametrisierung des Partial-Credit-Modells ergibt sich aus der Darstellung mit den Schwellenparametern durch Aufsummieren von $\lambda_{jc} = \sum_{k=1}^{c} \delta_{jk}$:

$$\begin{aligned}
P(u_{ij} = c\,|\theta_i, \delta_{j1}, \ldots, \delta_{jm_j}) &= \frac{e^{\sum_{k=0}^{c}(\theta_i - \delta_{jk})}}{\sum_{l=0}^{m_j} e^{\sum_{k=0}^{l}(\theta_i - \delta_{jk})}} \\
&\overset{①}{=} \frac{e^{c\cdot\theta_i - \sum_{k=1}^{c}\delta_{jk}}}{\sum_{l=0}^{m_j} e^{l\cdot\theta_i - \sum_{k=1}^{l}\delta_{jk}}} \\
P(u_{ij} = c\,|\theta_i, \lambda_{j1}, \ldots, \lambda_{jm_j}) &\overset{②}{=} \frac{e^{c\cdot\theta_i - \lambda_{jc}}}{\sum_{l=0}^{m_j} e^{l\cdot\theta_i - \lambda_{jl}}}
\end{aligned}$$

wobei $\lambda_{j0} = 0$. Die Darstellung der Modellgleichung in der letzten Zeile findet sich in vielen Büchern (wenn auch natürlich oft mit anderen griechischen Buchstaben anstelle von θ_i und λ_{jc}), weil die Formel weniger Summenzeichen enthält und dadurch einfacher aussieht. Zur Interpretation sind die kumulierten Schwellenparameter λ_{jc} aber nicht sonderlich gut geeignet.
Eine weitere mögliche Parametrisierung des Partial-Credit-Modells, die wieder relativ einfach zu interpretieren ist, ergibt sich aus den Schwellenparametern durch die Umrechnung in einen zentralen Lageparameter $\bar{\delta}_j = \sum_{k=1}^{m_j} \delta_{jk}/m_j$, dem Mittelwert der ursprünglichen Schwellenparameter, und den m_j Abweichungen der einzelnen Schwellenparameter von diesem Mittelwert, $\tau_{jk} = \bar{\delta}_j - \delta_{jk}$ (wobei von den m_j Differenzen τ_{jk} nur $m_j - 1$ frei variieren können):

① wegen $\sum_{k=0}^{l}(\theta_i - \delta_{jk}) = \sum_{k=1}^{l}(\theta_i - \delta_{jk})$ (wurde oben bei der Darstellung der Modellgleichung mit Schwellenparametern festgelegt) ist:
$\sum_{k=0}^{c}(\theta_i - \delta_{jk}) = \sum_{k=1}^{c}(\theta_i - \delta_{jk}) = \sum_{k=1}^{c}\theta_i - \sum_{k=1}^{c}\delta_{jk} = c\cdot\theta_i - \sum_{k=1}^{c}\delta_{jk}$
② $\sum_{k=1}^{c}\delta_{jk} = \lambda_{jc}$

$$P(u_{ij} = c\,|\,\theta_i, \bar{\delta}_j, \tau_{j1}, \dots, \tau_{jm_j}) = \frac{e^{\sum_{k=0}^{c}\left(\theta_i - (\bar{\delta}_j - \tau_{jk})\right)}}{\sum_{l=0}^{m_j} e^{\sum_{k=0}^{l}\left(\theta_i - (\bar{\delta}_j - \tau_{jk})\right)}}$$

Die CCCs des Partial-Credit-Modells in der Darstellung mit dem zentralen Lageparameter $\bar{\delta}_j$ und den Differenzen τ_{jk} sind in Abbildung 5.4 dargestellt.

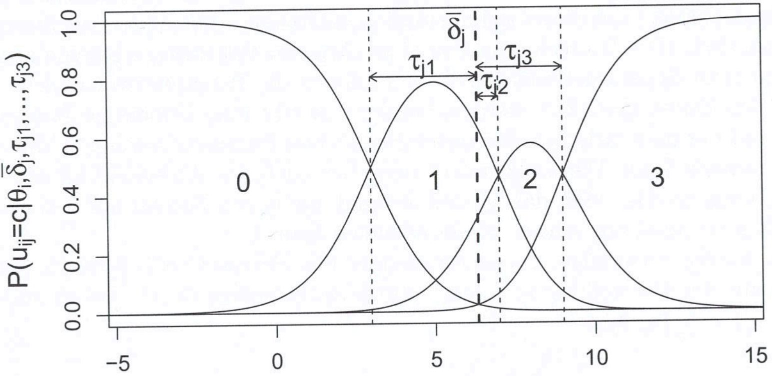

Abb. 5.4. CCCs für das Partial-Credit-Modell mit eingezeichneten $\bar{\delta}$- und τ-Parametern.

Ein Vorteil dieser Parametrisierung ist, dass sich das Partial-Credit-Modell in dieser Form sehr gut mit dem Rating-Scale-Modell vergleichen lässt, das im nächsten Abschnitt vorgestellt wird. Ein weiterer (vermeintlicher) Vorteil ist, dass der Parameter $\bar{\delta}_j$ die zentrale Lage der Aufgabe auf dem Fähigkeits- bzw. Schwierigkeits-Kontinuum ausdrückt, und man damit die Lage der Aufgabe mit einem einzelnen Wert beschreiben kann. Genau darin liegt aber auch das Risiko für Fehlinterpretationen bei dieser Parametrisierung: Eine Aufgabe mit einem höheren $\bar{\delta}_j$ kann man nicht einfach als schwieriger bezeichnen als eine Aufgabe mit einem niedrigeren $\bar{\delta}_j$ (wie es im Rasch-Modell für Aufgaben mit unterschiedlicher Schwierigkeit β_j möglich ist), weil bei unterschiedlichen Aufgaben die Schwellenparameter zwischen den Kategorien unterschiedlich weit auseinanderliegen können, wie in Abbildung 5.5 dargestellt.

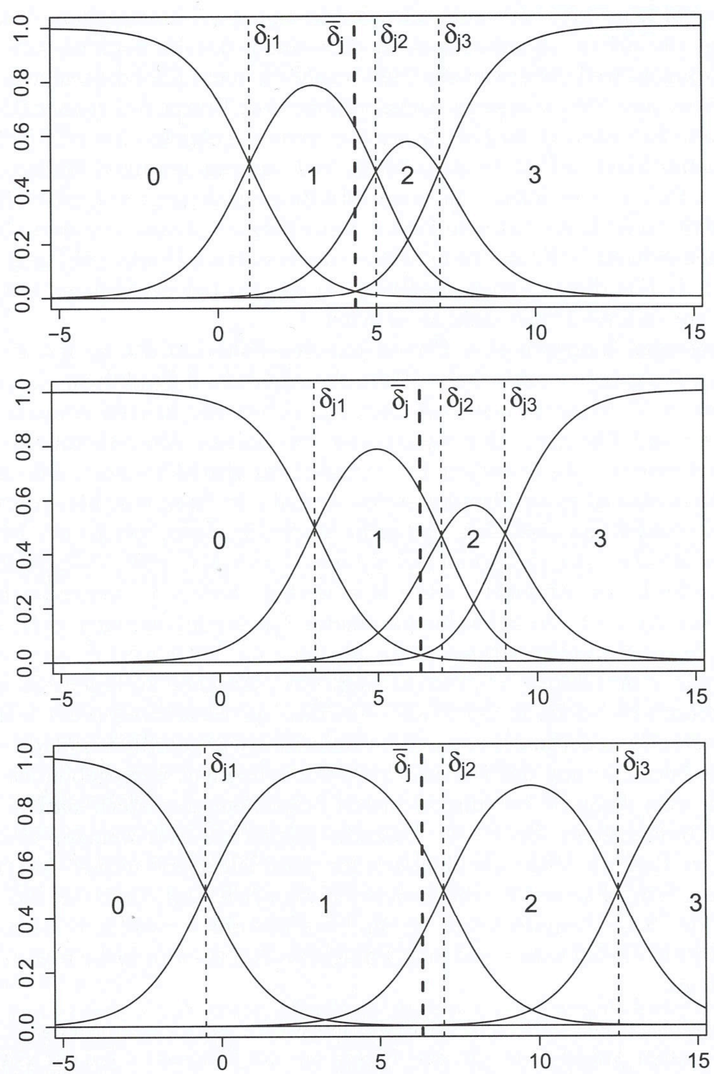

Abb. 5.5. CCCs für drei Aufgaben mit unterschiedlichen δ-Parametern: Vergleicht man die erste mit der zweiten Aufgabe, so liegen sowohl der zentrale Lageparameter $\bar{\delta}_j$ als auch jeder einzelne Schwellenparameter δ_{jk} bei der ersten Aufgabe weiter links als bei der zweiten Aufgabe. Die beiden Aufgaben lassen sich also in eine eindeutige Reihenfolge bringen. Vergleicht man hingegen die erste mit der dritten Aufgabe, so liegt der zentrale Lageparameter $\bar{\delta}_j$ zwar auch bei der dritten Aufgabe weiter rechts als bei der ersten Aufgabe, die Schwellenparameter zwischen den Kategorien liegen aber auch weiter auseinander. Dadurch lässt sich keine eindeutige Reihenfolge zwischen den beiden Aufgaben herstellen.

Auch formal kann man sich anhand von Abbildung 5.5 klarmachen, dass die erste und die dritte Aufgabe nicht in eine eindeutige Reihenfolge gebracht werden können, weil die erwartete Punktezahl[①] je nach Fähigkeitsniveau mal für die eine, mal für die andere Aufgabe höher ist: Betrachtet man z.B. eine Person mit Fähigkeit 0, so hat sie bei der ersten Aufgabe eine relativ hohe Wahrscheinlichkeit, 0 Punkte zu erzielen, und nur eine geringen Wahrscheinlichkeit, 1 Punkt zu erzielen – die erwartete Punktezahl liegt also näher bei 0. Bei der dritten Aufgabe hat eine Person mit Fähigkeit 0 hingegen eine höhere Wahrscheinlichkeit, 1 Punkt zu erzielen – die erwartete Punktezahl liegt also näher bei 1. Für dieses Fähigkeitsniveau ist es also bei der dritten Aufgabe leichter, eine höhere Punktezahl zu erzielen.

Betrachtet man hingegen eine Person mit der Fähigkeit 10, so hat sie bei der ersten Aufgabe eine sehr hohe Wahrscheinlichkeit, 3 Punkte zu erzielen – die erwartete Punktezahl liegt also nahe bei 3. Bei der dritten Aufgabe hat eine Person mit Fähigkeit 10 hingegen eine viel höhere Wahrscheinlichkeit, 2 Punkte zu erzielen – die erwartete Punktezahl liegt also näher bei 2. Für dieses Fähigkeitsniveau ist es also bei der ersten Aufgabe leichter, eine höhere Punktezahl zu erzielen. Es lässt sich also nicht eindeutig sagen, welche der beiden Aufgaben leichter ist (wie formal bei Sijtsma & Hemker, 1998, dargestellt).

Beim Vergleich von Aufgaben sollte man deshalb neben $\bar{\delta}_j$ immer auch die Differenzen τ_{jk} oder die Schwellenparameter δ_{jk} berücksichtigen (z.B. können sowohl die Schwellenparameter δ_{jk} als auch ihr Mittelwert $\bar{\delta}_j$ angegeben werden, wie in Abbildung 5.5, um die Lage einer Aufgabe zu beschreiben).

Häufig kommt es jedoch in der Praxis vor, dass die Schätzungen der Schwellenparameter δ_{jk} nicht in der erwarteten Reihenfolge angeordnet sind, da diese durch die Formulierung des Partial-Credit-Modells nicht vorgegeben ist. Die CCCs für eine Aufgabe mit ungeordneten Schwellenparametern sind in Abbildung 5.6 dargestellt. Bei dieser Aufgabe erzielen nur sehr wenige Personen 1 Punkt, so dass die Wahrscheinlichkeit für diese Kategorie immer unterhalb der Wahrscheinlichkeiten für die anderen Kategorien liegt, und sich die Reihenfolge der Schwellenparameter δ_{j1} und δ_{j2} umkehrt – wodurch auch ihre Interpretation als Schwellen auf dem Fähigkeits-Kontinuum nicht mehr sinnvoll erscheint.

① Die erwartete Punktezahl (d.h. der Erwartungswert) lässt sich einfach berechnen: Bei der ersten Aufgabe hat z.B. eine Person mit der Fähigkeit 0 (an der x-Achse ablesen) eine Wahrscheinlichkeit von ca. 73% (an der y-Achse ablesen – wobei man das an Abbildung 5.5 nicht so genau sehen kann wie es hier berechnet ist), 0 Punkte zu erzielen, eine Wahrscheinlichkeit von ca. 27%, 1 Punkt zu erzielen, und eine Wahrscheinlichkeit von je ca. 0%, 2 oder 3 Punkte zu erzielen. Die erwartete Punktezahl für eine Person mit der Fähigkeit 0 bei der ersten Aufgabe ist also $E(U_{i1}|\theta_i = 0) = 0,73 \cdot 0 + 0,27 \cdot 1 + 0 \cdot 2 + 0 \cdot 3 = 0,27$, u.s.w.:

	Fähigkeit 0	Fähigkeit 10
Aufgabe 1	$E(U_{i1}\|\theta_i = 0) = 0,27$	$E(U_{i1}\|\theta_i = 10) = 2,95$
Aufgabe 3	$E(U_{i3}\|\theta_i = 0) = 0,62$	$E(U_{i3}\|\theta_i = 10) = 2,03$

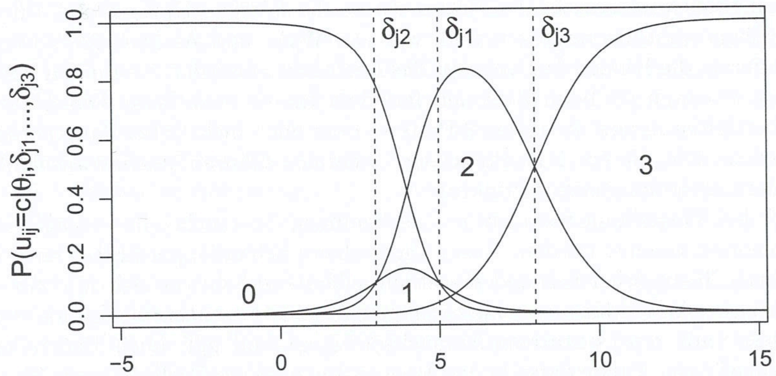

Abb. 5.6. CCCs für eine Aufgabe mit ungeordneten Schwellenparametern: δ_{j1}, der Schnittpunkt der Kurven für Kategorie 0 und Kategorie 1, liegt weiter rechts als δ_{j2}, der Schnittpunkt der Kurven für Kategorie 1 und Kategorie 2.

Über die formale und inhaltliche Bedeutung der ungeordneten Schwellenparameter gehen die Meinungen in der Fachwelt auseinander: Andrich (2010 und in früheren Arbeiten) wertet ihr Auftreten als Indiz dafür, dass die Antwortkategorien der betroffenen Aufgabe nicht in der intendierten, ordinalen Weise verwendet werden. Von diesem Standpunkt aus betrachtet, der von vielen anderen Autoren geteilt bzw. übernommen wurde (vgl. z.B. Rost, 2004), bietet das Partial-Credit-Modell also eine Möglichkeit, die intendierte Ordnung der Kategorien empirisch zu überprüfen. Adams, Wu und Wilson (2012) argumentieren jedoch gegen diese Sichtweise und führen zwei Definitionen von Ordnung an, die das Partial-Credit-Modell auch bei ungeordneten Schwellenparametern erfüllt.

Einig sind sich die Autoren allerdings darüber, dass ungeordnete Schwellenparameter darauf hinweisen bzw. daraus resultieren, dass die betroffenen Antwortkategorien in den Daten nur selten auftreten. Ursache hierfür ist oft eine ungeeignete Formulierung der Antwortkategorien bzw. des vorgegebenen Bewertungsschemas, wie z.B. Van Wyke und Andrich (2006) an der Beurteilung von geometrischen Zeichnungen von Schulkindern der dritten Klasse illustrieren:

Die Aufgabenstellung lautete, in einer Zeichnung mit mehreren Gänseblümchen einige davon mithilfe des Lineals zu einem Rechteck zu verbinden. Das Bewertungsschema sah vor, 0 Punkte zu vergeben, wenn die Antwort völlig unzureichend war, 1 Punkt, wenn zwar die richtigen Gänseblümchen ausgewählt wurden, das Rechteck aber schlecht gezeichnet war (z.B. mit ungeraden

oder nicht parallelen Seiten) und 2 Punkte, wenn das Rechteck komplett richtig eingezeichnet war. In der Praxis wurde die mittlere Kategorie (1 Punkt) allerdings wohl nie vergeben, weil, wie Van Wyke und Andrich (2006) argumentieren, die Kinder das Konzept des Rechtecks entweder verstanden hatten (dann machten sie beim Zeichnen mit dem Lineal auch keine Fehler mehr, und erzielten immer die vollen 2 Punkte) oder eben nicht (aber dann konnten sie auch nicht die für 1 Punkt nötigen richtigen Gänseblümchen auswählen, sondern erzielten gleich 0 Punkte).
Auch bei Einstellungstests kann es vorkommen, dass nicht alle vorgegebenen Kategorien genutzt werden. Wenn ungeordnete Schwellenparameter bzw. ungenutzte Kategorien vermieden werden sollen empfiehlt es sich also für die praktische Anwendung von mehrstufigen Antwortformaten besonders, einen Pretest (mit repräsentativen Versuchspersonen und ggf. auch Beurteilern) durchzuführen. Fallen dabei in der Auswertung mit dem Partial-Credit-Modell bereits ungeordnete Schwellenparameter auf, können die Antwortkategorien bzw. das Bewertungsschema noch entsprechend angepasst werden.
Wurden die Daten hingegen bereits erhoben, wird als pragmatische Lösung meist empfohlen, die betroffene Kategorie mit einer oder mehreren benachbarten Kategorien zusammenzufassen. Dazu müssen zunächst die Daten umkodiert, und dann das Partial-Credit-Modell neu geschätzt werden. Adams et al. (2012) raten allerdings vom routinemäßigen Zusammenlegen von Antwortkategorien nur aufgrund von ungeordneten Schwellenparametern ab.
Das Partial-Credit-Modell enthält das Rasch-Modell als Spezialfall mit zwei Antwortkategorien. Auch beim Partial-Credit-Modell haben alle Aufgaben dieselbe Trennschärfe und es existieren suffiziente Statistiken für alle Parameter, so dass es durch bedingte Maximum-Likelihood-Schätzung z.B. mit dem R-Paket `eRm` (Mair et al., 2011) geschätzt werden kann.
Eine Vielzahl von weiteren Modellen für mehrstufige Antwortkategorien wurde z.T. schon vor dem Partial-Credit-Modell vorgeschlagen. Die bekanntesten sollen im Folgenden noch kurz aufgeführt werden.

5.3.2 Das Rating-Scale-Modell

Das Rating-Scale-Modell von Andrich (1978) stellt einen Spezialfall des Partial-Credit-Modells dar, bei dem alle Aufgaben das gleiche Antwortformat aufweisen, d.h. insbesondere die gleiche Anzahl von Antwortkategorien.
Auch die Abstände zwischen den Schwellenparametern sind im Rating-Scale-Modell für alle Aufgaben gleich – im Gegensatz zum Partial-Credit-Modell, bei dem sich die Anzahl der Kategorien sowie deren Breite von Aufgabe zu Aufgabe unterscheiden können. Dies sieht man am besten in der Darstellung mit den Parametern $\bar{\delta}_j$ und $\tau_1, \ldots, \tau_m$:

$$P(u_{ij} = c\,|\theta_i, \bar{\delta}_j, \tau_1, \ldots, \tau_m) = \frac{e^{\sum_{k=0}^{c}\left(\theta_i - (\bar{\delta}_j - \tau_k)\right)}}{\sum_{l=0}^{m} e^{\sum_{k=0}^{l}\left(\theta_i - (\bar{\delta}_j - \tau_k)\right)}}$$

Jede Aufgabe j erhält dabei ihren eigenen zentralen Lageparameter $\bar{\delta}_j$. Sowohl die Anzahl der Kategorien $0, \ldots, m$, als auch die Abweichungen der Schwellen vom zentralen Lageparameter $\tau_1, \ldots, \tau_m$ sind aber nun für alle Aufgaben gleich (und haben deshalb auch keinen Index mehr für die Aufgabe j, wie beim Partial-Credit-Modell auf S. 58 oben). D.h. die Aufgaben können je nach ihrem Wert für $\bar{\delta}_j$ weiter rechts oder weiter links liegen, die Abstände zwischen den Kategorien sind aber für jede Aufgabe gleich (wie bei den oberen beiden Aufgaben in Abbildung 5.5). In diesem Sinne lassen sich beim Rating-Scale-Modell die Aufgaben auch eindeutig bzgl. ihrer Schwierigkeit anordnen (Sijtsma & Hemker, 1998).
Auch im Rating-Scale-Modell kann es theoretisch passieren, dass sich die Reihenfolge der Schwellenparameter umkehrt, weil einzelne Kategorien nur selten genutzt werden. Da die Schätzung der Schwellenparameter (bzw. der Differenzen $\tau_1, \ldots, \tau_m$) allerdings über alle Aufgaben hinweg erfolgt, wirkt sich eine Abweichung in einzelnen Aufgaben meist nicht merklich auf die aus allen Aufgaben geschätzten Schwellenparameter aus – wodurch ungeeignete Antwort- bzw. Bewertungsschemata ggf. unbemerkt bleiben können.
Die Entscheidung, ob ein Partial-Credit-Modell oder ein Rating-Scale-Modell zur Auswertung eines Tests verwendet werden soll, ist nicht immer ganz einfach. Wenn sich die Anzahl der Kategorien zwischen den Aufgaben unterscheidet, oder sich bei der Anpassung des Partial-Credit-Modells zeigt, dass die Breite der Kategorien von Aufgabe zu Aufgabe sehr unterschiedlich ist, sollte das Partial-Credit-Modell verwendet werden. Andererseits können inhaltliche Überlegungen für die Verwendung des Rating-Scale-Modells sprechen, z.B. wenn für alle Aufgaben dasselbe Antwortformat vorgegeben wurde (z.B. eine 5-stufige Likert-Skala mit denselben Labels von „stimme überhaupt nicht zu“ bis „stimme voll zu“).
Eine weitere empirische Entscheidungshilfe können Informationskriterien (wie das Akaike Information Criterion AIC und das Bayesian Information Criterion BIC; vgl. z.B. Tutz, 2000; Fahrmeir, Kneib & Lang, 2007) liefern, die sowohl die Güte der Anpassung an die Daten als auch die Anzahl der Parameter (d.h. der Komplexität des Modells) berücksichtigen. Oft kommt es dabei allerdings zu Widersprüchen zwischen dem AIC und dem BIC, weil das BIC die Anzahl der Parameter stärker bestraft und deshalb meist das sparsamere Modell bevorzugt (wie z.B. bei Baghaei, 2010).
Das Rating-Scale-Modell kann, wie das Partial-Credit-Modell, z.B. mit dem R-Paket `eRm` (Mair et al., 2011) geschätzt werden, das auch Informationskriterien zum Vergleich von Modellen ausgibt.

5.3.3 Das Graded-Response-Modell

Ein weiteres bekanntes Modell für mehrstufige Antwortkategorien ist das Graded-Response-Modell (Samejima, 1969). Bei diesem Modell handelt es sich um ein sogenanntes kumulatives Modell (vgl. Tutz, 2000). Modelliert wird nämlich die kumulierte Wahrscheinlichkeit dafür, dass eine Person mindestens