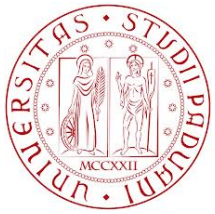

Università degli Studi di Padova

Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE

INDIRIZZO DI PSICOLOGIA SOCIALE E DELLA PERSONALITÀ

XXVII CICLO

Sviluppo e Applicazioni di Modelli Formali per la Valutazione Adattiva della Conoscenza e dell'Apprendimento nell'Ambito della Knowledge Space Theory.

Direttore della Scuola: Ch.mo Prof. Francesca Peressotti

Supervisore: Ch.mo Prof. Luca Stefanutti

Dottorando: Debora de Chiusole

Riassunto

Le cinque ricerche che si presentano in questa tesi si sviluppano entro la knowledge space theory, una teoria matematica recente che fornisce un importante quadro di riferimento formale per lo sviluppo di sistemi computerizzati web-based che abbiano l'obiettivo di *valutare la conoscenza e l'apprendimento degli individui*. La nozione al centro dell'intera teoria è quella di *stato di conoscenza*, cioè l'insieme dei problemi che uno studente è capace di risolvere, in un certo dominio di conoscenza. La collezione di tutti gli stati di conoscenza osservabili in una popolazione di studenti costituisce una *struttura di conoscenza*. Le strutture di conoscenza sono un modello deterministico teorico dell'organizzazione della conoscenza all'interno di un particolare dominio. La loro validazione empirica è resa possibile grazie alla verifica probabilistica della loro plausibilità. Il *basic local independence model* (BLIM) è un modello probabilistico che è stato sviluppato a questo scopo. Nonostante sia il modello più utilizzato nella KST, problemi relativi alla sua applicabilità rimanevano ancora aperti. L'obiettivo generale delle prime tre ricerche che si presentano in questa tesi, è stato quello di risolvere questi problemi per conferire una maggiore validità alle applicazioni empiriche del modello. Nella KST, la nozione di stato di conoscenza non fornisce alcun tipo di interpretazione cognitiva. Invece, nella *competence-based* KST (CbKST) l'obiettivo principale della valutazione diviene quello individuare lo *stato di competenza* dello studente, ovvero l'insieme delle abilità che possiede. Le altre due ricerche che si presentano nella tesi si collocano all'interno di questo quadro teorico. Esse hanno avuto l'obiettivo di colmare alcune mancanze relative della CbKST, una di tipo probabilistico e l'altra di tipo deterministico.

Abstract

The five studies presented in this thesis have been carried out in the area of knowledge space theory (KST), a recent mathematical theory providing an important framework for the formal development of computerized web-based systems aimed at *assessing individual knowledge and learning*. The basic concept at the core of the entire theory is that of *knowledge state*, that is the set of problems that a student is able to solve, in a certain filed of knowledge. The collection of all knowledge states that occurs in a population of students is called the *knowledge structure*. A knowledge structure is a deterministic model of the organization of knowledge in a particular domain. Its empirical validation is possible by a probabilistic assessment of its plausibility. The *basic local independence model* (BLIM) is a probabilistic model developed to this aim. Despite it is the most widely used model in KST, issues relating its applicability were open. The overall objective of the first three studies presented in this thesis was to solve some of these problems, in order to improve the validity of empirical applications of the model. In the KST framework, the notion of knowledge state does not provide cognitive interpretations. Instead, in the *competence-based* KST (CbKST) the main objective of the assessment becomes that of identifying the *competence state* of a student, which is the set of skills she masters. The other two studies that are introduced were developed within this extended theoretical framework. The general aim was to fill some gaps regarding both the probabilistic and the deterministic levels of CbKST's models.

Indice

Introduzione

Il denominatore comune delle ricerche che si presentano in questa tesi è la *Knoweldge Space Theory* (KST). La KST è una teoria matematica sviluppata a cavallo tra gli anni '80 e '90 grazie alla collaborazione dello psicologo Jean-Claude Falmagne con il matematico Jean-Paul Doignon. Potrebbe sembrare strano che due ricercatori con una formazione così diversa abbiano deciso di orientare i loro interessi verso un obiettivo comune, ovvero quello della valutazione della conoscenza degli individui. Eppure, fin da quando la psicologia è diventata una disciplina scientifica (per convenzione dal 1879, con l'apertura del laboratorio di Wundt a Leipzig), la matematica è stata uno strumento a disposizione di tutti quei ricercatori interessati a studiare i fenomeni e i processi psicologici. Tale connubio è diventato negli anni così forte da diventare una vera e propria branca della psicologia, ovvero la *psicologia matematica*. E' proprio entro questa disciplina che si sviluppano la KST e i contributi della presente tesi.

Prima di illustrare più da vicino le caratteristiche di questa teoria, nonché gli obiettivi della tesi, si vuole fare una piccola premessa sui vantaggi che derivano dall'applicazione del *modellamento formale*, lo strumento di indagine elettivo della psicologia matematica. Con il modellamento formale le ipotesi di ricerca vengono tradotte in una serie di assunzioni esplicite che portano alla costruzione di un modello matematico. Attraverso i metodi opportuni, tipicamente di tipo probabilistico, il modello viene applicato a dati empirici, che possono confermarlo o rifiutarlo. Nel caso in cui il modello sia validato, si ha disposizione una spiegazione estremamente precisa e rigorosa del

fenomeno in esame. Questo approccio si contrappone, ad esempio, a quello della verifica della significatività statistica dell'ipotesi nulla (NHST, dall'inglese null hypothesis significance testing), che ormai da decenni è sottoposto a innumerevoli critiche (si vedano ad esempio Pastore (2009); Wagenmakers (2007); Anderson, Burnham, e Thompson (2000); Nickerson (2000); Gill (1999)).

Una delle particolarità più sorprendenti della KST è che racchiude in un *quadro formale unificato* le nozioni più attuali circa le teorie psicologiche sulla valutazione della conoscenza e sull'apprendimento degli individui. Come si capirà nel Capitolo 1, concetti come *valutazione formativa*, *valutazione adattiva* e *apprendimento graduale* e *personalizzato*, sono infatti al tempo stesso fonte d'ispirazione e obiettivo da raggiungere dalle applicazioni della KST.

Vi è poi una novità assoluta circa gli strumenti formali di cui fa uso per rappresentare e valutare la conoscenza degli individui. Ci si riferisce alla *matematica discreta* e *combinatoria*. Questa scelta è profondamente legata alla teoria della misurazione in psicologia (Krantz, Luce, Suppes, & Tversky, 1971). In particolare, alla convinzione che le variabili psicologiche, come ad esempio la conoscenza, non siano per loro natura di tipo continuo e quindi "misurabili" attraverso una quantificazione numerica. Sebbene il pensiero come quello di Kelvin (1889): *"If you cannot measure it, then it is not science"*, abbia influenzato la ricerca scientifica in psicologia per un lungo periodo, gli strumenti attuali permettono di operare a un *livello non-numerico* (rispettando la natura delle variabili che si intende studiare), ciò nonostante scientifico.

Il passaggio da una misurazione numerica a una non-numerica è reso possibile dallo sviluppo tecnologico e in particolare dalla potenza di calcolo raggiunta al giorno d'oggi dalle macchine. E' proprio quest'ultimo aspetto ad essere un'altra delle caratteristiche essenziali della KST. Il rigore formale sul quale si sviluppa permette infatti una traduzione del tutto naturale dei

suoi concetti, dal linguaggio matematico a quello informatico. Lo sviluppo di una *macchina efficiente* per la valutazione della conoscenza diviene dunque non solo auspicabile ma anche possibile. Fra le applicazioni della KST che hanno avuto fino ad oggi più successo, si trova infatti un sistema intelligente web-based, chiamato `Aleks`, utilizzato da centinaia di migliaia di studenti negli USA per la valutazione della conoscenza nell'ambito della matematica, dell'economia, delle chimica e della statistica.

Emerge dunque un profilo della KST assolutamente *multidisciplinare*, che fa uso delle più recenti teorie nell'ambito della psicologia della valutazione e dell'apprendimento, della matematica discreta applicata alla psicologia, delle scienze informatiche e della psicologia matematica.

La prima parte della tesi è dunque interamente dedicata alla presentazione della KST, che costituisce la teoria di riferimento dei contributi scientifici presentati successivamente. Sebbene questa prima parte sia di fatto una presentazione dello *stato dell'arte* della KST, si vuole comunque sottolineare che si tratta di un contributo originale, che cerca di unificare i risultati scientifici riportati in svariati articoli e monografie, in una sintesi, seppur breve, il più completa possibile. Si è deciso di suddividere questa parte in due capitoli separati. Il Capitolo 1 riassume gli aspetti formali deterministici e probabilistici legati all'approccio tradizionale della KST, chiamato *comportamentale*, all'interno del quale confluiscono la gran parte dei contributi scientifici della teoria. Il Capitolo 2 riassume invece gli aspetti legati a un approccio più recente, chiamato *competence based* KST (CbKST), che cerca di collegare la base comportamentale della teoria con quella *cognitiva*, legata alla diagnosi delle abilità degli individui.

E' proprio all'interno di questo quadro teorico che si sviluppano le ricerche illustrate nella seconda parte della tesi. Nonostante abbia trent'anni, la KST è una teoria matematica giovane e come tale, non è affatto compiuta. Per quanto riguarda l'approccio tradizionale della teoria, se i suoi aspetti deterministici costituiscono un quadro formale pressoché completo, lo stesso non

si può dire per quanto riguarda gli aspetti probabilistici, ovvero quell'insieme di procedure e modelli che consentono l'applicazione della KST a contesti reali. Per quanto riguarda invece l'approccio più recente della teoria, ovvero la CbKST, sia gli aspetti deterministici che quelli probabilistici necessitano di maggior approfondimento. L'obiettivo generale della tesi è dunque quello di cercare di rispondere a una serie di quesiti ancora aperti sia nella KST che nella CbKST.

Nei Capitoli 3, 4 e 5 si presentano tre ricerche che sono state condotte entro l'approccio tradizionale della teoria. Queste tre ricerche rappresentano lo sviluppo di una serie di procedure e modelli legati al *basic local independence model* (BLIM), il modello più utilizzato nelle applicazioni della teoria. Nonostante siano ricerche il cui carattere formale ha un importanza centrale, risolvono problematiche di tipo pratico. In particolare, nel Capitolo 3 si derivano le formule analitiche per il calcolo della matrice di covarianza delle stime dei parametri del BLIM, che fino ad oggi non erano mai state derivate. Questa matrice è necessaria sia per calcolare la varianza delle stime dei parametri del modello, sia per calcolare gli intervalli di confidenza, e permette dunque una più chiara interpretazione delle stime dei parametri che si ottengono nelle applicazioni del modello.

Nel Capitolo 4 si propone una procedura per testare una delle assunzioni su cui si basa il BLIM. Infatti, quando si applica un modello ai dati non si conosce la sensibilità dei test statistici (come per esempio il Chi-quadro di Pearson o il rapporto di verosimiglianza) alle violazioni delle sue assunzioni. Per esserne certi, oltre che testare la sua bontà di adattamento ai dati occorre testare, attraverso le procedure opportune, anche le assunzioni su cui si basa.

Nel Capitolo 5 si propongono due estensioni del BLIM al caso di dati mancanti. I dati mancanti sono un problema ben noto nell'ambito dell'inferenza statistica. Questo perché, anche quando viene data massima attenzione alla fase della raccolta dei dati, le risposte mancanti a uno o più item di un test, sono piuttosto frequenti. Nonostante il problema si presenti anche nelle

applicazioni della KST, la questione non era mai stata affrontata in modo approfondito.

Nei Capitoli 6 e 7 si presentano invece due ricerche che sono state condotte entro l'approccio della CbKST. Nel Capitolo 6 si approfondiscono alcuni aspetti di tipo probabilistico della teoria, che ad oggi risultavano del tutto trascurati. Si propone nello specifico un nuovo modello probabilistico per la diagnosi cognitiva delle abilità sottostanti alla risoluzione di un insieme di problemi. Nel Capitolo 7 invece il focus è sugli aspetti deterministici della CbKST. Si presentano dei risultati teorici che consentono l'applicazione della teoria allo sviluppo di sistemi intelligenti (intelligent tutoring system) per la valutazione e per l'apprendimento delle abilità individuali.

La tesi si conclude con una discussione generale (Capitolo 8) nella quale si presentano, all'interno di un quadro unificato, risultati e limiti delle ricerche e si tracciano i percorsi che la ricerca futura in questo ambito dovrebbe intraprendere. Particolare attenzione verrà data all'ambito di applicazione della KST più promettente e anche molto attuale, ovvero quello dello sviluppo di intelligent tutoring system (ITS).

Parte I

Quadro di Riferimento Teorico

Capitolo 1

La Knowledge Space Theory

1.1 Introduzione

Nell'ambito dei contesti educativo/scolastici, la valutazione della conoscenza ha da sempre ricoperto un ruolo fondamentale. Lo scenario, ad esempio, di un'*interrogazione orale* è ben definito nell'immaginario di tutti:

> *un'insegnante sta interrogando uno studente per individuare quale sia il livello della sua conoscenza in un particolare dominio (come ad esempio la matematica, la chimica, ecc.). L'insegnante sceglierà quindi una domanda, e poi un'altra e un'altra ancora sulla base delle risposte dello studente a quelle precedenti. Dopo un certo numero di domande l'insegnante sarà in grado di fornire una valutazione.*

La knowledge space theory, nata nel 1985 per opera dello psicologo matematico Jean-Claude Falmagne e del matematico Jean-Paul Doignon, è una teoria matematica per la valutazione adattiva della conoscenza. Essa fornisce un quadro teorico formale per operazionalizzare, nello specifico, il contesto tipico di un'interrogazione orale e, in generale, il contesto della valutazione della conoscenza.

Le teorie psicometriche, come ad esempio la teoria classica dei test e l'item response theory (IRT; Glas & Pimentel, 2008; Holman & Glas, 2005; Little & Rubin, 2002; Mislevy & Chang, 2000; Schafer, 1997), nate molto prima della KST, si sono occupate a lungo della valutazione della conoscenza. Cosa c'è dunque di nuovo nell'approccio seguito dalla KST? La distinzione fondamentale riguarda il tipo di valutazione. La KST infatti, non vuole in alcun modo stabile un punteggio (lungo un continuum numerico) per ciascuno studente, che sia rappresentativo della sua conoscenza. Il suo intento piuttosto, è quello di *descrivere* in modo estremamente dettagliato e preciso *ciò che uno studente sa* e *ciò che uno studente è pronto ad apprendere*, in un determinato dominio di conoscenza.

La *valutazione* è quindi, per la KST, allo stesso tempo *qualitativa* e *adattiva*. Con il primo termine ci si riferisce al fatto che non è di tipo numerico, ma descrittivo; con il secondo termine ci si riferisce al fatto che si adatta al particolare studente in esame, dal momento che sceglie le domande sulla base delle risposte precedenti, proprio come farebbe un insegnante durante un'interrogazione orale. La peculiarità di questa teoria, che, secondo chi scrive, le conferisce fascino intellettuale, è che questo tipo di valutazione è perseguita seguendo un altissimo rigore formale.

Nelle sezioni che seguono si approfondiscono così, gli aspetti matematici alla base della teoria. In particolare, nella Sezione 1.2 si descrivono gli aspetti deterministi della KST, ovvero si capirà come viene rappresentata la conoscenza. Nella Sezione 1.3 si descriveranno invece gli aspetti probabilistici della teoria, ovvero gli aspetti che permettono la sua applicabilità a dati reali. Infine, nella Sezione 1.4, si presenteranno gli aspetti più innovativi della teoria, che permettono di sfruttare al meglio il suo formalismo. Ci si riferisce alle procedure adattive sviluppate per la costruzione di sistemi computerizzati intelligenti applicati alla valutazione della conoscenza.

1.2 Stato di Conoscenza e Struttura di Conoscenza

In questa sezione si presentano gli aspetti deterministici della KST. Si intende fin da subito evidenziare che, volendo fornire una descrizione della conoscenza di un individuo, la KST sviluppa la sua teoria utilizzando il concetto di insieme, piuttosto che quello di numero. Per questo motivo la rappresentazione della conoscenza è vista in senso discreto e non continuo.

La conoscenza nella KST, viene rappresentata attraverso una *struttura di conoscenza*.

Definizione 1. Una struttura di conoscenza è una coppia $(Q, \mathcal{K})$ in cui Q è un insieme non vuoto, e $\mathcal{K}$ è una famiglia di sottoinsiemi di Q, che contiene almeno Q e l'insieme vuoto $\emptyset$. L'insieme Q è chiamato *dominio* della struttura di conoscenza. I suoi elementi sono *problemi* (o item) e i sottoinsiemi della famiglia $\mathcal{K}$ sono chiamati *stati di conoscenza*.

In altre parole una struttura di conoscenza $\mathcal{K}$ di un dominio di conoscenza Q è la collezione di tutti gli stati di conoscenza $K \subseteq Q$ osservabili in una popolazione di studenti. Gli elementi della struttura $\mathcal{K}$ sono stati di conoscenza, ovvero l'insieme dei problemi che uno studente è capace di risolvere nel dominio Q, in un determinato momento.

Va precisato che gli stati di conoscenza non sono direttamente osservabili, ma sono latenti. Questo perché, in contesti reali, uno studente potrebbe essere influenzato dalla pressione del tempo, piuttosto che da questioni emotive dovute allo stress del momento valutativo, che possono portarlo a commettere errori di distrazione. Allo stesso tempo, nel caso di domande a scelta multipla, lo studente potrebbe indovinare la risposta, pur non sapendola. In questi casi le risposte dello studente non rappresentano fedelmente la sua conoscenza. Lo stato di conoscenza K invece rappresenta la conoscenza dello studente che emerge in condizioni ideali, e andrà dunque inferito dall'insieme

delle risposte osservate. Quest'ultimo aspetto sarà approfondito nella sezione successiva.

Una caratteristica essenziale delle strutture di conoscenza è che non corrispondono all'insieme potenza dei problemi 2^Q, ovvero alla collezione di tutti i sottoinsiemi che si possono ottenere da Q. Infatti, la collezione di tutti gli stati di conoscenza cattura quella che è l'organizzazione della conoscenza, e riflette dunque le relazioni che si possono trovare tra un insieme di problemi. Per fare un esempio, si consideri il dominio di conoscenza $Q = \{a, b, c, d, e\}$ composto da 5 problemi, e si ipotizzi la struttura

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, d\}, \{b, c\}, \{a, b, c\}, \{a, b, d\}, \{b, c, d\}, \{a, b, c, d\}, Q\}. \quad (1.1)$$

Per prima cosa si nota che l'insieme vuoto $\emptyset$, ovvero la situazione in cui uno studente non sa risolvere nessun problema, e l'insieme totale Q, ovvero la situazione in cui li sa risolvere tutti, sono elementi della struttura (così come da Definizione 1). Altra cosa che si nota è che tra i $2^5 = 32$ sottoinsiemi che è possibile formare con i 5 elementi di Q, solamente 11 appartengono a $\mathcal{K}$. Questo dipende dal fatto che tra i problemi $q \in Q$ esistono delle relazioni (come ad esempio relazioni di prerequisito) per le quali la presenza di un item nello stato, dipende dalla presenza di un altro. Per comprendere questa proprietà, è utile costruire una diagramma di Hasse della struttura $\mathcal{K}$ (Figura 1.1). Nella figura ciascuno stato di conoscenza è rappresentato da un pallino nero, chiamato vertice del grafo. I vertici sono collegati tra di loro da una linea continua, chiamata arco, che rappresenta la relazione insiemistica dell'inclusione: un arco collega uno stato K a uno stato K' che si trova alla sua destra se $K \subseteq K'$. L'insieme vuoto e l'insieme totale Q occupano rispettivamente il vertice iniziale e quello finale. Leggendo il grafo da sinistra verso destra, è possibile collegare questi due vertici seguendo diversi percorsi, che da un punto di vista educativo, riflettono diversi *percorsi d'apprendimento*: all'inizio di un corso, è molto probabile che uno studente non

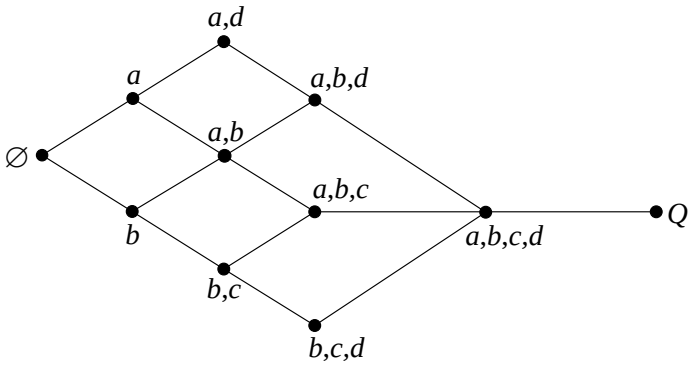

Figura 1.1: Diagramma di Hasse della struttura di conoscenza $\mathcal{K}$, riportata in Equazione (1.1)

conosca niente rispetto a quel particolare dominio di conoscenza, e si troverà pertanto nell'insieme vuoto; quello che si verifica durante l'apprendimento, è che questo studente si muoverà, gradualmente, da uno stato di conoscenza all'altro, seguendo uno dei percorsi rappresentati nel grafo.

Dal grafo di Figura 1.1 è possibile individuare anche una serie di relazioni tra gli item. Ad esempio, emerge che il problema b è un prerequisito del problema c. Questo perché b appartiene a tutti gli stati che contengono c. Lo stesso non vale, ad esempio, per i problemi a e b, dal momento che esistono sia lo stato $\{a\}$, che contiene solo il problema a , sia lo stato $\{b\}$, che contiene solo il problema b. In una struttura di conoscenza dunque è possibile definire una serie di relazioni di dipendenza tra i problemi $q \in Q$. E' proprio questa proprietà che consente alle strutture di contenere un numero ridotto di insiemi, rispetto a quelle che si potrebbero avere a livello teorico. Questo aspetto diviene estremamente importante nel caso di strutture di conoscenza costruite su un numero di problemi molto elevato, cosa piuttosto comune nelle applicazioni empiriche.

Vi è infine un ultimo aspetto che si vuole evidenziare, che richiama lo scenario dell'interrogazione orale citato nell'introduzione di questo capitolo. Rappresentare la conoscenza attraverso una struttura come quella esemplificata con $\mathcal{K}$, permette di dedurre, sulla base della risposta dello studente ad

un problema, quale sarà la risposta ad un altro problema. Se ad esempio, si osserva una risposta corretta per il problema c, assumendo (per ora) assenza di rumore, si deduce che anche la risposta al problema b sarà corretta. Una procedura computerizzata che sceglie le domande da somministrare a uno studente e registra le sue risposte, sulla base della struttura di conoscenza $\mathcal{K}$, sarà in grado di adattarsi a quello studente, e individuare il suo stato di conoscenza con un numero di domande minore a quelle che sono contenute in Q.

1.2.1 Le due frange di uno stato di conoscenza

La frangia esterna e la frangia interna sono due proprietà di uno stato di conoscenza K, che rivestono un ruolo fondamentale nella KST.

Definizione 2. La *frangia esterna* di uno stato K è l'insieme degli item q tali per cui aggiungendo q a K si ottiene un altro stato della struttura di conoscenza $\mathcal{K}$. Formalmente:

$$K^E = \{q \in Q \setminus K | K \setminus \{q\} \in \mathcal{K}\}. \tag{1.2}$$

Definizione 3. La *frangia interna* di uno stato K è l'insieme degli item q tali per cui eliminando q da K si ottiene un altro stato della struttura di conoscenza $\mathcal{K}$. Formalmente:

$$K^I = \{q \in K | K \setminus \{q\} \in \mathcal{K}\}. \tag{1.3}$$

Da un punto di vista educativo, la frangia esterna di uno stato rappresenta l'insieme degli item che uno studente è pronto ad apprendere. Si può affermare che questa frangia di fatto sia una rappresentazione formale di quella che Vygotsky (1978) ha chiamato *zona di sviluppo prossimale*. La frangia interna rappresenta invece l'insieme degli item da ultimi appresi.

Data questa interpretazione delle due frange, segue che, volendo individuare i bisogni educativi di uno studente, gli item che appartengono alla

frangia interna dello stato di conoscenza dello studente, saranno gli item che necessitano di un consolidamento, mentre quelli che appartengono alla frangia esterna, sono gli item che lo studente sarà in grado di apprendere con una relativa facilità, dal momento che ne possiede i prerequisiti. Nell'ottica di individuare il percorso d'apprendimento più adatto per uno studente, si apre la possibilità di utilizzare queste informazioni per selezionare i contenuti didattici da fornire allo studente.

Riassumendo, lo stato di conoscenza (*ciò che uno studente sa*), insieme alle sue due frange, ovvero *ciò che lo studente ha da ultimo appreso* e *ciò che è pronto ad apprendere*, oltre a fornire una rappresentazione completa della conoscenza di uno studente, consentono di selezionare il percorso d'apprendimento più opportuno.

Per fare un esempio, si consideri la struttura di conoscenza $\mathcal{K}$ riportata in (1.1), il cui diagramma ad Hasse è rappresentato in Figura 1.1. Si assuma ora che lo stato di conoscenza di uno studente sia $K = \{a, d\}$. Allora applicando la regola in (1.2), la frangia esterna di K è l'insieme $\{b\}$, infatti con $K \cup \{b\} = \{a, b, d\}$ si ottiene un altro stato della struttura $\mathcal{K}$. Mentre, applicando la regola in (1.3), la frangia interna di K è l'insieme $\{d\}$, infatti con $K \setminus \{d\} = \{a\}$, si ottiene un altro stato appartenente a $\mathcal{K}$. Osservando il grafo 1.1, si noti che lo stato $\{a, b, d\}$ è quello che si trova nel vertice successivo a quello dello stato K dello studente, mentre lo stato $\{a\}$ è quello che si trova nel vertice precedente a K.

Immaginando a questo punto, che venga fornito allo studente il materiale didattico per imparare a risolvere l'item b, lo studente passerà dallo stato $K = \{a, d\}$ allo stato $K' = \{a, b, d\}$. In questo modo, diviene possibile tracciare il percorso d'apprendimento dello studente:

$$\emptyset \subset \{a\} \subset \{a, d\} \subset \{a, b, d\},$$

che potrà poi continuare fino all'apprendimento dei restanti item del dominio di conoscenza Q.

Concludendo questa sezione, va evidenziato che, in generale, l'esistenza delle due frange di uno stato non è garantita. In altre parole un struttura di conoscenza qualsiasi potrebbe avere, per qualche stato, delle frange vuote. E' possibile evitare questa situazione considerando una classe speciale di strutture di conoscenza note come *spazi di conoscenza well-graded*, argomento della prossima sezione.

1.2.2 Spazi di conoscenza well-graded

Una classe speciale di strutture di conoscenza è nota come *spazi di conoscenza*. Per una trattazione completa dell'argomento si veda, ad esempio, Doignon e Falmagne (1999).

Definizione 4. Quando la famiglia $\mathcal{K}$ di strutture di conoscenza $(Q, \mathcal{K})$ è *chiusa all'unione* – ovvero quando $\bigcup \mathcal{F} \in \mathcal{K}$ per ogni $\mathcal{F} \subseteq \mathcal{K}$ – si può dire che $(Q, \mathcal{K})$ è uno spazio (di conoscenza), o, equivalentemente, che $\mathcal{K}$ è uno spazio di conoscenza su Q.

Gli spazi di conoscenza rivestono un ruolo chiave nella KST per la loro interpretazione empirica, che ha, di fatto, ispirato l'intero sviluppo formale della teoria. Si considerino, ad esempio, due studenti che, durante un corso, hanno interagito a lungo fra loro. Si supponga che che i loro stati di conoscenza iniziali, relativi a un certo dominio di conoscenza, fossero gli stati K e K'. E' lecito supporre che, a un certo punto durante il corso, i due studenti arrivino ad una conoscenza comune. Pertanto, il loro stato di conoscenza sarà $K \cup K'$. Ovviamente, questo potrebbe non verificarsi, ma è ragionevole supporre che lo stato derivante dall'unione di quei due insiemi debba esistere nella struttura.

Un'altra ragione che giustifica l'importanza degli spazi, dipende dal fatto che sono in relazione biunivoca con una classe di relazioni di prerequisito che prendono nome di *entail relation* (Koppen & Doignon, 1990). Non si vuole in questa sede discutere in dettaglio delle caratteristiche formali delle entail

relation. Si vuole comunque evidenziare che sono una classe di relazioni binarie che ammettono uno o più insiemi di prerequisiti per ogni stato di conoscenza appartenente a una struttura.

Vi è poi una classe speciale di spazi di conoscenza, chiamati *spazi di conoscenza weel-graded* o più recentemente definiti *spazi di apprendimento* (Falmagne & Doignon, 2011; Falmagne, Albert, Doble, Eppstein, & Hu, 2013).

Definizione 5. Uno spazio di apprendimento è uno spazio di conoscenza $\mathcal{K}$ su un insieme finito Q, in cui vale la condizione aggiuntiva

per ogni $K \in \mathcal{K}$ non vuoto, esiste un item $q \in K$ tale che $K \setminus \{q\} \in \mathcal{K}$.

La condizione aggiuntiva, consente a uno spazio di essere well-graded, ovvero garantisce che ogni stato della struttura possieda frange non vuote. Da un punto di vista educativo, tale proprietà garantisce la *gradualità dell'apprendimento*, infatti uno studente progredisce nell'apprendimento imparando un item per volta. L'esempio relativo alla struttura di conoscenza (1.1), rappresentata nel grafo di Figura 1.1, discusso nella sezione 1.2.1, è proprio uno spazio di conoscenza well-graded (o spazio di apprendimento).

Va infine evidenziato che gli spazi sono particolarmente utili anche nella fase di costruzione di una struttura di conoscenza, attraverso una tecnica chiamata *interrogazione di esperti* (QUERY, Kambouri, Koppen, Villano, & Falmagne, 1994; Stefanutti & Koppen, 2003). Nella prossima sezione si illustrano alcune tecniche di costruzione di una struttura di conoscenza.

1.2.3 Costruire una struttura di conoscenza

Non si vuole in questa sezione entrare in dettaglio dei metodi attualmente a disposizione per costruire le strutture di conoscenza, ma si fornisce una breve descrizione dello stato dell'arte dell'argomento.

I metodi che si pongono l'obiettivo di costruire una struttura di conoscenza seguono un principio generale comune: tra gli item q appartenenti a un

dominio di conoscenza Q, si possono trovare delle relazioni di prerequisito che determinano i vincoli per stabilire quali, nell'insieme potenza 2^Q, sono stati di conoscenza $K \in \mathcal{K}$ e quali non lo sono. I metodi fino ad oggi sviluppati si possono raggruppare in tre tipologie:

1. l'interrogazione di esperti, anche nota con il nome QUERY (Kambouri et al., 1994; Stefanutti & Koppen, 2003);

2. l'analisi del compito cognitivo, anche chiamata *cognitive task analysis* (Albert & Lukas, 1999);

3. metodi statistici di derivazione delle strutture dai dati, il più noto dei quali è conosciuto con il nome di *item tree analysis* (Schrepp, 1999, 2003).

1.3 Un Modello Probabilistico per Strutture di Conoscenza: il Basic Local Independence Model

Il basic local independence model (BLIM) è un modello probabilistico per strutture di conoscenza proposto da Falmagne e Doignon (1988a, 1988b). Come visto nella sezione precedente una struttura di conoscenza è un modello deterministico dell'organizzazione della conoscenza che, come evidenziato da Doignon e Falmagne (1999) e Falmagne e Doignon (2011), non può, in alcun modo, fare previsioni realistiche sulle risposte di uno studente a un insieme di item. Per questa ragione la validazione empirica delle strutture di conoscenza necessita di un approccio probabilistico.

In primo luogo, data una certa popolazione di studenti di riferimento, è plausibile supporre che ciascun stato di conoscenza si possa osservare nella popolazione con una certa frequenza. In altre parole, è possibile assumere l'esistenza di una distribuzione di probabilità sulla collezione degli stati.

Formalmente questa assunzione è data dall'introduzione di una *struttura di conoscenza probabilistica*, che è una terna $(Q, \mathcal{K}, \pi)$, dove la coppia (Q, K) è una struttura di conoscenza finita, mentre π è una distribuzione di probabilità su $\mathcal{K}$, ovvero $\pi_K \geq 0$ per tutti i $K \in \mathcal{K}$, e $\sum_{K \in \mathcal{K}} \pi_K = 1$.

In secondo luogo, va fatta una distinzione tra ciò che non è direttamente osservabile, ovvero lo stato di conoscenza $K \in \mathcal{K}$ di uno studente, e ciò che invece si osserva, ovvero il suo pattern di risposta $R \subseteq Q$. Il pattern di risposta di uno studente corrisponde alla collezione dei problemi che ricevono una risposta corretta. Naturalmente, da R è completamente deducibile anche l'insieme delle risposte errate, che corrisponde al suo complemento, ossia $Q \setminus R$. Nel BLIM, la relazione tra K ed R è di natura probabilistica ed è specificata dal modello a classi latenti

$$P(R) = \sum_{K \in \mathcal{K}} P(R|K)\pi_K, \tag{1.4}$$

dove:

- $P(R)$ è la probabilità di campionare uno studente il cui pattern di risposta è R;
- $P(R|K)$ è la probabilità condizionale di osservare il pattern R dato che lo stato di conoscenza è K;
- π_K è la probabilità di estrarre dalla popolazione uno studente nello stato K.

Si introducono poi specifiche assunzioni sulle probabilità condizionali $P(R|K)$. La prima, chiamata *regola di risposta*, riguarda la probabilità condizionale di ottenere una risposta corretta per un problema q, dato un certo stato di conoscenza K.

[**A1**] Sia $\mathbf{R}$ una variabile casuale (con realizzazioni nell'insieme potenza 2^Q) che rappresenta il pattern di risposta completo di uno studente campionato casualmente. Per ogni item $q \in Q$, e ogni stato di conoscenza $K \in \mathcal{K}$,

si denoti con $P(q \in \mathbf{R}|K)$ la probabilità condizionale che q appartenga al pattern di risposta dello studente, dato che il suo stato di conoscenza è K. Formalmente:

$$P(q \in \mathbf{R}|K) = \begin{cases} 1 - \beta_q & \text{se } q \in K \\ \eta_q & \text{se } q \in Q \setminus K, \end{cases} \tag{1.5}$$

dove:

- $\beta_q \in [0, 1)$ è la *probabilità di careless error* dell'item q;
- $\eta_q \in [0, 1)$ è la *probabilità di lucky guess* di q.

Sotto l'Assunzione [**A1**] e l'assunzione secondo cui le risposte agli item sono fra loro condizionalmente indipendenti, dato lo stato (indipendenza locale), la probabilità condizionale $P(R|K)$ del pattern di risposa $R \subseteq Q$, dato lo stato di conoscenza $K \in \mathcal{K}$, ha la forma

$$P(R|K) = \left[\prod_{q \in K \setminus R} \beta_q\right] \left[\prod_{q \in K \cap R} (1 - \beta_q)\right] \left[\prod_{q \in R \setminus K} \eta_q\right] \left[\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q)\right]. \tag{1.6}$$

L'Equazione (1.6) è composta dal prodotto di quattro termini:

1. $\prod_{q \in K \setminus R} \beta_q$ è il prodotto delle probabilità di careless error β_q di tutti gli item che appartengono a K ma non al pattern R ($q \in K \setminus R$);
2. $\prod_{q \in K \cap R} (1 - \beta_q)$ è il prodotto delle probabilità di non commettere un errore di distrazione $(1 - \beta_q)$ per gli item che appartengono sia a K che a R;
3. $\prod_{q \in R \setminus K} \eta_q$ è il prodotto delle probabilità di lucky guess (η_q) di tutti gli item che appartengono a R ma non a K;
4. $\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q)$ è il prodotto delle probabilità di non commettere un errore di distrazione $(1 - \eta_q)$ per tutti gli item che non appartengono né a R né a K.

E' chiaro che il secondo e il quarto termine rappresentano i due casi in cui il pattern di risposta R e lo stato di conoscenza K dello studente coincidono per un particolare item q, mentre il primo e il terzo termine rappresentano i due casi in cui si può osservare una discordanza fra i due. E' proprio per modellare questi due ultimi casi che si introducono i due parametri β_q ed η_q, che, per questo motivo, sono anche chiamati *parametri d'errore* del modello.

Per fare un esempio, si consideri il dominio di conoscenza $Q = \{a, b, c, d\}$, e si supponga di osservare il pattern di risposta $R = \{b, c\}$ di uno studente il cui stato di conoscenza è $K = \{a, b\}$. Confrontando R e K si può concludere che lo studente:

- ha commesso un errore di distrazione per il problema a, evento che ha probabilità β_a;
- non ha commesso errori di distrazione per il problema b, evento che ha probabilità $1 - \beta_b$;
- ha commesso una lucky guess per il problema c, evento che ha probabilità η_c;
- non ha commesso lucky guess per il problema d, evento che ha probabilità $1 - \eta_d$.

Applicando l'Equazione (1.6) si ottiene che la probabilità del pattern $\{b, c\}$, dato lo stato $\{a, b\}$ dello studente, è

$$P(\{b, c\}|\{a, b\}) = \beta_a(1 - \beta_b)\eta_c(1 - \eta_d).$$

1.3.1 Bontà di adattamento del BLIM

Due aspetti fondamentali per le applicazioni del BLIM a dati empirici, sono il test della bontà di adattamento ai dati e la stima dei parametri. In questo paragrafo si presenterà la tecnica più utilizzata per testare la bontà di adattamento di un modello, ovvero il test con la statistica di Chi-quadro.

Per il modello BLIM, la statistica di Chi-quadro assume la forma

$$\chi^2(\theta; \mathcal{D}, N) = \sum_{R \in \mathcal{R}} \frac{(F(R) - NP_\theta(R))^2}{NP_\theta(R)}, \tag{1.7}$$

dove $\theta = \beta, \eta, \pi$ è il vettore dei parametri del modello, $\mathcal{D}$ è un insieme di pattern osservati, $F(R)$ è la frequenza osservata del pattern $R \in \mathcal{R}$, N è la numerosità campionaria e $P_\theta(R)$ è la probabilità marginale di R.

Per N grandi e sotto l'assunzione che i pattern di risposta forniti da soggetti diversi siano indipendenti, la variabile casuale $\chi^2(\theta; \mathcal{D}, N)$ si approssima alla distribuzione teorica di Chi-quadro con

$$df = (2^{|Q|} - 1) - (|\mathcal{K}| - 1) - 2|Q| \tag{1.8}$$

gradi di libertà. Nell'Equazione (1.8) il simbolo $|Q|$ indica il numero degli item, mentre $|\mathcal{K}|$ indica il numero di stati della struttura di conoscenza. Il termine destro dell'equazione rappresenta la differenza tra il numero dei gradi di libertà dei dati $2^{|Q|} - 1$ e il numero dei parametri liberi di variare del modello $(|\mathcal{K}| - 1) + 2|Q|$. Se nell'esempio presentato alla fine della sezione 1.3, il dominio di 4 item $Q = \{a, b, c, d\}$ fosse stato rappresentato da una struttura di conoscenza con 7 stati, si avrebbero $df = (2^4 - 1) - (7 - 1) - 2 \cdot 4 = 1$ gradi di libertà.

Tipicamente, il modello viene accettato quando la probabilità di ottenere un valore di Chi-quadro maggiore di χ^2_{df} è *maggiore* di .05, rifiutato altrimenti. Si ricorda infatti che nella verifica di ipotesi applicata ai modelli formali, il modello rappresenta l'ipotesi nulla del ricercatore, anziché quella alternativa.

Va infine evidenziato che per matrici sparse di dati (matrici in cui molte celle hanno una frequenza osservata pari a zero) la distribuzione di $\chi^2(\theta; \mathcal{D}, N)$ non si approssima a quella teorica. Un modo per ovviare al problema è approssimare la distribuzione, attraverso una procedura di *bootstrap parametrico* (Efron, 1979). Il p-value associato al Chi-quadro viene quindi calcolato nel modo seguente: (1) si generano n campioni della stessa numerosità del campione di pattern osservati; (2) si stima il BLIM su ciascuno di essi; (3)

si calcola la proporzione di volte in cui il Chi-quadro è maggiore di quello ottenuto sui dati empirici. Tale proporzione rappresenta il p-value ottenuto via bootstrap.

1.3.2 Stima dei parametri del modello

In questa sezione si presentano due dei metodi presenti in letteratura, per stimare i parametri del BLIM, ovvero: (1) la stima per massima verosimiglianza (Stefanutti & Robusto, 2009) e (2) la stima per discrepanza minima (Heller & Wickelmaier, 2013).

Nello stimare i parametri di un modello è del tutto ragionevole scegliere i valori dei parametri che rendono i dati osservati più probabili. Questo tipo di stime sono chiamate in letteratura *stime per massima verosimiglianza* (MLE - maximum likelihhod estimates). Essendo il BLIM un modello a classi latenti (dove le classi latenti sono gli stati di conoscenza), è possibile ottenere le stime dei parametri con il metodo della massima verosimiglianza (ML - maximum likelihood) sviluppato per i modelli multinomiali. A questo scopo è utile presentare la funzione di verosimiglianza del BLIM, che per un campione finito $\mathcal{D}$ di numerosità N è

$$L(\theta|\mathcal{D}) = \prod_{R\subseteq Q} (P_\theta(R))^{F(R)}, \tag{1.9}$$

dove: θ è il vettore dei parametri del modello e $F(R)$ è la frequenza osservata del pattern $R \subseteq Q$. Le stime MLE dei parametri nel vettore θ, si ottengono massimizzando il logaritmo della funzione di verosimiglianza in (1.9).

Il procedimento di massimizzazione non ha una soluzione analitica, perché conduce a un sistema di equazioni non lineari. Per tale ragione si utilizzano algoritmi di tipo numerico. Nei modelli a classi latenti l'algoritmo più utilizzato a questo scopo è l'*expectation-maximization* (EM, Dempster, Laird, & Rubin, 1997). L'algoritmo itera su due step:

- E-step (expectation), nel quale viene calcolato il valore atteso della funzione di log-verosimiglianza dei dati completi, date le frequenze os-

servate $F(R)$ e le stime dei parametri attuali. Se si avessero a disposizione dati completi ciascuna singola osservazione è una coppia $(R, K) \in \mathcal{R} \times \mathcal{K}$ e il valore atteso della log-verosimiglianza dei dati completi è

$$E\left[\log \prod_{(R,K)\in\mathcal{R}\times\mathcal{K}} (P(R|K)\pi_K)^{F(R,K)} |\mathcal{D}, \theta_{i-1}\right] =$$
$$E\left[\sum_{(R,K)\in\mathcal{R}\times\mathcal{K}} F(R,K)\log(P(R|K)\pi_K)|\mathcal{D}, \theta_{i-1}\right],$$

dove $F(R, K)$ è la frequenza della coppia (R, K) nel campione di dati completi e Θ_{i-1} è il vettore dei parametri del modello stimati al passo $i - 1$;

- M-step (maximization), nel quale le stime ottenute all'E-step vengono massimizzate.

Questi due step si ripetono il numero necessario di volte a garantire l'aumento della verosimiglianza del modello e la stabilità delle stime.

Per quanto riguarda il *metodo della discrepanza minima* (MD - minimum discrepancy), si basa invece su un indice, molto usato nella KST, che descrive quanto i pattern di risposta osservati sono vicini alla struttura di conoscenza in esame. Questo indice è la distanza simmetrica $d(R, K)$ tra i pattern di risposta R e gli stati di conoscenza $K \in \mathcal{K}$, che si calcola

$$d(R, K) = |(R \setminus K) \cup (K \setminus R)|, \tag{1.10}$$

e rappresenta il numero di item che appartengono a uno dei due insiemi R o K, ma non all'altro. Ora, dato un pattern di risposta R e ogni stato di conoscenza $K \in \mathcal{K}$, si consideri il minimo di $d(R, K)$

$$d(R, \mathcal{K}) = min_{K\in\mathcal{K}} d(R, K).$$

Sotto le due assunzioni (1) lo stato di conoscenza K sottostante al pattern di risposta R è quello che sta a minima distanza, formalmente $d(R, K) =$

$d(R, \mathcal{K})$; (2) tutti gli stati di conoscenza K a distanza minima dal pattern R sono equiprobabili, le stime MD dei parametri d'errore del BLIM si calcolano nel modo seguente:

- $\beta_q = P(\mathcal{R}_{\bar{q}}|\mathcal{K}_q)$, dove $\mathcal{R}_{\bar{q}}$ è l'insieme di tutti i pattern che non contengono l'item q, e $\mathcal{K}_q$ è l'insieme di tutti gli stati che contengono l'item q;
- $\eta_q = P(\mathcal{R}_q|\mathcal{K}_{\bar{q}})$, dove $\mathcal{R}_q$ è l'insieme tutti i pattern che contengono l'item q, e $\mathcal{K}_{\bar{q}}$ è l'insieme tutti gli stati che non contengono l'item q.
- $\pi_K = \sum_{R \in \mathcal{R}} P(K|R)P(R)$, dove $P(R) = F(R)/N$ e $P(K|R)$ è ottenuta mediante l'applicazione del teorema di Bayes.

Si possono trovare vantaggi e svantaggi in ciascuno dei due metodi per la stima del BLIM, appena presentati. Per quanto riguarda l'efficienza, a parità del numero di parametri da stimare, il metodo MD è sicuramente più veloce del metodo ML. Dall'altro lato quest'ultimo ottiene stime più vicine a quelle vere, di quanto non faccia il metodo MD, che sembra invece sottostimare, anche se di poco, sistematicamente i parametri d'errore del modello. Heller e Wickelmaier (2013) hanno sviluppato un metodo ibrido chiamato MDML che combina i vantaggi di entrambi e ne limita gli svantaggi. Va comunque evidenziato che, a meno di chiare motivazioni per preferire l'efficienza della procedura all'accuratezza delle stime, il metodo ML è migliore. Per questa ragione, in tutti gli studi sul BLIM presentati nei prossimi capitoli si è scelto di stimare i parametri per massima verosimiglianza utilizzando l'EM.

1.3.3 Testare l'identificabilità del BLIM

Un aspetto fondamentale dei modelli matematici parametrici riguarda la loro *identificabilità*. Lo studio di questa proprietà dei modelli ha una lunga tradizione scientifica Bamber e van Santen (1985). Nonostante ciò, l'identificabilità è un aspetto che viene ancora oggi trascurato nella fase dell'appli-

cazione di un modello ai dati. Una motivazione di questa mancanza dipende dalla complessità formale che è richiesta per lo sviluppo di procedure che permettono di testare tale proprietà.

Prima di descrivere l'*identificabilità* di un modello, è opportuno fornire una definizione formale di un *modello parametrico*. Esso è una tripla (Θ, f, Ω), dove:

- $\Theta \subseteq \Re^m$ è chiamato spazio parametrico del modello;
- $\Omega \subseteq \Re^n$ è lo spazio degli esiti del modello;
- $f : \Theta \to \Omega$ è la funzione di previsione (prediction function) del modello.

L'insieme $\Re$ è l'insieme dei numeri reali, l'intero $m > 0$ è il numero di parametri del modello, mentre $n > 0$ è il numero degli esiti teoricamente osservabili. Una funzione di previsione $f(\theta)$ di un dato vettore di parametri $\theta \in \Theta$ è uno dei possibili esiti in Ω. Un modello (Θ, f, Ω) è identificabile se la è una corrispondenza biunivoca fra Θ e Ω è unica, al contrario non è identificabile se la stessa identica previsione si può ottenere da diversi punti dello spazio parametrico. In quest'ultimo caso l'interpretazione dei suoi parametri non è possibile. Se esistono differenti soluzioni per i parametri del modello, i loro valori non possono essere interpretati, e il modello perde la capacità esplicativa del fenomeno in esame.

Bamber e van Santen (1985, 2000) hanno proposto un metodo per testare l'identificabilità di un modello, che si basa sull'analisi della matrice Jacobiana della funzione di previsione del modello. Senza entrare nel merito degli aspetti formali del metodo, gli autori hanno trovato che, se il rango di questa matrice è minore del numero di parametri liberi del modello, allora il modello non è identificabile.

Recentemente, è stato sviluppato un metodo basato su quello proposto da Bamber e van Santen (2000), per il BLIM (Stefanutti, Heller, Anselmi, & Robusto, 2012). La procedura che propongono questi autori, consente

non solo di testare l'identificabilità del BLIM, ma, nel caso in cui il modello non sia identificabile, permette di individuare quali parametri ne sono la causa. Questa procedura è stata implementata in MATLAB, ed è chiamato BLIMIT (BLIM identification test). In tutti gli studi sul BLIM, presentati nei prossimi capitoli, l'identificazione del modello è stata testata utilizzando questa procedura.

Oltre ai vantaggi applicativi, lo studio dell'identificabilità del BLIM ha permesso lo studio delle cause che portano alla non identificazione del modello (Spoto, Stefanutti, & Vidotto, 2012). In particolare è stato scoperto che l'identificabilità del BLIM dipende dalla particolare struttura di conoscenza sulla quale viene definito. Gli stessi autori hanno trovato un metodo estremamente semplice per risolvere queste problematiche. Infatti, nel caso in cui i parametri di un item q non siano identificabili, è sufficiente aggiungere un item p tale che q e p sono ugualmente informativi. In questo modo si ripristina l'identificabilità.

1.4 La Valutazione Adattiva della Conoscenza

La valutazione adattiva della conoscenza è sicuramente la più importante applicazione della KST. Fin dalla prima pubblicazione (Doignon & Falmagne, 1985), apparsa sulla rivista *International Journal of Man-Machine Studies*, era ben chiaro l'obiettivo che i due autori perseguivano: costruire un *intellignet tutoring sistem* (sistema di tutoring intelligente) per la valutazione della conoscenza degli individui. Un sistema di questo tipo doveva avere una caratteristica essenziale: essere in grado di adattarsi al particolare individuo in esame, riuscendo a costruire una descrizione estremamente accurata della sue conoscenze, in modo efficiente, utilizzando il minor numero di domande possibile.

I vantaggi di pensare alla valutazione delle conoscenze in questi termini sono molteplici. In primo luogo permettono l'individualizzazione della valutazione: questo tipo di procedure infatti, si adattano al particolare studente esaminato, scegliendo le domande sulla base delle sue risposte, proprio come farebbe un insegnante durante un'interrogazione orale. Il secondo vantaggio segue direttamente dal primo: avere a disposizione una procedura computerizzata che mima il comportamento di un insegnante, è sicuramente più economico in termini di costi e di tempo, ma anche di accuratezza. I computer infatti, non soffrono né di problematiche legate alla stanchezza né delle *misconception* (idee o giudizi erronei) che talvolta gli insegnanti hanno sugli studenti e che possono influenzare la valutazione. Il terzo vantaggio dipende dal fatto che sulla base di una valutazione individuale è possibile sviluppare un percorso di apprendimento personalizzato, che incontra le particolare esigenze di ciascun studente, consentendo una didattica più efficace. L'ambiente classe infatti si compone di studenti che hanno bisogni ed esperienze anche molto diversi fra loro, e avere a disposizione uno strumento che risponda a queste differenze, può essere un valido supporto alla didattica degli insegnanti.

In questa sezione si prenderanno in esame le procedure adattive sviluppate nella KST. Esse si basano sul principio generale che non tutti i sottoinsiemi dei problemi, formulabili in un dominio di conoscenza, sono *stati di conoscenza* (come spiegato nella Sezione 1.2). Le strutture di conoscenza infatti, definiscono relazioni di dipendenza (come ad es. relazioni di prerequisito) tra i problemi del dominio. Tali relazioni permettono di fare una previsione sulla correttezza o meno della risposta di uno studente ad un problema, sulla base delle risposte già osservate ad altri problemi.

A seconda della procedura, tale previsione può essere di tipo deterministico oppure di tipo non deterministico, ma entrambe condividono l'obiettivo di individuare, con il minor numero di domande, lo stato di conoscenza di uno studente tra quelli appartenenti alla struttura di conoscenza del dominio

in esame. La differenza essenziale invece, è che le procedure deterministiche assumono che il comportamento di risposta di uno studente è completamente determinato dal suo stato di conoscenza. In questo caso, pattern di risposta e stato di conoscenza sarebbero identici. Le procedure non deterministiche, invece tengono conto del rumore che può intervenire sul pattern di risposta di uno studente, come per esempio errori di distrazione e lucky guess.

Nella Sezione 1.4.1 si analizzano le procedure deterministiche, mentre nella Sezione 1.4.2 si prendono in esame quelle non deterministiche.

1.4.1 Procedure deterministiche

La prime procedure ad essere state sviluppate sono quelle deterministiche, che sono sicuramente le più semplici, ma che soffrono, come si vedrà, di chiare limitazioni. In particolare vi sono due tipi di procedure deterministiche. Nella prima, sviluppata da Degreef, Doignon, Ducamp, e Falmagne (1986), ad ogni passo dell'assessment il problema da presentare allo studente viene scelto tra quelli che appartengono approssimativamente a metà degli stati di conoscenza della struttura. L'insieme dei possibili stati di conoscenza viene poi ridotto passo passo eliminando gli stati che non sono in accordo con le risposte osservate. La procedura continua finché rimane un unico stato, che sarà quello assegnato allo studente.

La Tabella 1.1 illustra il funzionamento della procedura deterministica attraverso un esempio. Si supponga di voler individuare lo stato di conoscenza di uno studente nel dominio di conoscenza $Q = \{a, b, c, d\}$. Si supponga inoltre che la struttura di conoscenza per Q sia $\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, Q\}$. Nella prima colonna della tabella sono elencati i sei stati di conoscenza appartenenti a $\mathcal{K}$. La seconda colonna rappresenta ciò che accade al primo step della procedura: viene selezionato l'item b e lo studente risponde correttamente ($R_b = 1$). Tutti gli stati che contengono l'item b (indicati con un segno di spunta) sono plausibili, essendo in accordo con la risposta osservata dello

Tabella 1.1: Esempio di applicazione della procedura deterministica di Degreef, et. al. (1986). Per i dettagli si faccia riferimento al testo.

$K \in \mathcal{K}$	$R_b = 1$	$R_a = 1$	$R_c = 0$
$\emptyset$			
$\{a\}$			
$\{b\}$	✓		
$\{a, b\}$	✓	✓	✓
$\{a, b, c\}$	✓	✓	
Q	✓	✓	

studente. Allo step numero due (terza colonna della tabella), la procedura seleziona l'item a e lo studente risponde correttamente ($R_a = 1$). Gli stati plausibili vengono ora aggiornati in accordo con la risposta osservata dello studente. Infine nello step 3, la procedura seleziona l'item c e lo studente risponde in modo errato ($R_c = 0$). A questo punto, fra gli stati plausibili individuati allo step precedente ne rimane solo uno, ovvero $\{a, b\}$, che sarà lo stato di conoscenza assegnato dalla procedura deterministica allo studente.

Nel caso di strutture di conoscenza con migliaia di stati, è chiaro che una procedura di questo tipo può diventare molto pesante dal punto di vista computazionale. Oltre a ciò risulta poco realistica. Uno studente infatti, potrebbe sbagliare un problema per distrazione, oppure al contrario potrebbe indovinare la risposta (si pensi ad esempio al caso delle domande con risposta a scelta multipla).

Le procedure non deterministiche superano questo problema, introducendo nella procedura aspetti di tipo probabilistico.

1.4.2 Procedure non deterministiche

Falmagne e Doignon (1988a, 1988b) hanno sviluppato due procedure non deterministiche, una di tipo discreto, l'altra di tipo continuo. Entrambe

considerano la possibilità che uno studente commetta errori di distrazione nel rispondere alle domande, oppure indovini le risposte ad alcune domande. Si darà di seguito una breve descrizione della procedura discreta, per poi approfondire quella continua.

La procedura non deterministica *discreta* (Falmagne & Doignon, 1988b; Doignon, 1994), inizia ad un primo step come una procedura deterministica, andando ad individuare uno stato di conoscenza preliminare, che si assume essere vicino a quello vero. In un secondo step vengono individuati, fra gli stati della struttura di conoscenza, quelli che stanno a distanza minima dallo stato preliminare. Questo insieme di stati plausibili si differenziano dallo stato preliminare per avere qualche item in più o in meno. Negli step successivi, gli item individuati allo step precedente vengono presentati allo studente, e lo stato di conoscenza preliminare viene aggiornato in accordo con le risposte osservate.

La procedura non deterministica *continua* (Falmagne & Doignon, 1988a), considera invece una funzione di verosimiglianza $\mathcal{L}_n$ sugli stati di conoscenza, che esprime la loro plausibilità per lo studente in esame, ad ogni step n della procedura. Per ogni stato $K \in \mathcal{K}$, si denota con $\mathcal{L}_n(K) \geq 0$ la verosimiglianza al passo n che lo studente sia nello stato di conoscenza K con $\sum_{K\in\mathcal{K}} \mathcal{L}_n(K) = 1$. La funzione di verosimiglianza Ł$_n$ può essere estesa a insiemi $\mathcal{F} \subseteq \mathcal{K}$ di stati, mediante la trasformazione:

$$\mathcal{L}_n(\mathcal{F}) = \sum_{K\in\mathcal{F}} \mathcal{L}_n(K). \tag{1.11}$$

Ogni applicazione della procedura di assessment è la realizzazione di un processo stocastico. Iniziando dallo step $\mathcal{L}_0$ l'assessment procede in modo iterativo (Figura 1.2). Ad ogni step $n > 0$ il processo deve rispondere a tre domande: (1) quale domanda presentare; (2) come tenere conto delle risposte dello studente; (3) quando termina la procedura. La risposta a queste domande è data da tre regole fondamentali (in figura rappresentate dagli ovali): (1) la regola di interrogazione; (2) la regola di aggiornamento; e (3)

la regola di stop. A partire dalla distribuzione di verosimiglianza iniziale

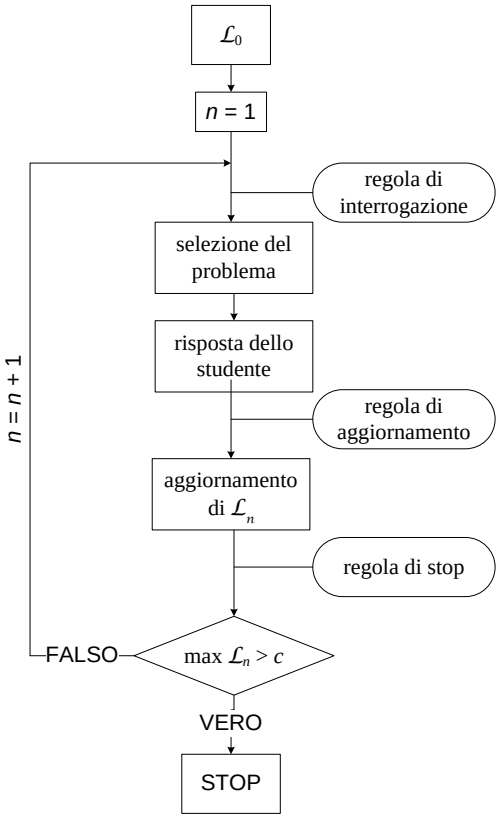

Figura 1.2: Algoritmo della procedura adattiva non deterministica continua sviluppata da Falmagne e Doignon (1988a).

$\mathcal{L}_0$, viene selezionata una domanda sulla base della regola di interrogazione. Tale domanda viene presentata allo studente, e sulla base della sua risposta, viene ricalcolata la funzione di verosimiglianza $\mathcal{L}_{n+1}$ allo step $n + 1$, in base alla regola di aggiornamento. Le iterazioni vengono ripetute fino a che non viene raggiunto il criterio c, stabilito dalla regola di stop. Esistono svariate versioni di queste tre regole. Di seguito si presentano quelle più utilizzate.

In primo luogo allo step $n = 0$ è necessario assumere una verosimiglianza a priori $\mathcal{L}_0$ per ogni stato di conoscenza. La scelta della distribuzione a priori $\mathcal{L}_0$ può dipendere da svariate condizioni, come ad esempio il livello scolastico degli studenti, piuttosto che da informazioni ottenute da somministrazioni precedenti. Questa scelta non è banale, come dimostrato da una serie di

studi (Heller & Repitsch, 2012; Hockemeyer, 2002). I ricercatori hanno condotto una serie di simulazioni utilizzando diverse tipologie di distribuzioni a priori. Ne è emerso che la scelta della distribuzione a priori può influenzare sia l'efficienza che l'accuratezza dell'assessment. Se non si hanno informazioni affidabili, gli autori suggeriscono di assumere equiprobabilità sugli di conoscenza e scegliere pertanto la distribuzione discreta uniforme, dove la verosimiglianza di ciascuno stato è $\mathcal{L}_n(K) = 1/|\mathcal{K}|$.

La regola di interrogazione *half-split*, ha l'obiettivo di scegliere gli item successivi in modo da rendere minimo il numero complessivo di domande necessarie per completare l'assessment. Per fare questo, allo step n viene somministrato l'item per il quale la risposta (corretta o errata) dello studente risulta massimamente informativa per l'aggiornamento della verosimglianza. Il che equivale a scegliere l'item q che suddivide l'insieme $\mathcal{K}$ in due sottoinsiemi $\mathcal{K}_q$ e $\mathcal{K}_{\bar{q}}$, aventi verosimiglianze più simili possibili. L'insieme $\mathcal{K}_q$ è la collezione di tutti gli stati di conoscenza che contengono l'item q, mentre $\mathcal{K}_{\bar{q}}$ è la collezione di tutti gli stati che non contengono l'item q. Questo significa che $\mathcal{L}_n(\mathcal{K}_q)$ deve essere più vicino possibile a $\mathcal{L}_n(\mathcal{K}_{\bar{q}}) = 1 - \mathcal{L}_n(\mathcal{K}_q)$. Di conseguenza, la regola di interrogazione half-split richiede che si somministri l'item q che rende minima la differenza assoluta

$$\left| \mathcal{L}_n(\mathcal{K}_q) - \frac{1}{2} \right|. \tag{1.12}$$

Se più item minimizzano questa differenza, si sceglie a caso un item fra questi. Da un punto di vista psicologico tale regola suggerisce di scegliere l'item che non né troppo facile (e quindi noioso), né troppo difficile (e quindi demotivante) per lo studente.

Per quanto riguarda la *regola di aggiornamento*, la funzione di verosimiglianza $\mathcal{L}_n$, viene aggiornata ad ogni step n della procedura secondo due principi:

- se si osserva una risposta corretta per l'item q, la verosimiglianza di tutti gli stati che lo contengono aumenta, mentre la verosimiglianza di

tutti gli stati che non lo contengono diminuisce;

- se si osserva una risposta errata per l'item q, la verosimiglianza di tutti i gli stati che lo contengono diminuisce, mentre la verosimiglianza di tutti gli stati che non lo contengono aumenta.

Da un punto di vista formale, la regola di aggiornamento stabilisce che la verosimiglianza di ciascun stato $K \in \mathcal{K}$ venga aggiornata, tenendo conto anche delle probabilità di careless error (β_q) e lucky guess (η_q) degli item $q \in Q$, secondo la seguente equazione:

$$\mathcal{L}_{n+1}(K) = \frac{P(R_q|K)\mathcal{L}_n(K)}{\sum_{K' \in \mathcal{K}} P(R_q|K')\mathcal{L}_n(K')}, \tag{1.13}$$

dove $P(R_q|K)$ è la probabilità condizionale della risposta osservata all'item q dato lo stato di conoscenza K. Indicando con R_q la risposta (1 = corretta, 0 = errata) fornita all'item q, tale probabilità è definita come

$$P(R_q|K) = \begin{cases} \beta_q & R_q = 0 \text{ e } q \in K, \\ 1 - \eta_q & R_q = 0 \text{ e } q \notin K, \\ 1 - \beta_q & R_q = 1 \text{ e } q \in K, \\ \eta_q & R_q = 1 \text{ e } q \notin K, \end{cases} \tag{1.14}$$

L'Equazione (1.13) è l'applicazione del teorema di Bayes, dove la probabilità a posteriori $\mathcal{L}_{n+1}(K)$ viene calcolata sulla base della probabilità a priori $\mathcal{L}_n(K)$ dello stato K e sulla base dell'evidenza empirica data dalla risposta osservata all'item q.

Infine, la *regola di stop* specifica il criterio di terminazione c della procedura di assessment, con $0 < c < 1$ (in Figura 1.2 rappresentato dal rombo). La procedura di valutazione termina quando la verosimiglianza di uno stato supera il criterio c, ovvero quando $\max(\mathcal{L}_n) > c$. A questo punto, lo stato di conoscenza la cui verosimiglianza ha raggiunto per prima il criterio di terminazione sarà lo stato assegnato allo studente. In linea di principio la scelta del criterio c è arbitraria. Tipicamente è consigliabile scegliere un criterio

superiore a .5, in modo tale che ci sia un solo stato ad avere la massima verosimiglianza. Inoltre, maggiore è il criterio di terminazione più accurato sarà il risultato della procedura di assessment.

Per fare un esempio, si supponga di voler individuare lo stato di conoscenza di uno studente nel dominio di conoscenza $Q = \{a, b, c, d\}$, composto da 4 item. Si supponga inoltre che la struttura di conoscenza associata al dominio Q sia

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, b, c\}, Q\}.$$

Si conoscono inoltre le probabilità di distrazione $\beta_a = .05$, $\beta_b = .07$, $\beta_c = .11$, $\beta_d = .06$ e le probabilità di lucky guess $\eta_a = .10$, $\eta_b = .09$, $\eta_c = .07$, $\eta_d = .02$ di ciascun item. Volendo simulare una procedura adattiva non deterministica continua, si considerino le seguenti regole: (1) regola di interrogazione *half-split*; (2) regola di aggiornamento bayesiano, descritta dall'Equazione (1.13); (3) regola di stop con criterio di terminazione $c = .80$.

La Figura 1.3 rappresenta il grafico relativo all'aggiornamento delle verosimiglianze nei diversi step della procedura. In assenza di informazione, la distribuzione di verosimiglianza iniziale $\mathcal{L}_0$ sarà uniforme e di conseguenza ciascuno stato di conoscenza ha una probabilità pari a $1/|K| = 1/6$ (diagramma in alto a sinistra della Figura). Procedendo secondo l'algoritmo illustrato in Figura 1.2, allo step $n = 1$ la regola di interrogazione deve scegliere l'item da somministrare allo studente. Dal momento che entrambi gli item a e b minimizzano la differenza assoluta riportata in Equazione (1.12), si procede a una selezione casuale di uno dei due. Si supponga che la procedura scelga di somministrare l'item a e che lo studente risponda correttamente all'item ($R_a = 1$). La verosimiglianza $\mathcal{L}_1$ viene aggiornata, per ciascuno stato, mediante l'applicazione dell'Equazione (1.13) ed assume la forma rappresentata nel digramma in alto a destra della Figura 1.3. Ciò che si osserva è un aumento della verosimiglianza di tutti gli stati che contengono l'item a e una diminuzione della verosimiglianza degli stati che non contengono l'item a. Allo step $n = 2$ la regola di interrogazione seleziona l'item b, lo studente

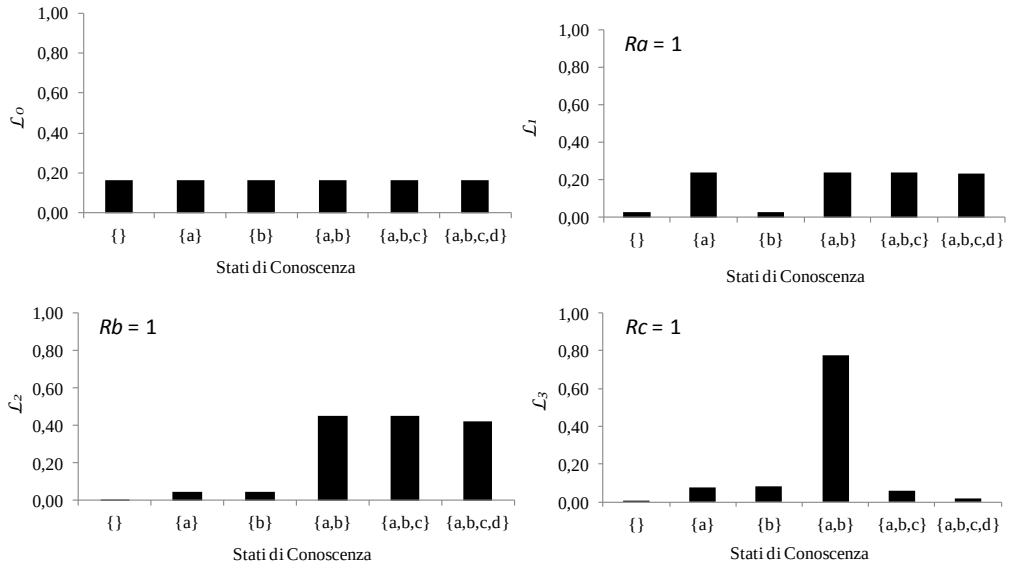

Figura 1.3: Aggiornamento della verosimiglianza degli stati di conoscenza di $\mathcal{K}$, secondo la procedura adattiva non deterministica continua.

risponde correttamente ($R_b = 1$) e la verosimiglianza viene aggiornata nuovamente ottenendo $\mathcal{L}_2$ (pannello in basso a sinistra della Figura 1.3). Infine allo step $n = 3$, viene selezionato l'unico item disponibile, ovvero c, lo studente risponde in modo errato ($R_c = 0$) e la verosimiglianza degli stati $\mathcal{L}_3$ viene aggiornata come illustrato nel pannello in basso a destra della figura. A questo punto, la verosimiglianza dello stato $\{a, b\}$ raggiunge il criterio di terminazione $c = .80$ e la procedura termina. Lo stato di conoscenza attribuito allo studente sarà pertanto $\{a, b\}$. Si vuole far notare che, in questo esempio giocattolo sono stati somministrati tutti gli item del dominio, ma in circostanze reali, dove il numero di item è molto grande, l'assessment termina, tipicamente, dopo aver somministrato una piccola parte dell'intero dominio di conoscenza.

La procedura discreta e quella continua sono state confrontate da Hockemeyer (2002) in una serie di simulazioni che avevano lo scopo di valutare l'accuratezza e l'efficienza delle due procedure. L'accuratezza è stata calcolata come proporzione di stati di conoscenza individuati correttamente, mentre l'effi-

cienza è stata misurata tenendo conto del numero di item necessari all'individuazione dello stato. Ne è emersa la superiorità della procedura continua in entrambi i casi.

1.5 Applicazioni della KST

Un esempio particolarmente importante delle applicazioni della KST è `ALEKS`, acronimo di Assessment and LEarning in Knowledge Spaces (www.aleks.com). `ALEKS` è un sistema di tutoring internet-based per la valutazione adattiva della conoscenza che include corsi in matematica, economia, statistica e chimica. Nato nel 1994 grazie a un finanziamento ottenuto dal National Science Foundation (USA) `ALEKS`, è stato acquistato dalla McGrow-Hill Education nel 2013 (www.wikipedia.org/wiki/ALEKS). Ad oggi è utilizzato da centinaia di migliaia di studenti americani, di tutti i livelli scolastici, per la valutazione delle proprie conoscenze.

Per quanto riguarda il versante europeo, la KST è stata utilizzata per lo sviluppo di due sistemi computerizzati game-based, chiamati *ELEXTRA* e *80-days*. Il primo è stato sviluppato per l'apprendimento della fisica. Il giocatore deve salvare due personaggi e per riuscire a farlo deve acquisire alcune nozioni di fisica aiutato dal fantasma di Galileo. Le azioni del giocatore forniscono al sistema informazioni relative al progredire delle sue conoscenze. Il sistema *80-days* invece, è stato sviluppato per l'apprendimento della geografia. Il giocatore deve aiutare un alieno a raccogliere informazioni sulla geografia del pianeta Terra, viaggiando a bordo di un UFO. Durante le esplorazioni il giocatore acquisisce informazioni di natura geografica.

Per quanto riguarda il versante italiano, non esistono, ad oggi, applicazioni simili della teoria.

1.6 Discussione

In questo capitolo sono stati presentati i concetti principali che stanno alla base della knowledge space theory. Tali concetti sono a fondamento delle ricerche presentate nei Capitoli 3, 4 e 5. L'obiettivo generale che ha fatto da sfondo a queste tre ricerche, è stato quello di approfondire gli aspetti probabilistici della teoria, rispondendo ad alcune domande che erano ancora aperte.

Nel Capitolo 3 si presentano gli sviluppi formali che consentono di *stimare la varianza dei parametri del* BLIM. Nel Capitolo 4 si presenta una procedura per *testare* una delle assunzioni su cui si basa il BLIM, ovvero *l'assunzione di invarianza dei parametri d'errore dagli stati di conoscenza.* Infine nel Capitolo 5 si presentano due *estensioni del* BLIM *ai dati mancanti.*

Tutte e tre queste ricerche sono un chiaro contributo per l'avanzamento delle conoscenze degli aspetti probabilistici della knowledge space theory, che conferiscono ulteriore validità alle sue applicazioni empiriche.

Capitolo 2

La Competence-based Knowledge Space Theory

2.1 Introduzione

Fin dalle sue origini, la KST è stata sviluppata seguendo un approccio *comportamentale*. Infatti, come visto nel capitolo 1, essa rappresenta la conoscenza di un individuo come l'*insieme di problemi* che è capace di risolvere in un determinato dominio di conoscenza. Questo approccio ha portato ad applicazioni di successo, si pensi ad esempio ad ALEKS, dove l'interesse era esclusivamente quello di individuare lo stato di conoscenza di un individuo.

Ciò non toglie che sia possibile collegare questo livello, che seguendo l'approccio di Korossy (1993, 1997, 1999) chiameremo *livello performance*, con un livello differente, che tiene conto delle abilità necessarie per risolvere quell'insieme di problemi. Chiameremo questo livello, *livello delle competenze* (Korossy, 1993, 1997, 1999). Così come avviene nei modelli psicometrici che tengono conto di variabili latenti, dalle risposte osservate a un insieme di item sarebbe possibile predire l'insieme delle abilità utilizzate per la loro risoluzione.

Il contesto formale sul quale è stata sviluppata la KST contiene tutti gi

elementi necessari all'integrazione dei due livelli. La teoria infatti è stata estesa da un certo numero di autori con l'obiettivo di fornire un'interpretazione cognitiva degli stati di conoscenza. I contributi maggiori in questa direzione sono stati dati da Falmagne, Koppen, Villano, Doignon, e Johanessen (1990), Doignon (1994), Düntsch e Gediga (1995), Korossy (1997), Gediga e Düntsch (2002). Recentemente, Heller, Ünlü, e Albert (2013) hanno contribuito allo sviluppo di un approccio unificato che integra i contributi citati sopra in un'unica teoria, chiamata *competence-based* knowledge space theory (CbKST).

In questo capitolo verranno presentate le nozioni alla base di questa teoria, che consentono di operazionalizzare la connessione tra il livello performance e quello delle competenze.

2.2 Assegnare le Abilità agli Item: Skill-multimap e Skill Function

L'idea alla base della CbKST è di individuare l'insieme di abilità s che sono necessarie per la risoluzione di un item q. Tali abilità sarebbero dunque in grado di spiegare in modo completo e univoco il comportamento di risoluzione, fornito da uno studente, a un item. Per fare un esempio si consideri il problema seguente

> *Calcolare il numero di combinazioni di 5 elementi presi 3 per volta.*

Per risolvere questo problema sulle combinazioni semplici, si può ipotizzare che siano necessarie le seguenti 3 abilità: (a) conoscere la formula del coefficiente binomiale; (b) calcolare il fattoriale di un numero; (b) risolvere frazioni con i fattoriali. Infatti, per poter arrivare alla soluzione

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

è necessario utilizzare contemporaneamente tutte e tre queste abilità. Se uno studente risponde correttamente a questo problema, a meno di lucky guess, siamo ragionevolmente certi che quello studente possiede le abilità a, b e c.

L'esempio considera solamente un item. Ovviamente questa idea va generalizzata al caso di un insieme di item. Quest'ultimo aspetto viene formalizzato attraverso la definizione di skill-multimap.

Definizione 6. Una *skill-multimap* è una terna (Q, S, μ), dove:

- Q è un insieme di problemi;
- S è un insieme di abilità;
- μ è una funzione da Q a 2^{2^S},

tale per cui ciascuna $\mu(q)$ è una collezione non vuota di sottoinsiemi non vuoti di S. Gli elementi $C \in \mu(q)$ sono chiamati competenze.

La Definizione 6 rappresenta il caso più generale possibile in cui un insieme di item può essere associato a un insieme di abilità, assumendo infatti che ogni item può essere risolto da almeno un sottoinsieme di abilità. Da un punto di vista educativo ciò significa che anche il caso in cui lo stesso item può essere risolto utilizzando strategie diverse di soluzione è contemplato. Vi sono poi due casi speciali degni di nota. Il primo è rappresentato dalla *skill-multimap congiuntiva*, nella quale, a ciascun item, viene assegnato esattamente un sottoinsieme non vuoto di abilità. Formalmente, per ogni item q: $\mu(q) = \{C\}$, con $\emptyset \subset C \subseteq S$. E' questo il caso dell'esempio visto sopra, in cui per risolvere correttamente l'item era necessario conoscere tutte le abilità ad esso associate. Vi è però un altro caso, ovvero quello delle *skill-multimap disgiuntive*, nelle quali ciascun sottoinsieme di abilità associate all'item è sufficiente per risolverlo. Formalmente, per ogni $q \in Q$: $\mu(q) = \{\{s\} : s \in M\}$, con $\emptyset \subset M \subseteq S$. In generale, una skill-multimap arbitraria può non appartenere né all'uno né all'altro caso.

Una rappresentazione più economica della skill-multimap, che elimina tutte le ridondanze, è la *skill function*. Essa si definisce esattamente come la prima, ma con l'ulteriore condizione che le competenze $C \in \mu(q)$ sono fra loro tutte incomparabili. In altre parole, in una skill function ciascuna competenza C è minimamente sufficiente per risolvere l'item q.

Per comprendere la differenza tra skill-multimap e skill function si consideri il seguente esempio sull'insieme $Q = \{a, b, c, d\}$ di 4 item e l'insieme $S = \{s, t, u\}$ di 3 abilità. Sia μ una skill-multimap definita come segue:

$$\begin{aligned} &\mu(a) = \{\{s,t\},\{s,u\}\} && \mu(b) = \{\{u\},\{s,u\}\} \\ &\mu(c) = \{\{s\},\{t\}\} && \mu(d) = \{\{t\}\}. \end{aligned} \tag{2.1}$$

In accordo con questa definizione è possibile affermare che:

- per risolvere l'item a uno studente deve possedere o la coppia di abilità $\{s, t\}$ o la coppia $\{s, u\}$. I due insiemi di abilità rappresentano due strategie alternative di risoluzione;
- l'item b viene risolto correttamente in due casi: se lo studente possiede l'abilità u o se possiede la coppia di abilità $\{s, u\}$.
- per risolvere l'item c è sufficiente possedere l'abilità s oppure l'abilità t;
- l'item d viene risolto correttamente se si possiede l'abilità t;

Si osserva che per l'item b la skill-multimap μ non soddisfa la condizione di incomparabilità, dal momento che le sue competenze sono annidate: $\{u\} \subseteq \{s, u\}$. E' possibile eliminare questa ridondanza, definendo una skill function $\tilde{\mu}$ che è una riduzione di μ. Esistono due skill function diverse per ridurre la skill-multimap μ: la skill function $\tilde{\mu}_1$ in cui $\tilde{\mu}_1(b) = \{\{u\}\}$ e la skill function $\tilde{\mu}_2$ in cui $\tilde{\mu}_2(b) = \{\{s, u\}\}$. Come si capirà nella sezione successiva esiste una sola delle due scelte è coerente con la skill multi-map μ in un senso ben preciso.

2.3 Stato di Competenza e Struttura di Competenza

Come visto nella sezione precedente, la skill function permette di collegare il *livello di performance* con il *livello delle competenze*. Questi due livelli forniscono due modi differenti per caratterizzare la conoscenza di uno studente. A livello performance, la conoscenza è rappresentata da un sottoinsieme $K \subseteq Q$ di tutti gli item che uno studente è capace di risolvere. A livello delle competenza è rappresentata da un sottoinsieme $C \subseteq S$ di tutte le abilità in S che lo studente possiede. Secondo questa distinzione, ogni studente è caratterizzato da una coppia (K, C), dove K è lo *stato di performance* (o stato di conoscenza) e C è lo *stato di competenza*. Quando si definisce una skill-multimap μ, lo stato di performance K è completamente deducibile dallo stato di competenza C se, per ogni item $q \in Q$, si applica la regola:

q appartiene a K se e solo se C include almeno una delle competenze in $\mu(q)$.

Questa regola si applica attraverso una funzione $p : 2^S \to 2^Q$ tale che, per ogni stato di competenza $C \in 2^S$,

$$p(C) = \{q \in Q : M \subseteq C \text{ per qualche } M \in \mu(q)\}. \tag{2.2}$$

Si può notare che p conserva l'ordine rispetto all'inclusione insiemistica, e che $p(\emptyset) = \emptyset$ e $p(S) = Q$. Una funzione che soddisfa queste proprietà è nota come *problem function*. Düntsch e Gediga (1995) hanno dimostrato che, per insiemi Q e S finiti, esiste una funzione biunivoca tra la famiglia di tutte le skill function μ e quella di tutte le problem function p.

Se p è la problem function corrispondente alla skill function μ, e $C \in 2^S$ è uno stato di competenza, allora $K = p(C)$ è lo stato di di conoscenza *delineato* da C attraverso la skill function μ. La collezione

$$\mathcal{K}_p = \{p(C) : C \in 2^S\}$$

di tutti gli stati di conoscenza delineati dagli stati di competenza in 2^S formano la *struttura di conoscenza* delineata da 2^S attraverso μ. Se μ è congiuntiva, allora $\mathcal{K}_p$ è chiusa all'intersezione. Se μ è disgiuntiva, allora $\mathcal{K}_p$ è chiusa all'unione e sarà pertanto uno spazio di conoscenza.

A questo proposito Doignon e Falmagne (1999) hanno dimostrato formalmente che ogni struttura di conoscenza costruita su un insieme Q è delineata da almeno una skill function (Q, S, μ). Questo implica che, ogni volta che nelle applicazioni empiriche si considera una certa struttura di conoscenza, anche la CbKST può essere applicata.

Per fare un esempio, si consideri la skill-multimap μ in (2.1), vista nella sezione precedente. E' possibile delineare la struttura di conoscenza $\mathcal{K}_p$ da μ, attraverso la regola della problem function definita in Equazione (2.2). Nella seconda colonna della Tabella 2.1 sono indicati tutti gli stati $K \in \mathcal{K}_p$ delineati dagli stati di competenza $C \in \mathcal{C}$.

Le terza e la quarta colonna della Tabella 2.1 illustrano invece, le strutture di conoscenza $\mathcal{K}'_p$ e $\mathcal{K}''_p$ delineate, rispettivamente, dalle skill function $\tilde{\mu}_1$ e $\tilde{\mu}_2$, che corrispondo ai due modi possibili di ridurre la skill-multimap μ. Come si

$\mathcal{C}$	$\mathcal{K}_p$	$\mathcal{K}'_p$	$\mathcal{K}''_p$
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\{s\}$	$\{c\}$	$\{c\}$	$\{c\}$
$\{t\}$	$\{c,d\}$	$\{c,d\}$	$\{c,d\}$
$\{u\}$	$\{b\}$	$\{b\}$	$\emptyset$
$\{s,t\}$	$\{a,c,d\}$	$\{a,c,d\}$	$\{a,c,d\}$
$\{s,u\}$	$\{a,b,c\}$	$\{a,b,c\}$	$\{a,b,c\}$
$\{t,u\}$	$\{b,c,d\}$	$\{b,c,d\}$	$\{c,d\}$
$\{s,t,u\}$	$\{a,b,c,d\}$	$\{a,b,c,d\}$	$\{a,b,c,d\}$

Tabella 2.1: Strutture di conoscenza $\mathcal{K}_p$, $\mathcal{K}'_p$ e $\mathcal{K}''_p$ delineate rispettivamente dalle skill-multimap μ2.1e dalle relative skill function $\tilde{\mu}_1$ e $\tilde{\mu}_2$.

può notare $\mathcal{K}_p = \mathcal{K}'_p \neq \mathcal{K}''_p$. Dovendo dunque scegliere una skill function che

sia allo stesso tempo una riduzione della skill-multimap μ e per la quale sia delineata la stessa struttura di conoscenza, la scelta cadrà sulla skill function $\tilde{\mu}_1$, infatti $\mathcal{K}_p = \mathcal{K}'_p$. Si osservi che la skill function $\tilde{\mu}_1$ è stata ottenuta da μ eliminando l'insieme più grande, tra quelli associati all'item b. L'insieme $\{u\}$ è infatti minimamente sufficiente per risolvere l'item.

Si vuole concludere questa sezione con un'ultima osservazione. Fino ad ora si è assunto implicitamente che ogni sottoinsieme di S sia uno stato di competenza. Così facendo la collezione degli stati di competenza sarebbe l'insieme potenza di 2^S. In certe situazioni però, questo potrebbe non essere vero, e la collezione degli stati di competenza potrebbe essere un sottoinsieme $\mathcal{C}$ dell'insieme potenza 2^S. Infatti, in contesti reali, è possibile trovare dipendenze di tipo logico o pedagogico tra le abilità, che portano all'esclusione di un certo numero di sottoinsiemi di S dalla collezione degli stati di competenza. Si supponga per esempio che l'abilità $s \in S$ non possa essere appresa prima dell'abilità $t \in S$. In questo caso, tutti i sottoinsieme di S che contengono l'abilità s ma non l'abilità t dovrebbero essere esclusi dalla collezione degli stati di competenza.

Secondo l'approccio che si segue nella presente tesi, in linea con Korossy (1993, 1997, 1999), si considera una *struttura di competenza* come la collezione $\mathcal{C} \subseteq 2^S$ di tutti gli stati di competenza che contengono almeno l'insieme vuoto e l'insieme totale.

2.4 Discussione

In questo capitolo sono stati riassunti gli aspetti deterministici della *competence-base* knowledge space theory (CbKST), un approccio che include nella teoria caratteristiche di tipo psicologico. L'obiettivo principale della CbKST infatti, è quello di fornire un'interpretazione cognitiva dello stato di conoscenza di uno studente, attraverso la definizione di uno *stato di competenza*, ovvero l'insieme delle abilità che possiede uno studente. La conoscenza di un indivi-

duo sarebbe dunque caratterizzata da uno stato di conoscenza, se si considera il *livello performance* e da uno stato di competenza, se si considera il *livello delle competenze*.

La CbKST fornisce gli aspetti formali per collegare questi due livelli, e lo fa attraverso i concetti di skill-function e di problem function. La prima è una funzione che collega un insieme di item a un insieme di abilità, mentre la seconda è una funzione che permette di delineare la struttura di conoscenza attraverso la skill-function. Infine è stato introdotto il concetto di struttura di competenza, cioè la collezione di tutti gli stati di competenza osservabili in una popolazione di studenti.

Questi concetti costituiscono la base teorica di riferimento per le ricerche che si presentano nei Capitoli 6 e 7 della tesi. In particolare, nel Capitolo 6 si presenterà un modello probabilistico per la validazione empirica di strutture di competenza. Questa ricerca risponde alla necessità di sviluppare gli aspetti probabilistici della CbKST. Nel Capitolo 7 si discute di una mancanza della corrispondenza biunivoca tra stato di competenza e stato di conoscenza, e si fornisce un contributo teorico a questo problema, nel caso in cui l'interesse sia quello di individuare lo stato di competenza di uno studente a partire dalle risposte osservate a un insieme di item. Questo contributo è fondamentale nel caso in cui si voglia costruire un intelligen tutoring system basato sulla CbKST.

Parte II

Sviluppo di Procedure e Modelli: Contributi

Capitolo 3

Procedure Analitiche per la Stima della Varianza dei Parametri del BLIM

Da quando il BLIM è stato introdotto per la prima volta, molte domande riguardo la sua applicabilità hanno trovato una risposta. Sono stati sviluppati alcuni metodi per stimare i suoi parametri (Heller & Wickelmaier, 2011; Stefanutti & Robusto, 2009) e per testare la sua identificabilità (Spoto et al., 2012; Stefanutti et al., 2012). Ciò nonostante rimangono aperte altre questioni molto importanti. In particolare, non si conosce la matrice di covarianza delle stime di massima verosimiglianza (in seguito chiamate stime ML) dei parametri del modello. Questa matrice è necessaria sia per calcolare la varianza delle stime dei parametri sia per calcolare gli intervalli di confidenza. Oltre agli aspetti di tipo applicativo, avere a disposizione le formule analitiche della varianza delle stime dei parametri, permette lo studio teorico del loro comportamento asintotico, sotto differenti condizioni. L'obiettivo della ricerca che si presenta in questo capitolo, è dunque quello di derivare queste matrici per le stime ML dei parametri del BLIM.

La matrice di covarianza è strettamente collegata ad altre due matrici,

chiamate matrice Hessiana e matrice dell'informazione di Fischer (si veda ad es. Lehmann, 1999; Lehmann & Casella, 1998). Più precisamente, se il metodo di stima dei parametri è per massima verosimiglianza, la matrice di covarianza si ottiene dall'inverso della matrice dell'informazione di Fischer che, sotto certe condizioni di regolarità, corrisponde al valore atteso della matrice Hessiana. Per arrivare dunque alla matrice di covarianza del BLIM si devono derivare la matrice dell'informazione di Fischer e la matrice Hessiana.

Il capitolo è organizzato come segue. Dopo aver introdotto alcune nozioni di base sulla matrice dell'informazione di Fischer e sulla corrispondente matrice di covarianza di un modello (Sezione 3.1), si derivano le formule analitiche per il calcolo di queste matrici nel caso del BLIM (Sezione 3.3). Nella Sezione 3.4 si presenta un metodo per il calcolo degli intervalli di confidenza dei suoi parametri. Nella Sezione 3.5 si illustra uno studio asintotico, in cui si analizza il comportamento della varianza dei parametri del BLIM, sotto alcune condizioni particolari. Infine, si discutono i risultati di un applicazione empirica (Sezione 3.6), dove la varianza e gli intervalli di confidenza delle stime dei parametri del BLIM sono state calcolate nell'applicazione del modello a dati reali.

3.1 L'Informazione di Fisher e la Matrice di Covarianza di un Modello

Queste sezione contiene il materiale relativo all'informazione di Fischer e alla matrice di Covarianza che si ottengono dalla stima per massima verosimiglianza dei parametri di un modello multinomiale. Queste nozioni saranno poi la base per derivare queste due matrici nel caso del BLIM, un modello multinomiale con particolari restrizioni.

Ricordando la definizione data nella Sezione 1.3.3, un modello parametrico è la terna (Θ, f, Ω). Se il modello è multinomiale, dato un campione di os-

servazioni X, e un vettore di parametri $\theta \in \Theta$, la funzione di verosimiglianza del modello è nota, e assume la forma:

$$L(X;\theta) = \frac{N!}{\prod_{x\in\Omega} F(x)!} \prod_{x\in\Omega} f_x(\theta)^{F(x)}, \tag{3.1}$$

dove $F(x)$ è la frequenza x osservata nel campione X, e $N = \sum_{x\in\Omega} F(x)$ è la numerosità campionaria. La stima per massima verosimiglianza dei parametri del modello consiste nel trovare il punto $\hat{\theta} \in \Theta$ che massimizza la funzione di verosimiglianza $L(X;\theta)$.

Oltre ad essere asintoticamente priva di bias, la stima per massima verosimiglianza (chiamata in seguito MLE, dall'inglese maximum likelihood estimetion) ha la proprietà fondamentale di essere lo stimatore che ottiene la varianza più piccola. La varianza di qualsiasi stimatore ha un limite inferiore, conosciuto come *limite di Kramér-Rao*, che dipende dalla curvatura attesa della log-verosimiglianza del punto $\hat{\theta}$, al crescere della numerosità campionaria. La stima per massima verosimiglianza si avvicina a questo limite inferiore.

Sotto certe condizioni di regolarità la curvatura attesa della verosimiglianza logaritmica $\log L(X;\theta)$, si misura dalla matrice dell'informazione di Fischer (FIM), che è la matrice simmetrica $\mathbf{I}_\theta$ in cui ogni elemento è

$$\mathbf{I}_{\theta(i,j)} = E\left[\left(\frac{\partial \log L(X;\theta)}{\partial \theta_i}\right)\left(\frac{\partial \log L(X;\theta)}{\partial \theta_j}\right)\Big|\theta\right], \tag{3.2}$$

dove, per ogni variabile casuale Z che dipende dai parametri θ, $E(Z|\theta)$ denota l'attesa condizionale di Z nel punto θ. Le condizioni di regolarità sono (Lehmann, 1999, p. 456):

(C1) il modello è identificabile, ovvero $f(\theta) = f(\theta')$ implica $\theta = \theta'$ per ogni $\theta, \theta' \in \Theta$;

(C2) lo spazio parametrico Θ è un insieme aperto;

(C3) le osservazioni $X_i, i = 1, 2, \ldots, N$ sono indipendenti e identicamente distribuite con probabilità $f_{X_i}(\theta)$;

(C4) l'insieme $A = \{x \in \Omega : f_x(\theta) > 0\}$ è indipendente da θ;

(C5) per ogni $x \in A$ esistono le derivate parziali

$$\frac{\partial f_x(\theta)}{\partial \theta_j}, j = 1, 2, \ldots, m$$

dove m è il numero di parametri del modello.

Una stima della matrice di covarianza nel punto $\hat{\theta}$ è data dall'inversa della FIM: $\mathbf{C}_{\hat{\theta}} = \mathbf{I}_{\hat{\theta}}^{-1}$. La varianza delle stime ML $\hat{\theta}$ è la diagonale principale della matrice di covarianza $\mathbf{C}_{\hat{\theta}}$.

Il calcolo della FIM secondo l'Equazione (3.2) può essere molto difficile da ottenere, dato che ricavare la formula analitica del valore atteso del prodotto di due derivate non è banale. Ciò nonostante, sotto la condizione aggiuntiva

(C6) $f(\theta)$ è due volte differenziabile,

ciascun elemento della FIM assume la forma (Lehmann & Casella, 1998)

$$\mathbf{I}_{\theta(i,j)} = -E\left[\frac{\partial^2 \log L(X;\theta)}{\partial \theta_i \partial \theta_j} \;\middle|\; \theta\right]. \tag{3.3}$$

Il calcolo di (3.3) è solitamente più semplice di quello di (3.2).

Una proprietà estremamente utile della FIM riguarda la sua trasformazione a seguito di una riparametrizzazione del modello.

Definizione 7. Sia (Θ, f, Ω) un modello parametrico. Si consideri ora un nuovo spazio parametrico $\Psi \subseteq \Re^k, k > 0$, e una funzione differenziabile continua $g : \Psi \to \Theta$, allora il modello $(\Psi, f \circ g, \Omega)$ è una *riparametrizzazione* di (Θ, f, Ω).

Lehmann (1999) e Lehmann e Casella (1998) hanno dimostrato che la FIM della riparametrizzazione $f \circ g$ nel punto $\psi \in \Psi$, è data dalla trasformazione

$$\mathbf{I}_{\psi} = \mathbf{J}_{\psi}^T \mathbf{I}_{g(\psi)} \mathbf{J}_{\psi}, \tag{3.4}$$

dove:

- $\mathbf{J}_\psi$ è la matrice Jacobiana della funzione g nel punto ψ, cioè la matrice delle derivate prime parziali di g rispetto ai parametri ψ:

$$\mathbf{J}_\psi = \left\{ \frac{\partial g_i(\psi)}{\partial \psi_j} \right\}.$$

  Questa matrice ha un numero di righe pari al numero di parametri liberi di variare del modello originale f, e ha un numero di colonne pari al numero di parametri liberi di variare nella riparametrizzazione $f \circ g$;

- $\mathbf{I}_{g(\psi)}$ è la matrice FIM nel punto $g(\psi)$, che appartiene allo spazio parametrico Θ del modello originale.

3.2 Lo Spazio Parametrico del BLIM

Con l'obiettivo di derivare l'informazione di Fisher e la matrice di covarianza del BLIM (oggetto della prossima sezione), è utile avere a disposizione una definizione formale di spazio parametrico e della funzione di predizione del modello. In questa sezione si riassumono brevemente i risultati ottenuti da Stefanutti et al. (2012).

Il numero complessivo delle dimensioni dello spazio parametrico del BLIM uguaglia il numero di parametri liberi del modello. Il BLIM si compone di tre tipi di parametri:

- una probabilità β_q per ciascun item $q \in Q$;

- una probabilità η_q per ciascun item $q \in Q$;

- una probabilità π_K per ogni stato della struttura di conoscenza $\mathcal{K}$.

Inoltre, fra le $|\mathcal{K}|$ probabilità π_K degli stati, solamente un numero $|\mathcal{K}| - 1$ sono libere di variare, dato il vincolo $\sum_{K \in \mathcal{K}} \pi_K = 1$. Per convenzione e senza perdita di generalità, si prende $\mathcal{K}^* = \mathcal{K} \setminus \{Q\}$, come la collezione di stati i cui parametri sono liberi di variare. Segue che il BLIM ha un numero totale di

parametri liberi di variare pari a $m = 2|Q| + |\mathcal{K}| - 1$, che corrisponde anche al numero di dimensioni dello spazio parametrico.

I parametri del BLIM devono soddisfare le seguenti restrizioni:

(R1) tutti gli stati in $\mathcal{K}$ devono avere una probabilità $\pi_K \neq 0$, dato che se uno stato avesse probabilità pari a zero, ciò corrisponderebbe ad eliminarlo dalla struttura;

(R2) i parametri β_q e η_q devono essere positivi per tutti gli item $q \in Q$, dal momento che se fossero uguali a zero, non ci sarebbe nessun parametro da stimare;

(R3) la disequazione $\beta_q + \eta_q < 1$ deve essere vera per tutti gli item $q \in Q$. Questo è dovuto al fatto che la probabilità di rispondere correttamente indovinando la risposta deve essere minore della probabilità di rispondere correttamente perché si conosce la risposta ($\eta < 1 - \beta$), e parallelamente la probabilità di sbagliare una risposta per distrazione deve essere minore della probabilità di rispondere in modo errato non conoscendo la risposta ($\beta < 1 - \eta$).

Siano $\beta = (\beta_q)_{q \in Q}$ e $\eta = (\eta_q)_{q \in Q}$ i vettori delle probabilità di careless error e lucky guess, rispettivamente. Sotto le restrizioni (R1), (R2) e (R3), lo spazio parametrico del BLIM diventa $\Theta = E \times \Pi$, dove:

- l'insieme E è il sottospazio parametrico delle probabilità β_q e η_q del modello, formalmente:

$$E = \left\{ (\beta, \eta) \in (0,1)^{2|Q|} \middle| \eta_q + \beta_q < 1 \text{ per ogni } q \in Q \right\};$$

- l'insieme Π è il sottospazio parametrico delle probabilità degli stati π_K del modello, formalmente

$$\Pi = \left\{ \pi \in (0,1]^{|\mathcal{K}^*|} \middle| \sum_{K \in \mathcal{K}^*} \pi_K < 1 \right\}.$$

Una singola osservazione nel BLIM è un pattern di risposta $R \subseteq Q$, quindi l'insieme di tutte le osservazioni possibili è l'insieme potenza $\mathcal{R} = 2^Q$. Tuttavia, dato che le probabilità dei pattern di risposta sommano a 1, ci sono solo $|\mathcal{R}|-1$ osservazioni indipendenti. Indicando con $\mathcal{R}^* = \mathcal{R} \setminus \{Q\}$ l'insieme delle osservazioni indipendenti, lo spazio degli esiti del BLIM prende la forma

$$\Omega = \left\{ p \in [0,1]^{|\mathcal{R}^*|} \Big| \sum_{R \in \mathcal{R}^*} p_R < 1 \right\}.$$

La funzione di predizione $f : \Theta \to \Omega$ mappa un punto $\theta \in \Theta$ in una distribuzione di probabilità $f(\theta) \in \Omega$ sui pattern di risposta. Indicando con $f_R(\theta)$ il singolo elemento di $f(\theta)$ corrispondente al pattern di risposta $R \in \mathcal{R}^*$, dall'Equazione (1.6), la funzione di predizione assume la forma

$$f_R(\theta) = \sum_{K \in \mathcal{K}} P(R|K)\pi_K. \tag{3.5}$$

Dal momento che una delle probabilità degli stati di conoscenza è ridondante (dato il vincolo della somma a 1), è conveniente riscrivere la funzione di predizione f nella forma *non-ridondante* seguente:

$$f_R(\theta) = \sum_{K \in \mathcal{K}^*} P(R|K)\pi_K + P(R|Q)\left(1 - \sum_{K \in \mathcal{K}^*} \pi_K\right).$$

Per quanto riguarda la matrice Jacobiana $\mathbf{J}_\theta$ della funzione di predizione del BLIM, è una matrice $n \times m$, dove $m = 2|Q| + |\mathcal{K}| - 1$ è il numero di parametri liberi del BLIM e $n = 2^{|Q|} - 1$ è il numero di pattern di risposta indipendenti. La matrice Jacobiana del BLIM derivata da Stefanutti et al. (2012), per ogni pattern di risposta $R \in \mathcal{R}^*$, si compone delle tre parti:

$$\frac{\partial f_R(\theta)}{\partial \pi_K} = P(R|K) - P(R|Q),$$

per ogni stato $K \in \mathcal{K}^*$;

$$\frac{\partial f_R(\theta)}{\partial \beta_q} = \begin{cases} \frac{1}{\beta_q}\left[\sum_{K \in \mathcal{K}_q^*} P(R|K)\pi_K + P(R|Q)\left(1 - \sum_{K \in \mathcal{K}^*} \pi_K\right)\right] & \text{se } q \notin R \\ -\frac{1}{1-\beta_q}\left[\sum_{K \in \mathcal{K}_q^*} P(R|K)\pi_K + P(R|Q)\left(1 - \sum_{K \in \mathcal{K}^*} \pi_K\right)\right] & \text{se } q \in R \end{cases}$$

per ogni parametro β_q, dove $\mathcal{K}^*_q = \{K \in \mathcal{K}^* : q \in K\}$;

$$\frac{\partial f_R(\theta)}{\partial \eta_q} = \begin{cases} \frac{1}{\eta_q} \sum_{K \in \bar{\mathcal{K}}^*_q} P(R|K)\pi_K & \text{if } q \in R \\ -\frac{1}{1-\eta_q} \sum_{K \in \bar{\mathcal{K}}^*_q} P(R|K)\pi_K & \text{if } q \notin R \end{cases}$$

per ogni parametro η_q, dove $\bar{\mathcal{K}}^*_q = \{K \in \mathcal{K}^* : q \notin K\}$.

3.3 La Matrice d'Informazione di Fisher del BLIM

In questa sezione si deriva l'informazione di Fisher per le stime ML dei parametri del BLIM. E' possibile derivare la matrice FIM, riportata in Equazione (3.3), se si dimostra che tutte le condizioni (C1)–(C6) sono rispettate dal modello. Si prendono ad esame una ad una.

Come evidenziato nella Sezione 3.2, sotto le restrizioni (R1)–(R3), lo spazio parametrico del BLIM è un insieme aperto. Questo significa che la condizioni (C2) è rispettata. Essendo poi, la funzione di predizione f del BLIM una funzione analitica, questo significa che anche le condizioni (C5) e (C6) sono soddisfatte dal modello, dal momento che una funzione analitica ammette sempre le derivate prime e le derivate seconde. La condizione (C3) è rispettata, in generale, dal campionamento multinomiale, che assicura l'indipendenza delle osservazioni.

Per quanto riguarda la condizione (C4), può essere riformulata come segue: il sottoinsieme $\mathcal{A} = \{R \in \mathcal{R}^* : f_R(\theta) > 0\}$ non dipende dalla scelta di θ. Date le restrizioni (R1)–(R3), non è difficile osservare che l'Equazione (1.6) assegna sempre una probabilità positiva ad ogni pattern di risposta R, indipendentemente dai parametri scelti per θ. Ciò significa che l'uguaglianza $\mathcal{A} = \mathcal{R}^*$ è sempre vera, indipendentemente da θ. Per questo motivo anche la condizione (C4) è rispettata dal BLIM.

Infine, la condizione (C1) è la più critica. Questo dipende dal fatto che, come evidenziato nella Sezione 1.3.3, il BLIM non è sempre identificabile, e

che la sua identificabilità dipende dalla particolare struttura di conoscenza $(Q, \mathcal{K})$ sulla quale il BLIM è stato applicato (Spoto et al., 2012; Stefanutti et al., 2012). Di conseguenza l'identificazione del modello non è rispettata in generale, ma deve essere adeguatamente testata per la particolare struttura di conoscenza in esame.

E' ora possibile derivare la matrice FIM relativa al BLIM. A questo scopo, si consideri la seguente proposizione.

Proposizione 1. *Data una qualsiasi numerosità campionaria $N > 0$, la matrice d'informazione di Fisher del BLIM, nel punto $\theta \in \Theta$ dello spazio parametrico è*

$$\mathbf{I}_\theta = \mathbf{J}'_\theta \mathbf{H}_{f(\theta)} \mathbf{J}_\theta,$$

dove: $\mathbf{J}_\theta$ è la matrice Jacobiana della funzione di predizione f del BLIM nel punto θ, e $\mathbf{H}_{f(\theta)}$ è una matrice simmetrica $|\mathcal{R}^*| \times |\mathcal{R}^*|$, in cui ogni riga (rispettivamente colonna) è indicizzata da un pattern di risposta $R_i \in \mathcal{R}^*$, e in ogni cella si trova

$$\mathbf{H}_{f(\theta)(ij)} = \begin{cases} N\left(\dfrac{1}{f_{R_i}(\theta)} + \dfrac{1}{f_Q(\theta)}\right) & \text{se } i = j \\ \dfrac{N}{f_Q(\theta)} & \text{se } i \neq j. \end{cases}$$

Dimostrazione. Si consideri il modello multinomiale saturo (SMM) i cui parametri liberi di variare sono le probabilità p_R dei pattern di risposta $R \in \mathcal{R}^*$ (ovvero $p_R > 0$ per ogni $R \in \mathcal{R}^*$, e$\sum_{R\in\mathcal{R}} p_R = 1$). Dato un campione X di numerosità N, la funzione di verosimiglianza del modello SMM è

$$L(X; p) = \frac{N!}{\prod_{R\in\mathcal{R}} F(R)!} \prod_{R\in\mathcal{R}^*} p_R^{F(R)} \left(1 - \sum_{R\in\mathcal{R}^*} p_R\right)^{F(Q)}, \tag{3.6}$$

mentre la funzione di predizione è l'identità $g(p) = p$, dove p è il vettore delle probabilità p_R. E' facile notare che il BLIM è la riparametrizzazione $g \circ f$ del modello SMM, in cui f è la funzione di predizione del BLIM. Pertanto, dall'Equazione (3.4), la matrice d'informazione di Fisher $\mathbf{I}_\theta$ del BLIM nel

punto $\theta \in \Theta$, si può ottenere da

$$\mathbf{I}_\theta = \mathbf{J}'_\theta \mathbf{H}_{f(\theta)} \mathbf{J}_\theta,$$

dove $\mathbf{J}_\theta$ è la matrice Jacobiana di f in θ, e $\mathbf{H}_{f(\theta)}$ è la matrice d'informazione di Fisher del modello SMM nel punto $f(\theta)$. Quest'ultima matrice si può ottenere calcolando il valore atteso della derivata seconda del logaritmo della funzione di verosiglianza (3.6). Quest'ultimo passaggio corrisponde all'applicazione dell'Equazione (3.3). In particolare, dati due pattern di risposta qualsiasi $R_i, R_j \in \mathcal{R}^*$, e ponendo $p = f(\theta)$, la FIM del BLIM è

$$\mathbf{H}_{p(ij)} = -E\left[\frac{\partial^2 \log L(X;p)}{\partial p_{R_i} \partial p_{R_j}} \;\middle|\; p\right] = \begin{cases} N\left(p_{R_i}^{-1} + p_Q^{-1}\right) & \text{if } i = j \\ N p_Q^{-1} & \text{if } i \neq j, \end{cases}$$

dove $p_Q = 1 - \sum_{R \in \mathcal{R}^*} p_R$. □

Va ricordato infine, che la matrice di covarianza $\mathbf{C}_{\hat{\theta}}$ per le stime ML dei parametri di un modello, si ottiene dall'inversa della FIM.

3.4 Intervalli di Confidenza

I vantaggi maggiori dell'avere a disposizione la matrice di covarianza per le stime ML dei parametri, si hanno nelle applicazione empiriche di un modello. Infatti, l'inferenza statistica, sia nella forma della verifica di ipotesi che nella forma degli intervalli di confidenza, diviene possibile.

In questa sezione, a partire dalla matrice di covarianza del BLIM, si costruiscono gli intervalli di confidenza per le stime ML dei suoi parametri d'errore β_q e η_q. Dal momento che nel BLIM ci si aspetta che questi parametri siano il più piccoli possibile (si veda per esempio, Stefanutti & Robusto, 2009), avere a disposizione gli intervalli di confidenza è particolarmente utile per una loro corretta interpretazione.

Una delle proprietà note della stima ML è la normalità. Se $\hat{\theta}$ è la stima ML del parametro vero θ, allora la sua distribuzione è normale con media θ e

varianza $var(\hat{\theta})$. Sotto queste condizioni i limiti dell'intervallo di confidenza per $\hat{\theta}$ si ottengono con

$$\hat{\theta} \pm \sqrt{var(\hat{\theta})}\, z_{1-\alpha/2}, \tag{3.7}$$

dove $z_{1-\alpha/2}$ è il $1-\alpha/2$ quantile della distribuzione standard normale. Ciò nonostante, se lo spazio parametrico ha dei limiti, come nel caso dei parametri β_q e η_q, la normalità viene rispettata solamente se i parametri cadono lontano da questi limiti.

Come accennato sopra i parametri d'errore del BLIM sono piccoli, e quindi vicini al limite inferiore del loro spazio parametrico, che è zero. In questa situazione dunque, la normalità delle stime ML verrà violata piuttosto spesso e la formula (3.7) risulterà inadeguata.

L'approccio che si segue consiste nel considerare una riparametrizzazione di β_q e η_q, in uno spazio parametrico senza limiti, e di calcolare i limiti dell'intervallo di confidenza per le stime ML dei nuovi parametri, con la Formula (3.7). Una "riparametrizzazione adeguata" è una qualsiasi trasformazione monotona strettamente crescente $h : \Re \to (0,1)$. Una trasformazione di questo tipo è, ad esempio, una funzione logistica definita da $h(x) = \exp(x)/[1+\exp(x)]$. I nuovi parametri saranno dunque i log-odds $\omega_q, \tau_q \in \Re$ dei parametri d'errore β_q e η_q che soddisfano le due equazioni:

$$\omega_q = \ln \frac{\beta_q}{1-\beta_q} \quad \text{e} \quad \tau_q = \ln \frac{\eta_q}{1-\eta_q}.$$

Date le stime ML $\hat{\omega}_q$ e $\hat{\tau}_q$ dei nuovi parametri, i limiti di confidenza per $\hat{\beta}_q$ e $\hat{\eta}_q$ sono quindi dati dalle due formule

$$h\left[\hat{\omega}_q \pm \sqrt{var(\hat{\omega}_q)} z_{1-\alpha/2}\right] \quad \text{e} \quad h\left[\hat{\tau}_q \pm \sqrt{var(\hat{\tau}_q)} z_{1-\alpha/2}\right].$$

Indicando con $\phi = (\omega, \tau, \pi)$ il vettore dei parametri del modello riparametrizzato, la matrice dell'informazione di Fischer di quest'ultimo si ottiene da

$$\mathbf{I}_\phi = \mathbf{J}_\phi^T \mathbf{I}_\theta \mathbf{J}_\phi,$$

dove $\mathbf{J}_\phi$ è la matrice delle derivate prime parziali dei parametri originali θ, rispetto ai nuovi parametri ϕ, ovvero una matrice diagonale i cui elementi della diagonale principale sono

$$\frac{\partial \beta_q}{\partial \omega_q} = \beta_q(1-\beta_q), \quad \frac{\partial \eta_q}{\partial \tau_q} = \eta_q(1-\eta_q) \text{ e } \quad \frac{\partial \pi_K}{\partial \pi_K} = 1.$$

Infine, la matrice di covarianza si ottiene dall'inverso di $\mathbf{I}_\phi$.

3.5 Studio Asintotico della Varianza dei Parametri del BLIM

La procedura per calcolare la FIM, e dunque la matrice di covarianza del BLIM, è stata implementata in un tool MATLAB. La procedura richiede in ingresso tre argomenti

- una struttura di conoscenza$(Q, \mathcal{K})$;
- un vettore di parametri θ;
- una numerosità campionaria $N > 0$.

L'obiettivo dello studio riguardava l'analisi del comportamento della varianza delle stime ML dei parametri d'errore del BLIM sotto differenti condizioni.

In tutte le condizioni è stata utilizzata una struttura $\mathcal{K}$, composta da 45 stati di conoscenza, su un insieme di 12 problemi. In particolare si voleva studiare l'andamento della varianza delle stime di β_q e η_q, nel caso di due tipi di item, chiamati *item target*, ovvero:

1. item che appartengono a molti stati di conoscenza (*item "bottom"*);
2. item che appartengono a pochi stati di conoscenza (*item "top"*).

Nella struttura $\mathcal{K}$ è stato quindi individuato un item di ciascuna categoria: l'item "bottom" (item 4), apparteneva a 26 stati di conoscenza, mentre l'item "top" (item 5), apparteneva solamente a 4 stati. Ogni altro item di

$\mathcal{K}$ apparteneva a un numero di stati maggiore o uguale 4, oppure minore o uguale a 26.

Sono stati poi considerati 4 valori differenti per i parametri β ed η degli item target, ovvero $\beta_t, \eta_t \in \{.010, .037, .136, .500\}$. I parametri β_q e η_q degli item non target sono invece stati fissati, in tutte le condizioni, a valori casuali scelti nell'intervallo $(0, .10)$. Per quanto riguarda le probabilità degli item target, sono stati scelti 50 valori crescenti, scelti nell'intervallo aperto $(0, 1)$. La probabilità $P(q)$ di tutti gli altri item $q \in Q$, è stata infine calcolata come la somma delle probabilità di tutti gli stati contenenti l'item. Formalmente:

$$P(q) = \sum_{K \in \mathcal{K}_q} \pi_K. \tag{3.8}$$

Questa probabilità non può essere manipolata direttamente, dal momento che non è un parametro libero del modello. Essa è invece una funzione lineare dei parametri π_K. Per questo motivo, $P(q)$ è stata manipolata indirettamente nel modo seguente: (1) è stato generato casualmente il vettore π_K degli stati di conoscenza; (2) su questo vettore, è stata applicata una trasformazione lineare che soddisfa la condizione $\sum_{K \in \mathcal{K}_t} \pi_K = P(t)$, dove $P(t)$ è il valore di probabilità scelto per l'item target. Questa procedura è stata applicata per ciascuno dei 50 valori $P(t)$ scelti per gli item target t. Si è deciso infine di considerare due numerosità campionarie differenti $N \in \{200, 1000\}$. Riassumendo, sono stati manipolati:

1. gli item target;
2. i valori dei parametri β_t e η_t;
3. la probabilità $P(t)$ degli item target;
4. la numerosità campionaria.

Utilizzando la procedura implementata in MATLAB, la matrice di covarianza del BLIM è stata calcolata per ognuna delle $2 \times 4 \times 50 \times 2 = 800$ condizioni differenti.

3.5.1 Risultati

I risultati relativi al parametro di careless error dell'item "bottom" (item 4) sono illustrati nei due grafici in alto della Figura 3.1. In entrambi i grafici si rappresenta la varianza del parametro β_4 (asse y), in funzione della probabilità dell'item 4 (asse x). Ognuna delle quattro curve corrisponde a un valore diverso del parametro β_4. Dalla due grafici si può osservare che la

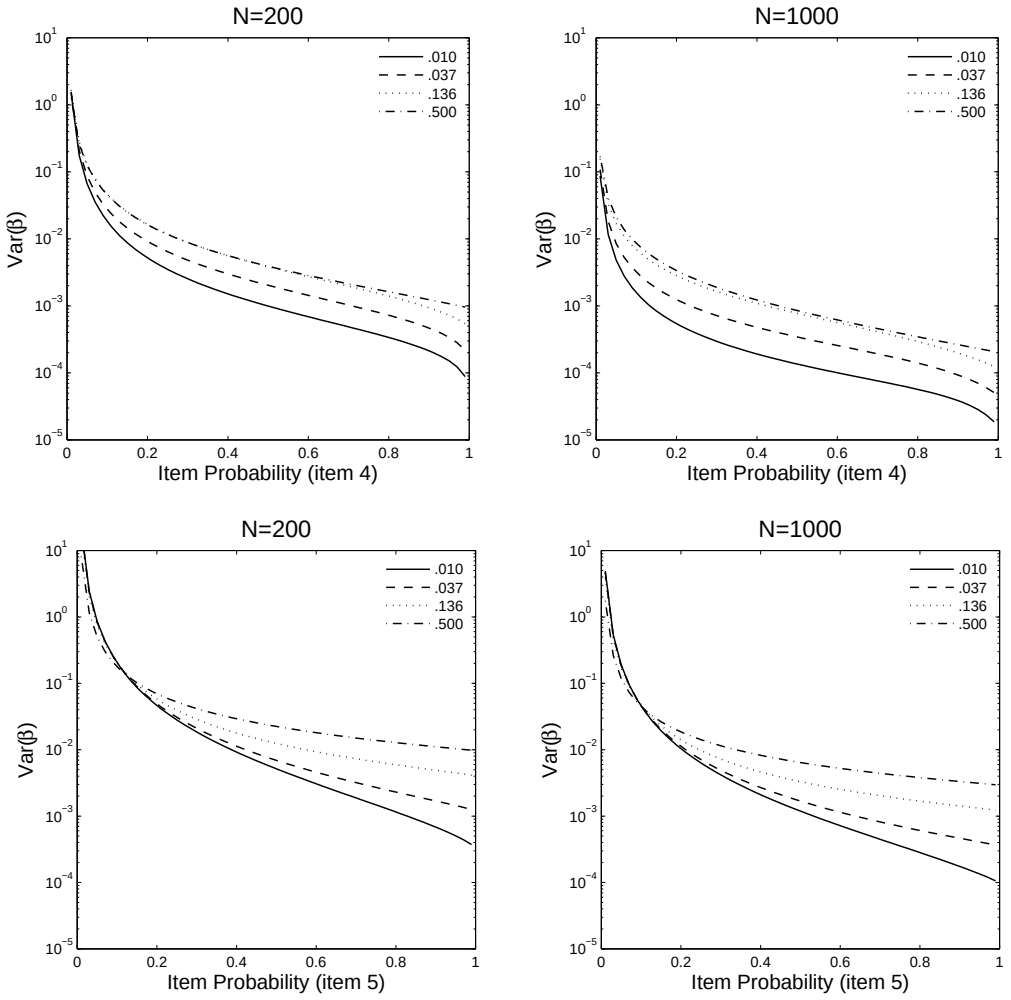

Figura 3.1: Varianza di $\hat{\beta}_q$ per l'item *"bottom"* (pannelli in alto) e per l'item *"top"* (pannelli in basso) in funzione della loro probabilità (x-axis) e del valore vero di β_q. L'asse y è in scala logaritmica.

varianza del parametro di careless error è monotona decrescente rispetto alla

probabilità dell'item. In particolare, valori bassi della probabilità dell'item corrispondono a valori elevati della varianza del parametro, indipendentemente dalla numerosità campionaria considerata ($N = 200$ nel pannello di sinistra e $N = 1000$ nel pannello di destra). Un confronto tra i due pannelli conferma semplicemente che la varianza diminuisce al crescere della numerosità campionaria.

I risultati relativi al parametro β_5 (item *"top"*), sono illustrati nei due pannelli in basso della Figura 3.1. Le considerazioni fatte per il parametro careless error dell'item "bottom" si applicano anche all'item "top", con la sola differenza che la varianza di β_q è sistematicamente più alta per l'item "top".

Per quanto riguarda il parametro di lucky guess dei due item, si osserva una tendenza opposta a quella del parametro di careless error. I due diagrammi in alto della Figura 3.2, illustrano i risultati dell'item "bottom", mentre i due diagrammi in basso della stessa figura, illustrano quelli dell'item "top". Come si può vedere, la varianza del parametro è monotona crescente rispetto alla probabilità dell'item. Questo significa che, valori elevati della probabilità dell'item corrispondono a valori elevati della varianza dei parametri, indipendentemente dall'item target e dalla numerosità campionaria considerati. Ciò che cambia tra i due tipi di item è che la varianza è sistematicamente più alta per l'item "top" (diagrammi in basso di Figura 3.2), se confrontata con quella dell'item "bottom" (diagrammi in alto di Figura 3.2).

Nelle figure esaminate fino ad ora, l'effetto dei valori dei parametri d'errore non può essere letto in modo accurato. Per questa ragione nella Figura 3.3 la varianza di $\hat{\beta}_q$ è stata rappresentata come funzione del parametro β_q dell'item "top", con $N = 200$. Le quattro curve del grafico si riferiscono a un valore diverso della probabilità dell'item, scelta fra $\{.010, .500, .745, .990\}$. In tutte e quattro le curve, il valore massimo dell'ascissa si avvicina a .5 e, allontanandosi da esso verso i due limiti zero e uno, la varianza dei parametri diminuisce monotonicamente.

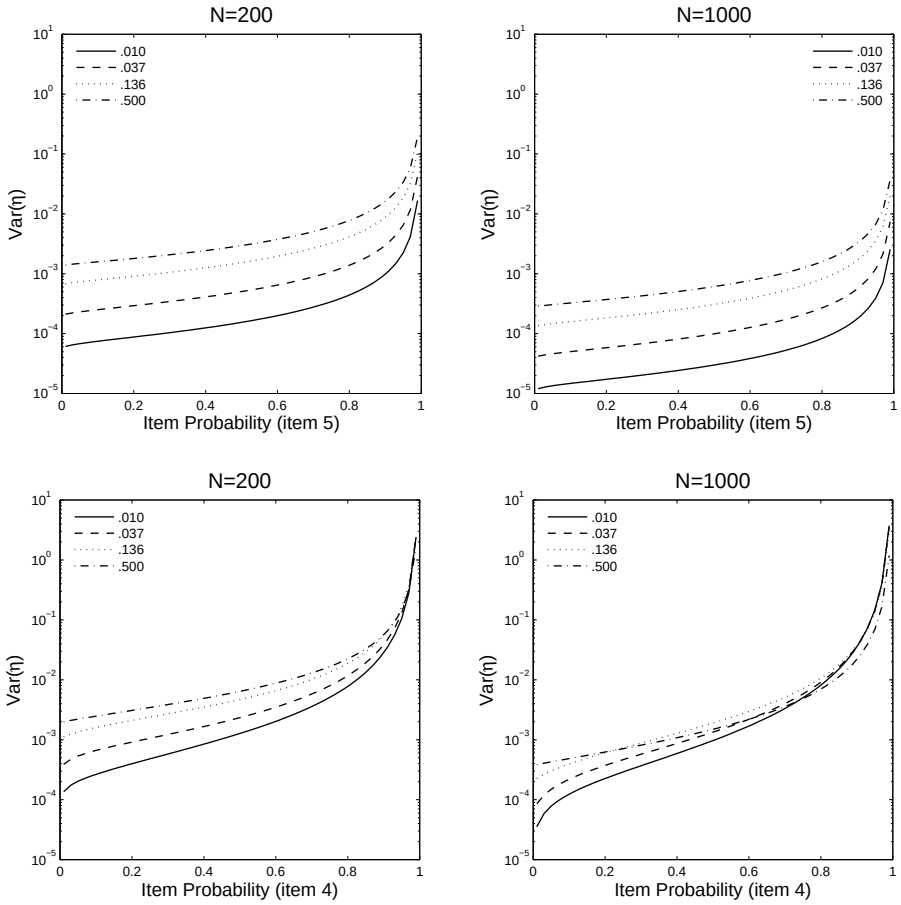

Figura 3.2: Varianza di $\hat{\eta}_q$ per l'item *"top"* (diagrammi in basso) e per l'item *"bottom"* (diagrammi in alto) in funzione della loro probabilità (x-axis) e del valore vero di η_q. L'asse y è in scala logaritmica.

Complessivamente, emerge che l'effetto dei parametri d'errore sulla loro varianza è relativamente piccolo se confrontato con quello della probabilità dell'item. Quest'ultima, fra le quattro variabili manipolate, sembra essere quella che ha un effetto maggiore sulla varianza dei parametri di lucky guess e careless error.

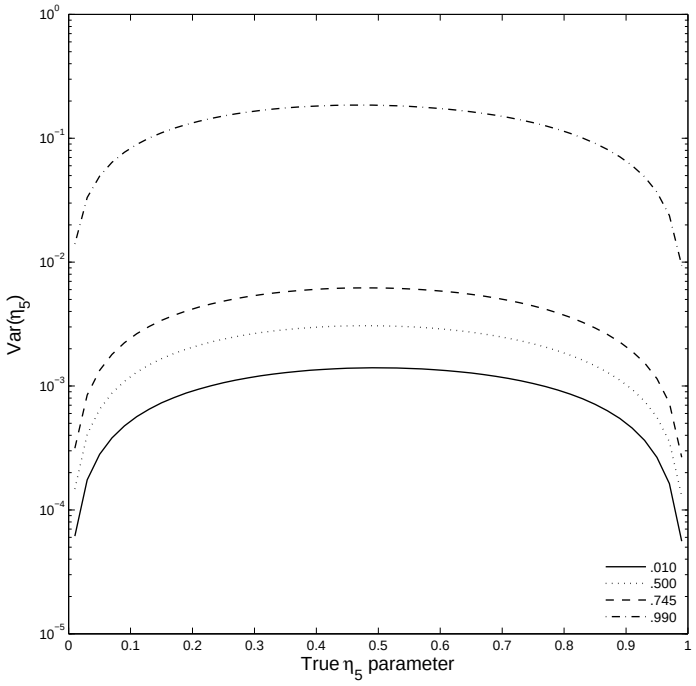

Figura 3.3: Relazione tra i parametri di lucky guess e la varianza delle stime MLE. Le curve corrispondono a differenti valori di probabilità dell'item. L'asse y è in scala logaritmica.

3.6 Applicazione Empirica

E' stato condotto uno studio empirico con l'obiettivo di studiare la varianza delle stime (ottenute per massima verosimiglianza) dei parametri del BLIM, nel caso di applicazioni reali. A questo scopo è stato somministrato un test carta-matita che valutava le abilità nell'ambito *problem solving aritmetico*, a un campione di 173 bambini italiani frequentanti la terza classe della scuola primaria.

Il test si componeva di 18 problemi tra cui 16 a risposta aperta e 4 a scelta multipla. I problemi a risposta aperta erano i classici problemi aritmetici di tipo testuale, che richiedevano l'applicazione di una, due o più delle quattro operazioni e alcune semplici attività di pianificazione della soluzione. Nei problemi a scelta multipla invece, il bambino, leggendo il testo di un problema, doveva semplicemente individuare quale operazione avrebbe utilizzato

per risolverlo.

Complessivamente si è ipotizzato che per risolvere l'insieme dei 18 problemi fossero necessarie 11 abilità, ovvero: manipolazione di tre addendi (NA), individuare la divisione come l'operazione per risolvere il problema (D), individuare l'addizione come l'operazione per risolvere il problema (A), individuare la sottrazione come l'operazione per risolvere il problema (S), individuare la moltiplicazione come l'operazione per risolvere il problema (M), individuare la sottrazione come operazione corretta, in un contesto ingannevole (SI), livello base di pianificazione (un operazione coinvolta – P1), livello intermedio di pianificazione (due operazioni coinvolte – P2), livello avanzato di pianificazione (tre operazioni coinvolte, – P3), risolvere problemi che richiedono almeno un calcolo in colonna (C). I 18 problemi sono stati associati alle 11 abilità come illustrato nella skill-multimap in Tabella 4.2. Per fare un esempio di problema a scelta multipla, si consideri il problema numero 2:

> *Clara ha 44 anni e ha 11 anni più di Angela. Per sapere quanti anni ha Angela, devo usare: (a) addizione; (b) sottrazione; (c) moltiplicazione; (d) divisione.*

In questo problema il bambino deve scegliere quale, tra le quattro operazioni, userebbe per individuare l'età di Angela. Va notato però che siamo in presenza di un contesto ingannevole, dal momento che l'operazione corretta è la sottrazione, ma nel testo del problema compare la parola *più*. Questa situazione comporta tipicamente una difficoltà aggiuntiva. Le abilità coinvolte nella soluzione di questo problema sono dunque due: (S) individuare la sottrazione come l'operazione per risolvere il problema e (SI) individuare la sottrazione come operazione corretta, in un contesto ingannevole. Per fare un esempio di problema a risposta aperta, si consideri il numero 10:

> *Un pittore ha conservato in tutto 122 barattoli di pittura rossa, 78 di gialla e 93 di blu. Successivamente decide di pitturare una tela e usa 86 barattoli. Quanti barattoli gli sono rimasti?*

Tabella 3.1: Skill-multimap usata per delineare la struttura di conoscenza dell'applicazione empirica. I problemi nelle colonne 1 e 3 richiedono le abilità nelle colonne 2 e 4.

Problemi	Insieme delle abilità	Problemi	Insieme delle abilità
1	{D}	10	{NA,A,S,P2,C}
2	{SI,S}	11	{S,C}
3	{M}	12	{S,M,P1,C}
4	{S}	13	{A,S,M,P1,P2,C}
5	{NA,A,S,P1,P2,C}	14	{D,C}
6	{D,M,P1,C} {D,A,P1}	15	{M,C}
7	{D,M,P1,P2,C}	16	{D,A,P1,P2,P3,C}
8	{S,M,P1,P2,P3,C}	17	{D,A,P1,C}
9	{SI,S,C}	18	{D,A,P1,P2,C}

NA = manipolazione di tre addendi, D = individuare la divisione come l'operazione per risolvere il problema, A = individuare l'addizione come l'operazione per risolvere il problema, S = individuare la sottrazione come l'operazione per risolvere il problema, M = individuare la moltiplicazione come l'operazione per risolvere il problema, SI = individuare la sottrazione come operazione corretta, in un contesto ingannevole, P1 = livello base di pianificazione (un operazione coinvolta), P2 = livello intermedio di pianificazione (due operazioni coinvolte), P3 = livello avanzato di pianificazione (tre operazioni coinvolte), C = risolvere problemi che richiedono almeno un calcolo in colonna.

In questo problema il bambino deve svolgere due operazioni per arrivare alla soluzione. Prima deve calcolare quanti barattoli ha in tutto il pittore, svolgendo un'addizione a tre addendi. Dopo di che potrà sottrarre al risultato il numero di barattoli che ha utilizzato per pitturare la tela. Le abilità necessarie per la risoluzione del problema, fra le 11 dell'intero test sono dunque: (A) individuare l'addizione come l'operazione per risolvere il problema; (NA) manipolazione di tre addendi; (S) individuare la sottrazione come l'operazione per risolvere il problema; (P2) livello intermedio di pianificazione (due operazioni coinvolte), (C) risolvere problemi che richiedono il calcolo in colonna.

Usando l'approccio *competence-performance* di Korossy (1999), descritto nel Capitolo 2, a partire dalla skill-multimap illustrata in Tabella 4.2, è stata delineata una struttura di conoscenza contenente 78 stati. Le risposte di ciascun bambino ai problemi sono state codificate come corrette (1) o errate (0) e sono state usate per costruire la matrice dei dati.

Il BLIM è stato quindi applicato ai dati e i suoi parametri sono stati stimati per massima verosimiglianza, utilizzando l'algoritmo EM (Stefanutti & Robusto, 2009), descritto nella Sezione 1.3.2. La bontà di adattamento del BLIM è stata testata con il Chi-quadro di Pearson il cui p-value, a causa della matrice sparsa dei dati, è stato ottenuto utilizzando una procedura di bootstrap parametrico su 500 replicazioni (Sezione 1.3.1). Infine, utilizzando la procedura descritta nella Sezione 3.3, è stata stimata la varianza delle stime ML dei parametri, per poi calcolare i relativi intervalli di confidenza (IC).

3.6.1 Risultati

Dall'applicazione del BLIM ai dati si è ottenuta una buona fit, con un p-value bootstrap pari a .78. E' stato quindi possibile calcolare la varianza delle stime ML dei parametri d'errore. La Figura 3.4 illustra gli intervalli di confidenza al 95% delle stime dei parametri di careless error e lucky guess ottenute per massima verosimiglianza. In Figura, sull'asse delle x vi sono i 18 item del test, mentre le stime dei parametri d'errore sono lungo l'asse y. Le barre grigie rappresentano gli IC delle stime dei parametri di lucky guess, mentre le barre bianche rappresentano gli IC delle stime dei parametri di careless error. Le stime puntuali dei parametri sono indicate con una linea orizzontale all'interno delle barre. La linea continua orizzontale che attraversa il grafico indica una probabilità d'errore del 50%.

Si elencano di seguito alcuni commenti relativi alla Figura 3.4. In primo luogo è possibile notare che la maggior parte dei parametri d'errore degli item,

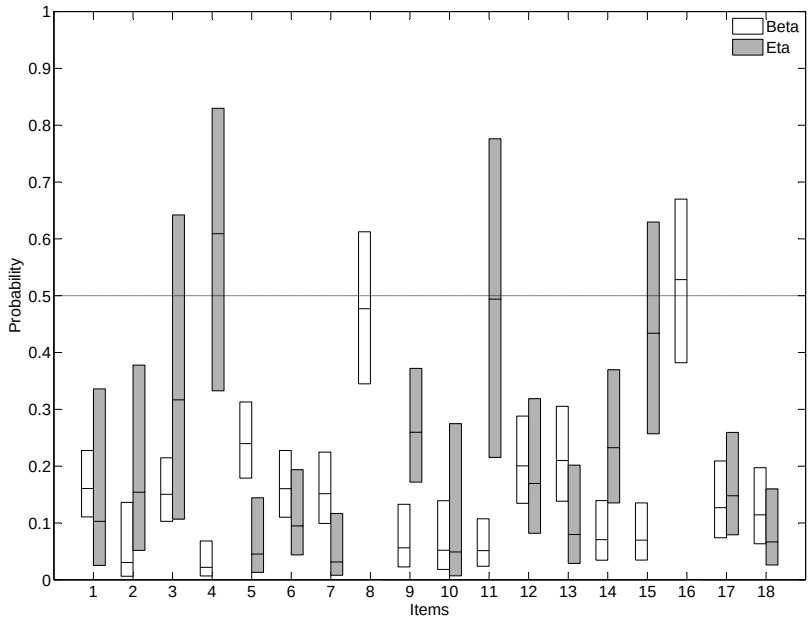

Figura 3.4: Intervalli di confidenza al 95% relativi alle stime dei parametri d'errore degli item, ottenuti nell'applicazione empirica. Sull'asse delle x vi sono i 18 item del test, mentre le stime dei parametri d'errore sono lungo l'asse y. Le barre grigie rappresentano gli IC delle stime dei parametri di lucky guess, mentre le barre bianche rappresentano gli IC delle stime dei parametri di careless error. Le stime puntuali dei parametri sono indicate con una linea orizzontale all'interno delle barre. La linea continua orizzontale che attraversa il grafico indica una probabilità d'errore del 50%.

hanno stime puntuali ragionevolmente basse (inferiori a .20) e intervalli di confidenza piuttosto piccoli. Ricordando che ci si aspettano valori piccoli dei parametri d'errore del BLIM, questo è un risultato positivo, che conferisce validità al modello.

Questa situazione però, non è vera per tutti i parametri del modello. Infatti alcuni di essi ottengono stime puntuali piuttosto elevate, con intervalli di confidenza piccoli (si vedano ad es. i parametri β_8 e β_{16}). Pertanto, per gli item 8 e 16 potrebbe esserci una specificazione errata del modello, dovuta probabilmente ad un'associazione tra questi due item e le abilità assunte alla

base della loro risoluzione, non corretta.

Per altri parametri si nota, invece, che valori elevati delle stime puntuali sono accompagnati da intervalli di confidenza piuttosto ampi ($\hat{\eta}_4$ e $\hat{\eta}_{11}$). In questi casi la stima puntuale è elevata ma c'è anche molta incertezza sull'esatta posizione del parametro vero. A causa di questa incertezza non si può concludere nulla su eventuali specificazioni errate del modello che coinvolgerebbero gli item 4 e 11. Diversi possono essere i motivi per cui le stime di questi parametri e delle relative varianze risultano elevati. Come evidenziato nella Sezione 4.4 la varianza del parametro di lucky guess di un item tende a crescere al crescere della della sua probabilità. Non sorprende che, fra i 18 item del test, quelli per i quali la probabilità $P(q)$ è maggiore siano proprio il 4 e l'11: $P(4) = P(11) = .87$.

Gli intervalli di confidenza dei parametri $\hat{\eta}_8$ e $\hat{\eta}_{16}$ non sono visibili nel grafico di Figura 3.4. Questo è dovuto al fatto che le stime puntuali di questi due item sono prossime allo zero ($\hat{\eta}_8 = 3.13 \times 10^{-10}$, $\hat{\eta}_{16} = 2.87 \times 10^{-11}$) e i loro intervalli di confidenza sono del tutto trascurabili.

Per quanto riguarda la relazione tra la probabilità degli item e la loro varianza, la Figura 3.5 mostra i risultati ottenuti dall'applicazione empirica. In figura le stime delle probabilità degli item si collocano lungo l'asse x e l'errore standard delle stime dei parametri β_q (punti neri) e η_q (punti bianchi) si colloca lungo l'asse y. Dalla figura si osserva che:

- l'errore standard delle stime dei parametri di lucky guess *crescono* al crescere della probabilità dell'item;
- l'errore standard dei parametri di careless error *decrescono* al crescere della probabilità dell'item.

Questi risultati sono del tutto in linea con quelli ottenuti nello studio simulativo. Ciò che si osserva di diverso rispetto alle simulazioni è che l'errore standard dei parametri $\hat{\beta}_q$ è, mediamente, più basso di quello dei parametri $\hat{\eta}_q$. Questo risultato è dovuto al fatto che le probabilità degli item sono tutte

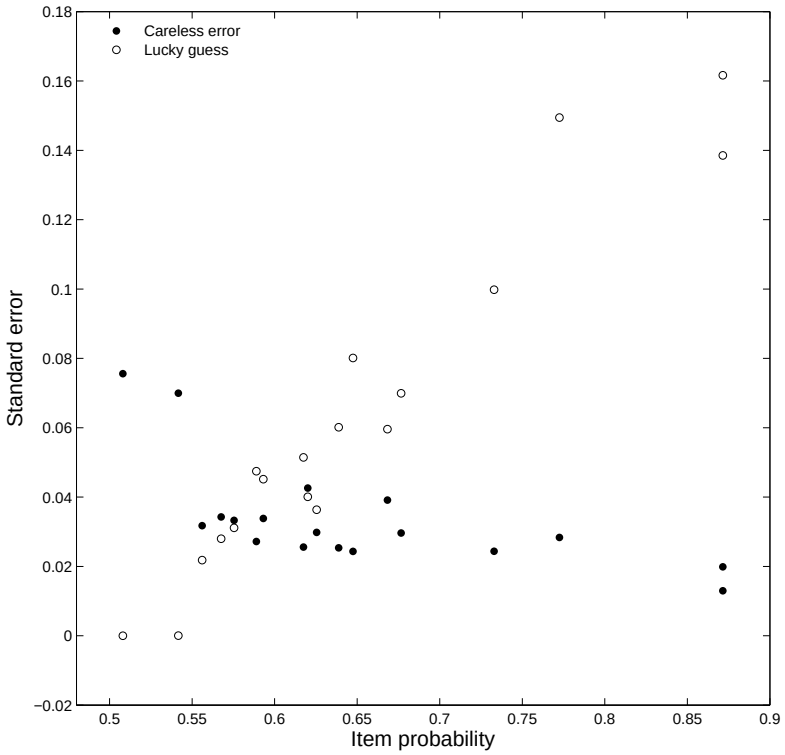

Figura 3.5: Errore standard delle stime dei parametri d'errore vs. le probabilità dei 18 item utilizzati nell'applicazione empirica del BLIM. Le stime delle probabilità degli item si collocano lungo l'asse x e l'errore standard delle stime dei parametri β_q (punti neri) mentre η_q (punti bianchi) si colloca lungo l'asse y.

comprese tra .50 e 1. Per questa ragione la varianza delle stime dei parametri di careless error non è mai troppo elevata.

3.7 Discussione

Le formule analitiche per il calcolo della matrice di covarianza del BLIM non erano mai state derivate nella letteratura scientifica su questo modello, sebbene abbiano un ruolo molto importante non solo per lo studio del comportamento della varianza dei parametri del modello, ma anche per il calcolo degli intervalli di confidenza delle stime dei parametri.

L'obiettivo della ricerca descritta in questo capitolo, è stato quindi quello di derivare le matrici dell'informazione di Fischer (FIM) e di covarianza del BLIM e di utilizzarle per studiare il comportamento asintotico della varianza delle stime ML dei parametri d'errore del modello. Conoscere queste caratteristiche di un modello è molto importante anche nella fase della sua applicazione ai dati, e in particolare nell'interpretazione dei parametri. Da queste matrici infatti si possono costruire gli intervalli di confidenza.

Nella Sezione 3.3 si sono dunque derivate le due matrici per il BLIM. I risultati teorici sono stati poi impiegati in uno studio simulativo (Sezione 3.5), con l'obiettivo di analizzare il comportamento della varianza delle stime dei parametri del BLIM sotto diverse condizioni. Un'attenzione particolare è stata data ai parametri di careless error e di lucky guess di due tipi di item: uno primo tipo appartenente a molti stati di conoscenza (chiamato "bottom" item) e l'altro tipo appartenente a pochi stati di conoscenza (chiamato "top" item). I risultati dello studio hanno evidenziato che tra la numerosità campionaria, il tipo di item ("bottom" o "top"), i valori veri dei parametri e la probabilità dell'item, è quest'ultima ad avere un effetto maggiore sulla varianza delle stime dei parametri. In particolare si osserva che al crescere della probabilità di un item cresce la varianza della stima dei parametro di lucky guess, mentre al diminuire delle probabilità di un item, cresce la varianza delle stime dei parametri di careless error. In altre parole, ci si aspetta un'incertezza elevata della probabilità:

- di lucky guess, nel caso di item *molto facili*;
- di careless error nel caso di item *molto difficili*.

Le procedure per il calcolo della varianza analitica del BLIM sono state utilizzate anche in un'applicazione empirica, per ottenere gli intervalli di confidenza delle stime dei parametri del modello. I risultati dell'applicazione suggeriscono che gli intervalli di confidenza hanno un valore diagnostico nell'individuazione di specificazioni errate del modello, ma non solo. La stima

puntuale particolarmente elevata del parametro di un item associata a un intervallo di confidenza piccolo, è indice di una possibile specificazione errata del modello per quel particolare item. Ma è anche vero che una stima elevata associata una varianza ampia del parametro può anche indicare altri tipi di considerazioni. Parallelamente a quanto osservato per lo studio simulato infatti, le stime dei parametri degli item la cui probabilità è elevata hanno una varianza elevata, indicando incertezza sulla posizione del parametro vero. In questi casi sarebbe del tutto fuorviante pensare a specificazioni errate del modello.

Oltre a considerazioni di carattere interpretativo, i risultati della ricerca presentata, possono essere impiegati anche per altri scopi. L'informazione di Fischer e la matrice Hessiana hanno infatti un ruolo importante nelle applicazioni per quanto riguarda:

- l'efficienza delle procedure di stima per massima verosimiglianza, come per esempio l'algoritmo Newton-Raphson (Press, Teukolsky, Vetterling, & Flannery, 2007);
- l'identificabilità locale di un modello (Goodman, 1974).

Le ricerche future sul BLIM, dovrebbero dare un'attenzione particolare a questi due aspetti.

Capitolo 4

Testare l'Invarianza dei Parametri del BLIM: Modelli a Bipartizione

In questo capitolo si presentano alcune generalizzazioni dei risultati ottenuti da de Chiusole, Stefanutti, Anselmi, e Robusto (2013). Il lavoro nasce dalla volontà di studiare in maniera approfondita alcune proprietà del BLIM, il modello più utilizzato nella KST per la validazione empirica delle strutture di conoscenza.

Nonostante alcune problematiche legate all'applicabilità del modello ai contesti reali, siano state ad oggi risolte, altre rimangono ancora inesplorate. Mancano infatti procedure per testare l'assunzione di invarianza dei parametri d'errore del BLIM. In tutte le sezioni del presente capitolo ci si riferirà a questa assunzione come *assunzione di invarianza*. Secondo tale assunzione la probabilità di distrazione e la probabilità di indovinare la risposta di un item, non dipenderebbero dallo stato di conoscenza dello studente, ma sarebbero una proprietà dell'item.

Ci si è chiesti allora se questa assunzione non fosse troppo forte e se potesse invece, in certe circostanze, essere violata dai dati. E' plausibile ad

esempio ipotizzare che la probabilità di distrazione possa variare non solo attraverso gli item ma anche attraverso gli individui.

In questo capitolo si presenta una procedura per testare le violazioni di questa assunzione.

4.1 Introduzione

In generale, quando si applica un modello ai dati non si conosce la sensibilità dei test statistici (come per esempio il Chi-quadro di Pearson o il rapporto di verosimiglianza) alle violazioni delle sue assunzioni. Per esserne certi, oltre che testare la sua fit occorre testare, attraverso le procedure opportune, anche le assunzioni su cui si basa.

Nel caso specifico del BLIM, diverse possono essere le ragioni di uno scarso adattamento ai dati. Potrebbe dipendere da una violazione dell'assunzione di invarianza, ma anche da eventuali errori di specificazione del modello, ovvero da una errata definizione degli stati di conoscenza. Dall'altro lato, quando il BLIM ha una buona fit, non si può essere certi che tutte le assunzioni su cui si basa siano rispettate dai dati. Quest'ultima è sicuramente la situazione più allarmante perché porterebbe a commettere degli errori di valutazione piuttosto grossolani. La ricerca qui presentata ha l'obiettivo di mettere a disposizione una procedura utile all'individuazione delle violazioni dell'assunzione di invarianza del BLIM, per evitare di commettere errori circa la bontà di adattamento del modello ai dati.

La procedura, prende ispirazione dall'approccio utilizzato nell'item response theory (IRT) per testare l'assunzione di indipendenza delle difficoltà degli item dalle abilità degli individui, tipica dei modelli di Rasch. L'approccio utilizzato nell'IRT (Andersen, 1973; Glas & Verhelst, 1995) si scompone nelle seguenti fasi: (1) suddivisione dei dati osservati in due o più gruppi. Ad esempio si possono formare due gruppi: il primo contenente soggetti con punteggio minore o uguale alla mediana e il secondo contenente soggetti con

punteggi maggiori alla mediana; (2) stima dei parametri del modello in ciascun gruppo di dati; (3) applicazione di test statistici della differenza tra le stime dei parametri ottenute nei diversi gruppi. Se il test è statisticamente significativo si conclude che l'assunzione di indipendenza delle difficoltà degli item dalle abilità dei soggetti è violata dai dati.

Il problema della violazione dell'assunzione di invarianza emerge anche nell'area di ricerca dei *cognitive diagnostic models* (Bolt, 2007; de la Torre & Douglas, 2004; de la Torre, 2009; DiBello & Stout, 2007; C. Tatsuoka, 2002; K. Tatsuoka, 1990), in particolare quando si applica il modello *deterministic inputs, noisy AND-gate* (DINA, Junker & Sijtsma, 2001). In questo modello si assume che le probabilità careless error e lucky guess degli item siano indipendenti dalle abilità possedute da un individuo. In un articolo recente, de la Torre e Lee (2010) propongono un metodo per testare l'assunzione di invarianza dei parametri del DINA. Nelle loro analisi, anziché creare due gruppi "puri" di soggetti con punteggi minori o maggiori della mediana, hanno costruito due differenti data set ottenuti da una combinazione dei due gruppi "puri": un primo gruppo era composto dal 60% dei soggetti che avevano un punteggio sotto la mediana e il 40% dei soggetti con un punteggio sopra la mediana; il secondo gruppo era composto dal 60% dei soggetti che avevano un punteggio sopra la mediana e un 40% sotto la mediana. A ciascun gruppo hanno poi applicato il DINA e testato la differenza dei parametri del modello stimati nei due gruppi. I risultati hanno evidenziato che l'assunzione di invarianza dei parametri del DINA può essere violata dai dati, il che significa che le stime delle difficoltà degli item possono variare con il punteggio degli individui.

Sia nel caso dell'IRT, con i modelli di Rasch, sia in quello dei CDM, con il DINA, è possibile trovare delle analogie con il BLIM. Nel primo caso i parametri d'errore degli item nel BLIM ricordano le difficoltà degli item dei modelli di Rasch, mentre le probabilità degli stati di conoscenza nel BLIM ricordano le abilità dei soggetti dei modelli di Rasch. Nel secondo caso,

è stato recentemente evidenziato da Heller, Stefanutti, Anselmi, e Robusto (under revision) che, a livello performance, il DINA è equivalente al BLIM. Sembrava così ragionevole trasferire le procedure per testare l'assunzione di invarianza, sviluppate in questi ambiti, a quello della KST. Ma come spesso accade, la scelta più semplice è quella sbagliata. Infatti, un adattamento delle procedure sopra descritte, non funziona correttamente se applicato alla KST. Nella Sezione 4.2 si illustra l'adattamento di questo metodo alla KST, evidenziandone le problematicità.

Si è pensato allora a altro modo per testare questa assunzione, forse più in linea con l'approccio modellistico. Nella Sezione 4.3 si presenta un metodo che consiste nel confrontare il BLIM con un modello alternativo, chiamato *Bipartition Model* (BPM), nel quale la dipendenza dei parametri degli item dagli stati di conoscenza è un'assunzione esplicita del modello. Per testare la capacità del BPM nell'individuare le dipendenze tra i parametri degli item e gli stati di conoscenza, è stato condotto uno studio simulativo (Sezione 4.4). Infine il BLIM e il BPM sono stati confrontati in un'applicazione empirica (Sezione 4.5) per controllare se, in contesti reali, l'assunzione di invarianza possa essere violata dai dati anche quando il BLIM ottiene una buona fit.

4.2 Test Naïve dell'Invarianza del BLIM

Come anticipato nell'introduzione, un modo per valutare le violazioni dell'assunzione di invarianza dei parametri del BLIM, è quello di dividere l'intero data set in due gruppi indipendenti, che chiamiamo Gruppo 1 e Gruppo 2, applicare il BLIM a ciascuno di essi, e utilizzare un test statistico adatto alla valutazione della differenza tra le stime dei parametri dei due gruppi. Se il test è statisticamente significativo allora si conclude che l'assunzione di invarianza è violata dai dati.

Si consideri un qualunque cutoff $c \in \{0, 1, \ldots, |Q|-1\}$, inoltre $\mathcal{R}^{\downarrow} = \{R \in \mathcal{R} : |R| \leq c\}$ sia la collezione di tutti i pattern di risposta la cui cardinalità

è minore o uguale a c, e $\mathcal{R}^{\uparrow} = \{R \in \mathcal{R} : |R| > c\}$ sia la collezione di tutti i pattern la cui cardinalità è maggiore di c. Allora, secondo il BLIM, le probabilità condizionali che, nell'estrazione casuale di un pattern di risposta, un item q venga sbagliato per careless error, dato che R è sotto o sopra il cutoff, sono determinate rispettivamente dalle due equazioni che seguono:

$$\beta_q^{\downarrow} = \frac{\sum\limits_{R \in \bar{\mathcal{R}}_q \cap \mathcal{R}^{\downarrow}} \sum\limits_{K \in \mathcal{K}_q} P(R,K)\pi_K}{\sum\limits_{R \in \mathcal{R}^{\downarrow}} \sum\limits_{K \in \mathcal{K}_q} P(R,K)\pi_K}, \quad \beta_q^{\uparrow} = \frac{\sum\limits_{R \in \bar{\mathcal{R}}_q \cap \mathcal{R}^{\uparrow}} \sum\limits_{K \in \mathcal{K}_q} P(R,K)\pi_K}{\sum\limits_{R \in \mathcal{R}^{\uparrow}} \sum\limits_{K \in \mathcal{K}_q} P(R,K)\pi_K} \tag{4.1}$$

Parallelamente, le probabilità condizionali che, nell'estrazione casuale di un pattern di risposta R, l'item q sia corretto per lucky guess, dato che R è sotto o sopra il cutoff, sono determinate, rispettivamente, dalle due equazioni che seguono:

$$\eta_q^{\downarrow} = \frac{\sum\limits_{R \in \mathcal{R}_q \cap \mathcal{R}^{\downarrow}} \sum\limits_{K \in \bar{\mathcal{K}}_q} P(R,K)\pi_K}{\sum\limits_{R \in \mathcal{R}^{\downarrow}} \sum\limits_{K \in \bar{\mathcal{K}}_q} P(R,K)\pi_K}, \quad \eta_q^{\uparrow} = \frac{\sum\limits_{R \in \mathcal{R}_q \cap \mathcal{R}^{\uparrow}} \sum\limits_{K \in \bar{\mathcal{K}}_q} P(R,K)\pi_K}{\sum\limits_{R \in \mathcal{R}^{\uparrow}} \sum\limits_{K \in \bar{\mathcal{K}}_q} P(R,K)\pi_K}. \tag{4.2}$$

Per qualsiasi scelta del cutoff c, per qualsiasi item $q \in Q$ e per ogni $\beta_q, \eta_q \in (0,1)$, si ottengono le seguenti disequazioni:

$$\beta_q^{\uparrow} < \beta_q < \beta_q^{\downarrow}, \quad \eta_q^{\downarrow} < \eta_q < \eta_q^{\uparrow}. \tag{4.3}$$

Si rimanda il lettore all'articolo di de Chiusole et al. (2013) per le dimostrazioni. Le considerazioni che emergono dalle disequazioni in 4.3, hanno chiare conseguenze sull'inappropriatezza di questo metodo. Si evince infatti che, anche quando vale l'assunzione di invarianza, è più probabile osservare una careless error quando si campiona sotto il cutoff (condizione di sinistra), mentre è più probabile osservare una lucky guess quando si campiona sopra il cutoff (condizione di destra).

In sintesi, stimare i parametri del BLIM su una parte del campione, sia essa sotto o sopra il cutoff, porta a stime affette da bias. Questo può essere spiegato anche intuitivamente se si considera che i parametri $\beta_q^{\downarrow}, \beta_q^{\uparrow}, \eta_q^{\downarrow} e \eta_q^{\uparrow}$

sarebbero stimati assumendo una distribuzione troncata sui pattern di risposta: se, ad esempio, si campiona sopra il cutoff c, ogni pattern di risposta sotto a c avrebbe una probabilità pari a zero di essere osservato, e viceversa nel caso in cui si campionasse sotto a c. Nel BLIM invece, i parametri β_q ed η_q vengono stimati sotto l'assunzione che ogni pattern di risposta abbia probabilità di essere osservato diversa da zero.

Dalle considerazioni fin qui emerse, si potrebbe ipotizzare che sia la regola utilizzata per formare i due gruppi sopra e sotto il cutoff a determinare il bias nelle stime dei parametri. Ci si è chiesti allora se potesse esserci un altro modo per campionare i soggetti, che consentisse di evitare il bias.

In linea con il metodo utilizzato da de la Torre e Lee (2010), si è deciso di considerare la regola seguente: in un'estrazione casuale con reinserimento, data la proporzione $p^{\downarrow}$, con $0 \leq p^{\downarrow} \leq 1$, si estrae un numero $n^{\downarrow}$ sufficientemente grande di pattern di riposta da $\mathcal{R}^{\downarrow}$, e si assegna al Gruppo $\mathcal{G}_1$ con probabilità $p^{\downarrow}$, e al Gruppo $\mathcal{G}_2$ con probabilità $1 - p^{\downarrow}$. In maniera del tutto analoga, data la proporzione $p^{\uparrow}$, $0 \leq p^{\uparrow} \leq 1$, $n^{\uparrow}$ pattern di risposta vengono campionati casualmente con reinserimento da $\mathcal{R}^{\uparrow}$, e ognuno di essi viene assegnato al gruppo $\mathcal{G}_1$ con probabilità $p^{\uparrow}$, e a $\mathcal{G}_2$ con probabilità $1 - p^{\uparrow}$.

Come stabilito dalla proposizione che segue, anche questa regola più generale soffre degli stessi problemi illustrati nel paragrafo precedente. Si faccia riferimento a de Chiusole, Stefanutti, Anselmi, e Robusto (in press) per la dimostrazione della Proposizione 2 e tutte quelle che seguiranno.

Proposizione 2. *Sia $\beta_q^{(1)}$ la probabilità che, in un pattern di risposta estratto casualmente, un item q venga sbagliato per distrazione, dato che il pattern appartiene al Gruppo $\mathcal{G}_1$, e $\eta_q^{(1)}$ sia la probabilità che l'item q venga svolto correttamente per lucky guess, dato che il pattern appartiene a $\mathcal{G}_1$. Allora $\beta_q^{(1)} \leq \beta_q$ se e solo se*

$$\frac{p^{\downarrow}}{p^{\uparrow}} \geq \frac{\beta_q P(\mathcal{R}^{\downarrow}, \mathcal{K}_q) - P(\bar{\mathcal{R}}_q^{\downarrow}, \mathcal{K}_q)}{P(\bar{\mathcal{R}}_q^{\uparrow}, \mathcal{K}_q) - \beta_q P(\mathcal{R}^{\uparrow}, \mathcal{K}_q)},$$

dove $\bar{\mathcal{R}}_q^{\downarrow} = \bar{\mathcal{R}}_q \cap \mathcal{R}^{\downarrow}$ *e* $\bar{\mathcal{R}}_q^{\uparrow} = \bar{\mathcal{R}}_q \cap \mathcal{R}^{\uparrow}$. *Inoltre,* $\eta_q^{(1)} \leq \eta_q$ *se e solo se*

$$\frac{p^{\downarrow}}{p^{\uparrow}} \leq \frac{\eta_q P(\mathcal{R}^{\uparrow}, \bar{\mathcal{K}}_q) - P(\mathcal{R}_q^{\downarrow}, \bar{\mathcal{K}}_q)}{P(\mathcal{R}_q^{\downarrow}, \bar{\mathcal{K}}_q) - \eta_q P(\mathcal{R}^{\downarrow}, \bar{\mathcal{K}}_q)},$$

dove $\mathcal{R}_q^{\downarrow} = \mathcal{R}_q \cap \mathcal{R}^{\downarrow}$ *e* $\mathcal{R}_q^{\uparrow} = \mathcal{R}_q \cap \mathcal{R}^{\uparrow}$.

Ciò che segue dalla Proposizione 2 è che per qualunque scelta dei parametri del BLIM β_q, η_q e π_K, esiste un'unica scelta di $p^{\downarrow}/p^{\uparrow}$ per la quale l'equazione $\beta_q^{(1)} = \beta_q$ risulti vera, e un'unica scelta di $p^{\uparrow}/p^{\downarrow}$, per la quale l'equazione $\eta_q^{(1)} = \eta_q$ sia vera. Va sottolineato che, in generale, i due rapporti differiranno da questa scelta, e conseguentemente i parametri β_q e η_q saranno quasi sempre affetti da bias.

4.3 Modelli a Bipartizione

Un modo più diretto per testare l'assunzione di invarianza del BLIM è di confrontare questo modello con un modello alternativo in cui la dipendenza dei parametri d'errore dagli stati di conoscenza diventa un'assunzione esplicita. Se il modello alternativo si adatta ai dati meglio del BLIM, l'assunzione di invarianza è violata.

In una formulazione generale di questo *modello di dipendenza*, le probabilità di careless error e di lucky guess sono libere di variare attraverso gli stati di conoscenza. Questa particolare condizione, può essere espressa attraverso la seguente funzione di risposta dei pattern dati gli stati di conoscenza:

$$P(R,K) = \left[\prod_{q \in K \setminus R} \beta_{qK}\right] \left[\prod_{q \in K \cap R} (1 - \beta_{qK})\right] \left[\prod_{q \in R \setminus K} \eta_{qK}\right] \left[\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_{qK})\right], \quad (4.4)$$

dove le careless error $\beta_{qK} \in (0,1)$ e le lucky guess $\eta_{qK} \in (0,1)$ variano sia attraverso gli item sia attraverso gli stati di conoscenza.

Con il solo obiettivo di individuare eventuali dipendenze dei parametri d'errore dagli stati di conoscenza, un modello di questo tipo potrebbe essere troppo complesso. Si possono ottenere modelli più semplici, ponendo specifiche restrizioni al modello generale definito nell'equazione 4.4. Per ogni item $q \in Q$, queste restrizioni sono rappresentate da vincoli di uguaglianza posti sui parametri d'errore di q attraverso gli stati di conoscenza $K_1, K_2, \ldots, K_n$, appartenenti a un qualunque sottoinsieme di $\mathcal{K}$. Tali vincoli assumono la forma generale

$$\beta_{qK_1} = \beta_{qK_2} = \cdots = \beta_{qK_n}$$

$$\eta_{qK_1} = \eta_{qK_2} = \cdots = \eta_{qK_n}$$

E' possibile definire una gerarchia tra tutti questi modelli, dove ciascuno di essi si specifica come segue: per ogni item $q \in Q$, siano $\overset{\beta}{\sim}_q$ e $\overset{\eta}{\sim}_q$ due relazioni di equivalenza sugli stati di conoscenza appartenenti a $\mathcal{K}$. Allora, per ogni $K, L \in \mathcal{K}$ si richiede che:

$$K \overset{\beta}{\sim}_q L \implies \beta_{qK} = \beta_{qL},$$

e

$$K \overset{\eta}{\sim}_q L \implies \eta_{qK} = \eta_{qL}.$$

Così facendo, ciascun modello della gerarchia corrisponde a una differente collezione $\{\langle \overset{\beta}{\sim}_q, \overset{\eta}{\sim}_q \rangle : q \in Q\}$ di coppie delle relazioni di equivalenza per $\mathcal{K}$. Si noti come sia il BLIM che il modello di dipendenza generale (in seguito chiamato GDM, dall'inglese general dependence model), appartengono alla gerarchia.

Con l'obiettivo di testare l'assunzione di invarianza, si considera qui l'elemento più semplice della gerarchia, dove i vincoli di uguaglianza sono rappresentati da una bipartizione della struttura di conoscenza $\mathcal{K}$. Questi modelli, chiamati di seguito *modelli a bipartizione* (BPM), sono l'elemento della gerarchia successivo al BLIM e dovrebbero essere sufficienti per individuare le dipendenze, qualora esistano.

Per suddividere la struttura di conoscenza $\mathcal{K}$ in due classi di equivalenza, occorre scegliere un certo cutoff $c \in \{0, 1, \ldots, |Q| - 1\}$ sulla cardinalità degli stati di conoscenza. Qualsiasi scelta di c induce una partizione di $\mathcal{K}$ nelle due classi di equivalenza

$$\mathcal{K}^- = \{K \in \mathcal{K} : |K| \leq c\},$$
$$\mathcal{K}^+ = \{K \in \mathcal{K} : |K| > c\}.$$

In ciascuna delle due classi $\mathcal{K}^-$ e $\mathcal{K}^+$, per ogni $q \in Q$ e ogni $K \in \mathcal{K}$, si avrà una coppia di parametri careless error e lucky guess, i cui valori sono dati dalle seguenti equazioni:

$$\beta_{qK} = \begin{cases} \beta_q^+ & \text{se } K \in \mathcal{K}^+ \\ \beta_q^- & \text{se } K \in \mathcal{K}^- \end{cases} \quad \text{e} \quad \eta_{qK} = \begin{cases} \eta_q^+ & \text{se } K \in \mathcal{K}^+ \\ \eta_q^- & \text{se } K \in \mathcal{K}^-, \end{cases}$$

dove β_q^- e η_q^- sono, rispettivamente, le probabilità di careless error e di lucky guess dell'item q per gli stati di conoscenza *sotto* il cutoff, e β_q^+ e η_q^+ sono, rispettivamente, le probabilità di careless error e di lucky guess dell'item q per gli stati di conoscenza *sopra* il cutoff.

E' importante evidenziare come i parametri β_q^-, η_q^-, β_q^+ e η_q^+ abbiano particolari condizioni di esistenza, infatti non sono sempre definiti. In un modello a bipartizione le probabilità d'errore β_q^- e β_q^+ soddisfano le seguenti equazioni:

$$\beta_q^- = \frac{\sum_{R \in \bar{\mathcal{R}}_q} \sum_{K \in \mathcal{K}_q^-} P(R, K)\pi_K}{\sum_{K \in \mathcal{K}_q^-} \pi_K}, \quad \beta_q^+ = \frac{\sum_{R \in \bar{\mathcal{R}}_q} \sum_{K \in \mathcal{K}_q^+} P(R, K)\pi_K}{\sum_{K \in \mathcal{K}_q^+} \pi_K}, \tag{4.5}$$

dove $\mathcal{K}_q^- = \{K \in \mathcal{K}^- : q \in K\}$, e $\mathcal{K}_q^+ = \{K \in \mathcal{K}^+ : q \in K\}$. A condizione che le probabilità degli stati di conoscenza $K \in \mathcal{K}$ siano tutte maggiori a zero, dall'equazione di sinistra in (4.5), si nota che la probabilità di careless error β_q^- è definita solamente nel caso in cui la collezione $\mathcal{K}_q^-$ non è vuota; in caso contrario, sia il numeratore che il denominatore dell'equazione andrebbero

a zero, lasciando β_q^- indefinito. Questo dipende dal cutoff c scelto per la partizione. Per fare un esempio, si consideri il cutoff $c = 1$ e la struttura di conoscenza

$$\mathcal{K}_1 = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,3\}, \{1,2,3\}\}.$$

Tra gli stati contenuti in $\mathcal{K}_1$, nessuno con cardinalità minore o uguale a 1 contiene l'item 3. Il parametro β_3^- rimane così indefinito per $\mathcal{K}_1$, quando il cutoff è 1. Invece, nel caso in cui il cutoff sia $c = 2$, esiste almeno uno stato di conoscenza sotto c che contiene l'item 3, ad indicare che il parametro β_3^- è definito.

Si consideri ora l'equazione di destra in (4.5). La probabilità di careless error β_q^+ è definita solamente se $\mathcal{K}_q^+$ non è vuota. Questa condizione, al contrario della precedente è sempre vera perché, indipendentemente dal cutoff scelto, l'insieme totale Q apparterà sempre a $\mathcal{K}^+$.

Un ragionamento simmetrico si applica ai parametri di lucky guess. Le equazioni che seguono soddisfano le probabilità di lucky guess η_q^- e η_q^+:

$$\eta_q^- = \frac{\sum_{R \in \mathcal{R}_q} \sum_{K \in \bar{\mathcal{K}}_q^-} P(R,K)\pi_K}{\sum_{K \in \bar{\mathcal{K}}_q^-} \pi_K}, \quad \eta_q^+ = \frac{\sum_{R \in \mathcal{R}_q} \sum_{K \in \bar{\mathcal{K}}_q^+} P(R,K)\pi_K}{\sum_{K \in \bar{\mathcal{K}}_q^+} \pi_K}, \tag{4.6}$$

dove $\bar{\mathcal{K}}_q^- = \{K \in \mathcal{K}^- : q \notin K\}$, e $\bar{\mathcal{K}}_q^+ = \{K \in \mathcal{K}^+ : q \notin K\}$. Dall'equazione di sinistra in (4.6) si nota che il parametro di lucky guess η_q^- è definito solo se $\bar{\mathcal{K}}_q^-$ non è vuoto, e questa condizione è sempre vera perché, indipendentemente dal cutoff scelto, almeno l'insieme vuoto apparterà sempre a $\mathcal{K}^-$. Nell'equazione di destra in (4.6), la probabilità di lucky guess η_q^+ è definita solo se $\bar{\mathcal{K}}_q^+$ non è vuoto e, in determinate circostanze, questa condizione potrebbe essere falsa. Considerando l'esempio precedente sulla struttura di conoscenza $\mathcal{K}_1$. Con il cutoff $c = 1$ la probabilità di lucky guess dell'item 1 sopra il cutoff (η_1^+) è indefinita, perché non ci sono stati di conoscenza in $\mathcal{K}^+$ che non contengono q, ovvero $\bar{\mathcal{K}}_q^+ = \emptyset$.

Per riassumere, nei modelli a bipartizione i parametri β_q^+ e η_q^- sono sempre definiti, indipendentemente dal cutoff scelto. Dall'altro lato, gli altri due parametri sono definiti sotto le seguenti condizioni:

- β_q^- è definito solo se $\mathcal{K}_q^- \neq \emptyset$
- η_q^+ è definito solo se $\bar{\mathcal{K}}_q^+ \neq \emptyset$.

Il punto cruciale ora è capire cosa succede quando alcuni parametri di un modello a bipartizione sono indefiniti. Quello che accade è che i parametri *sotto il cutoff* non si possono separare da quelli *sopra il cutoff*, e conseguentemente, per gli item in questione esisterà un solo parametro. Questo significa che, in generale, data una certa struttura di conoscenza, i modelli a bipartizione che si ottengono variando il valore del cutoff c, possono differire sostanzialmente l'uno dall'altro, in termini di numero dei parametri d'errore.

Per chiarire questo punto, si consideri la seguente struttura di conoscenza su 5 item

$$\mathcal{K}_2 = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}, \{1,2,4\}, \{1,2,3,4\}, \{1,2,3,4,5\}\}. \quad (4.7)$$

Con cinque item, il cutoff può essere scelto tra i valori $\{0,1,2,3,4\}$. La Tabella 4.1 mostra come, per ogni scelta possibile di c, alcuni parametri sono definiti (indicati in tabella con il simbolo di spunta), mentre altri non lo sono. Per esempio con $c = 0$, i parametri di careless error sotto il cutoff sono tutti indefiniti, mentre le lucky guess sopra il cutoff sono tutte definite. Dalla Tabella 4.1 è possibile inoltre notare che, per la struttura considerata, non esiste un particolare valore del cutoff per il quale tutti i parametri β_q^- e η_q^+ siano definiti.

Tabella 4.1: Parametri d'errore definiti e indefiniti nei modelli a bipartizione per la struttura di conoscenza $\mathcal{K}_2$, in funzione del cutoff. Per ogni valore del cutoff, i parametri definiti sono indicati con il simbolo di spunta ✓.

Cutoff	β_1^-	β_2^-	β_3^-	β_4^-	β_5^-	η_1^+	η_2^+	η_3^+	η_4^+	η_5^+
0						✓	✓	✓	✓	✓
1	✓	✓				✓	✓	✓	✓	✓
2	✓	✓	✓					✓	✓	✓
3	✓	✓	✓	✓						✓
4	✓	✓	✓	✓						

4.4 Studio Simulativo

Attraverso una serie di simulazioni, si è voluto testare la capacità del BPM di individuare le eventuali dipendenze dei parametri d'errore β_q ed η_q dagli stati di conoscenza. A questo scopo, è stato condotto uno studio simulativo nel quale il BPM e il BLIM sono stati confrontati attraverso alcuni indici di selezione dei modelli, tra cui il rapporto di verosimiglianza (Likelihood Ratio - LR), l'Akaike Information Criterion (AIC; Akaike, 1973), l'AIC con la correzione per campioni di piccole dimensioni ($AICc$; Hurvich & Tsai, 1989) e il Bayesian Information Criterion (BIC; Schwarz, 1978).

Nel confronto, sono state considerate differenti condizioni, in cui il modello che generava i dati (modello generativo) poteva essere il BPM o il BLIM. Inoltre, quando il modello generativo era il BPM, i dati sono stati simulati utilizzando differenti valori di β_q e η_q, che potevano essere molto piccoli o molti grandi. L'obiettivo era testare la fit dei due modelli indipendentemente dai valori dei parametri d'errore utilizzati per simulare i dati. Questo ha permesso di testare la capacità degli indici di selezione dei modelli nell'individuare il modello che ha generato i dati (modello *vero*), quando i parametri d'errore sopra e sotto il cutoff sono più o meno vicini l'uno all'altro. Infine la differenza tra i parametri d'errore sopra e sotto il cutoff è stata testata

applicando il test di Wald.

4.4.1 Disegno delle simulazioni e stima dei parametri

La struttura di conoscenza utilizzata nelle simulazioni è stata costruita adottando l'approccio *competence-performance* (Doignon, 1994; Korossy, 1999), descritto nel Capitolo 2, ed è stata ottenuta a partire dalla skill function congiuntiva illustrata nella Tabella 4.2, che associa 11 abilità a 18 problemi. Si è deciso di utilizzare questa particolare skill function perché è la stessa del-

Tabella 4.2: Skill function congiuntiva utilizzata per delineare la struttura di conoscenza delle simulazioni e dell'applicazione empirica. Per ciascun problema si riporta l'insieme delle abilità ipotizzate essere alla base della sua soluzione.

Problemi	Stati di Competenza	Problemi	Stati di Competenza
1	{F,FF}	10	{F,FF,BC,C,U,P1}
2	{F,FF,BC}	11	{B}
3	{F,FF,BC,C}	12	{F,FF,BC,C,P1,B}
4	{F,D}	13	{P,RC}
5	{F,FF,BC,C,U}	14	{F,FF,BC,C,U,P1,P2,B}
6	{F,FF,D}	15	{P,RC}
7	{F,FF,BC,C,D,P1,P2}	16	{F,FF,BC,C,P1,P2,P,CR}
8	{F,FF,BC,C,P1}	17	{F,FF,BC,C,U,P1,P2,P,CR}
9	{F,FF,BC,C,D,P1,P2}	18	{F,FF,BC,C,U,P1,P2,P,CR}

F = Calcolare il fattoriale di un numero, FF = ridurre una frazione contenente fattoriali, BC = calcolare il coefficiente binomiale, C = risolvere problemi sulle combinazioni semplici, D = risolvere problemi sulle disposizioni, U = applicare la regola di unione di eventi, B = applicare la regola delle prove Bernoulliane, P = calcolare la probabilità di un evento, CR = applicare la regola di concatenazione, P1 = risolvere problemi che richiedono lo svolgimento in almeno due passaggi e P2 = risolvere problemi che richiedono lo svolgimento in tre passaggi.

l'applicazione empirica (Sezione 4.5). I 18 problemi di cui si compone sono stati costruiti per valutare la conoscenza nell'area della *teoria della probabi-*

lità e le 11 abilità ad essi associati sono state ritenute necessarie per la loro risoluzione. Le abilità sono le seguenti: calcolare il fattoriale di un numero (F), ridurre una frazione contenente fattoriali (FF), calcolare il coefficiente binomiale (BC), risolvere problemi sulle combinazioni semplici(C) e sulle disposizioni (D), applicare la regola di unione di eventi (U), applicare la regola delle prove Bernoulliane (B), calcolare la probabilità di un evento(P), applicare la regola di concatenazione(CR), risolvere problemi che richiedono lo svolgimento in almeno due passaggi (P1) o tre passaggi(P2).

Per fare un esempio, si consideri il problema numero 11:

> *Si lanci una moneta truccata per 9 volte. Se la probabilità di ottenere croce (C) è 2/3 ad ogni lancio, qual'è la probabilità di ottenere esattamente la sequenza TCCTCTCCT?*

Per risolvere correttamente questo problema è necessario conoscere solamente l'applicazione della regola delle prove Bernoulliane.

La skill function congiuntiva definita in Tabella 4.2 ha delineato una struttura di conoscenza con 69 stati di conoscenza.

Per quanto riguarda il disegno delle simulazioni, i campioni casuali sono stati generati sotto sei differenti condizioni, utilizzando un numero fisso di 500 campioni, ciascuno composto da 1,000 pattern di risposta. Anche le probabilità degli stati di conoscenza sono state fissate attraverso le condizioni, per facilitarne il confronto. La Tabella 4.3 riassume il disegno delle simulazioni e descrive nello specifico ciò che varia tra le condizioni.

Tabella 4.3: Disegno delle Simulazioni

Condizione	Modello Generativo	Intervalli dei Parametri			
C1	BLIM			(0,.3]	
Da C2 a C5	BPM	(0,.1]	(0,.2]	(0,.3]	(0,.5]
C6	GDM			(0,.3]	

Dalla Tabella si può notare che, ciò che varia attraverso le condizioni, sono i valori utilizzati per i parametri d'errore. Essi sono stati generati da una distribuzione uniforme nell'intervallo $(0, a]$, dove il limite superiore a variava a seconda della condizione. Entrando più in dettaglio nella condizione 1 il modello che generava i dati era il BLIM e i valori dei parametri β e η erano nell'intervallo (0,.3]. Nelle condizioni dalla 2 alla 5, il modello che generava i dati era il BPM e gli intervalli dei parametri erano rispettivamente (0,.1], (0,.2], (0,.3] e (0,.5]. In questo modo, nelle diverse condizioni, la differenza massima tra i parametri sotto e sopra il cutoff poteva essere .1, .2, .3 or .5. E' stato scelto di utilizzare diversi intervalli per testare la sensibilità del BPM nell'individuare le differenze tra parametri sopra e sotto il cutoff, al crescere della loro differenza. Inoltre, per avere la stessa media sui parametri sopra e sotto il cutoff, prima sono stati generati casualmente i parametri sopra, e poi è stata fatta una permutazione di questi valori per generare i parametri sotto il cutoff. L'ultima condizione, la condizione C6, è stata considerata per valutare la capacità del BPM nel testare le dipendenze quando si osservano nel caso più generale possibile, ovvero quando i parametri d'errore variano attraverso tutti gli stati di conoscenza. In questo caso il modello generativo era dunque il GDM (4.3). Per generare i parametri si è utilizzato, in questo caso, solo l'intervallo (0,.3] .

Per quanto riguarda le stime dei parametri, in entrambi i modelli sono stati stimati per massima verosimiglianza utilizzando l'algoritmo expectation-maximization (EM), in ognuno dei $500 \times 6 = 3,000$ campioni casuali.

4.4.2 Fit e selezione dei modelli

Il BLIM e il BPM sono stati stimati su ognuno dei 3,000 data set, e la loro fit è stata confrontata attraverso i seguenti indici di selezione: LR, AIC, $AICc$ e BIC. Il test LR è stato calcolato come segue:

$$LR = -2ln\frac{L_{BLIM}}{L_{BPM}},$$

dove L_{BLIM} e L_{BPM} sono i valori della funzione di verosimiglianza dei modelli stimati. Tipicamente, questa statistica si approssima a una distribuzione di Chi-quadro con $df = df_{BLIM} - df_{BPM}$ gradi di libertà. L'indice AIC dei due modelli è stato calcolato secondo la formula

$$AIC = 2k - 2ln(L),$$

dove k è il numero dei parametri del modello considerato e L è la massimizzazione della funzione di verosimiglianza. L'indice $AICc$ è stato calcolato secondo la formula

$$AICc = AIC + \frac{2k(k+1)}{n-k-1},$$

dove n è la numerosità campionaria. Infine il BIC è stato calcolato secondo l'equazione

$$BIC = -2ln(L) + kln(n).$$

Nel caso degli ultimi tre indici, per ciascuna delle 3,000 replicazioni dello studio, sono state poi calcolate le differenze

$$\Delta_{AIC} = AIC_{BPM} - AIC_{BLIM}$$
$$\Delta_{AICc} = AICc_{BPM} - AICc_{BLIM}$$
$$\Delta_{BIC} = BIC_{BPM} - BIC_{BLIM}.$$

Infine, in ognuna delle sei condizioni è stata costruita la distribuzione Δ_{AIC} sulle 500 repicazioni ed è stata calcolata la proporzione di volte in cui $\Delta_{AIC} \leq 0$, ad indicare che il BPM si addatta ai dati meglio del BLIM. Per Δ_{AICc} e Δ_{BIC} è stata applicata la stessa procedura.

4.4.3 Test di Wald per i parametri del BPM

In tutte le condizioni in cui i modelli generativi erano il BPM o il GDM (dalla $C2$ alla $C6$) è stato applicato il test di Wald per testare la differenza delle stime dei parametri sopra e sotto il cutoff. L'obiettivo era di valutare

la potenza del test nell'individuare differenze significative tra i due gruppi di parametri, indipendentemente dall'intervallo usato per generarli.

In ognuna delle 2,500 replicazioni d'interesse la statistica di Wald è stata calcolata come segue. Siano, rispettivamente

$$\phi^- = (\beta_1^-, \beta_2^-, \ldots, \beta_n^-, \eta_1^-, \eta_2^-, \ldots, \eta_n^-)^T$$

e

$$\phi^+ = (\beta_1^+, \beta_2^+, \ldots, \beta_n^+, \eta_1^+, \eta_2^+, \ldots, \eta_n^+)^T$$

i vettori dei parametri sotto e sopra il cutoff. Inoltre, siano $\mathbf{\Sigma}^-$ e $\mathbf{\Sigma}^+$, rispettivamente, le matrici di covarianza dei parametri sopra e sotto il cutoff. Allora la statistica di Wald è:

$$W = (\phi^- - \phi^+)^T \mathbf{\Sigma}^{-1} (\phi^- - \phi^+),$$

dove $\mathbf{\Sigma} = \mathbf{\Sigma}^- + \mathbf{\Sigma}^+$. Questo procedimento permette di avere un test della DIF, perché W è asintoticamente χ^2-distribuito con $df = n - 1$ gradi di libertà, dove n è il numero di item per i quali si fa il confronto. I valori $\mathbf{\Sigma}^-$ e $\mathbf{\Sigma}^+$ sono stati approssimati attraverso una procedura bootstrap. Infine, in ognuna delle cinque condizioni, è stata calcolata la proporzione di replicazioni per le quali il test di Wald è risultato significativo.

4.4.4 Risultati

Il grafico in Figura 4.1 mostra sei curve, una per ciascuna condizione. Le curve sono state ottenute nel modo seguente: (1) è stato calcolato il p-value teorico relativo al rapporto di verosimiglianza, per ciascuna delle 500 replicazioni; (2) i 500 p-value sono stati ordinati dal più piccolo al più grande; (3) sono state costruite le curve empiriche. Nel grafico, sull'asse delle x si trova il p-value mentre su quello delle y ci sono le proporzioni cumulate.

Dalla figura si può notare che, nella condizione C1 il 6% delle replicazioni ottiene un p-value minore o uguale a .05, a significare che, quando il modello

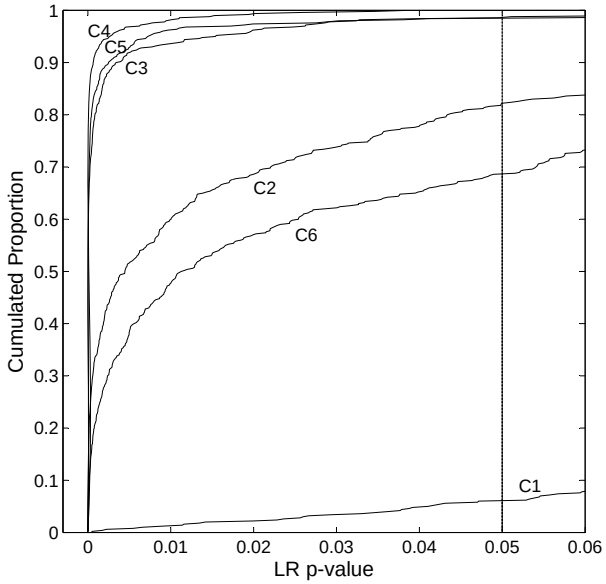

Figura 4.1: Distribuzione cumulativa bootstrap del p-value relativo all'indice LR: confronto attraverso le sei condizioni dello studio simulativo. La linea continua verticale è il riferimento per p-value = .05.

generativo è il BLIM, il BPM si adatta ai dati meglio di quanto faccia il BLIM solamente nel 6% dei casi. Questo primo risultato conferisce validità all'indice LR, dal momento che il p-value bootstrap (6%) è molto vicino a quello teorico (5%).

Nelle condizioni dalla C2 alla C4, nelle quali il modello generativo era il BPM, la proporzione di replicazioni in cui il p-value è minore uguale .05 è monotona crescente in funzione della differenza tra i parametri d'errore sopra e sotto il cutoff. In altre parole, maggiore è la differenza tra parametri sotto e sopra il cutoff, maggiore è la proporzione di replicazioni in cui il BPM ottiene una fit migliore del BLIM: .82 nella condizione C2, .99 nella condizione C3 e 1.00 nella condizione C4. Per quanto riguarda la condizione C5, dove il modello generativo era sempre il BPM ma l'intervallo usato per generare i parametri era piuttosto ampio $(0, .5]$, questa proporzione è .98.

Nella condizione C6, dove il modello generativo era il GDM, nel 69% delle replicazioni il p-value dell'indice LR era minore o uguale .05, ad indicare che,

in tutti questi casi, il BPM si adatta meglio ai dati di quanto non faccia il BLIM.

Nell'insieme, questi primi risultati indicano che il BPM sembra capace di individuare le dipendenze tra i parametri sopra e sotto il cutoff, anche quando la loro distanza è molto piccola (C2) e il numero di dipendenze dei parametri d'errore dagli stati di conoscenza, è massimo (C6). Inoltre, i risultati conferiscono attendibilità all'indice LR, quando viene utilizzato per operare il confronto tra il BLIM e il BPM. Gli altri indici di selezione dei modelli non hanno dato prova di essere altrettanto attendibili.

Nella Figura 4.2, il pannello di sinistra mostra la distribuzione cumulativa bootstrap di $\Delta_{AIC} = AIC_{BPM} - AIC_{BLIM}$, ottenuta in ognuna delle sei condizioni dello studio. Nella figura, sull'asse delle ascisse si collocano i valori di Δ_{AIC} mentre sull'asse delle ordinate si trova la proporzione cumulativa delle 500 replicazioni bootstrap. Ciascuna curva del diagramma rappresenta una specifica condizione dello studio.

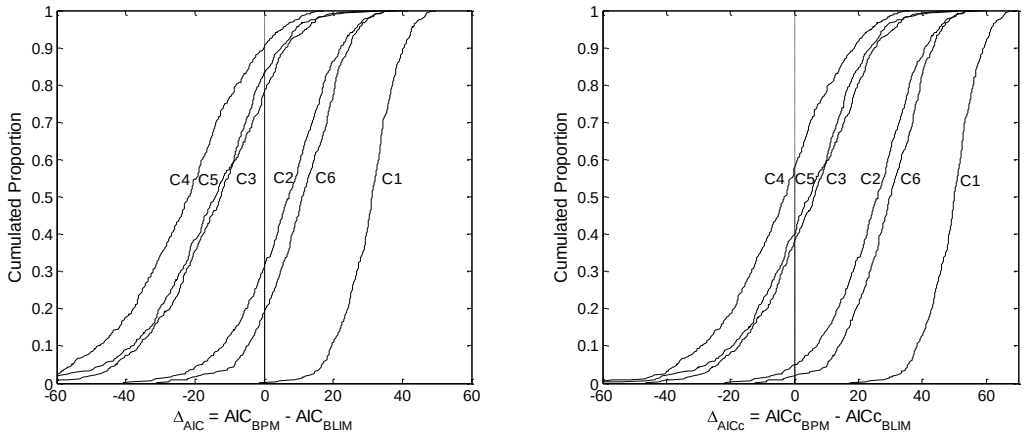

Figura 4.2: Distribuzione cumulativa bootstrap relativa a Δ_{AIC} (pannello di sinistra) e a Δ_{AICc} (pannello di destra): confronto attraverso le sei condizioni dello studio simulativo. La linea verticale indica una differenza pari a zero. Valori negativi indicano una preferenza per il BPM, mentre valori positivi indicano una preferenza per il BLIM.

Si può notare che, nella condizione C1 il Δ_{AIC} è sempre positivo a si-

gnificare che, quando il modello generativo è il BLIM il criterio di selezione AIC riconosce correttamente il modello vero. Nelle condizioni dalla C2 alla C5, dove il modello generativo era il BPM, ci si aspettava invece che il Δ_{AIC} fosse negativo, ad indicare il BPM ottiene un a fit migliore del BLIM. Le proporzioni delle replicazioni in cui questo accade non sono invece così alte: .31 in C2, .78 in C3, .90 in C4 and .83 in C5. Nella condizione C6, dove il modello generativo era il GDM, le cose peggiorano ulteriormente: il Δ_{AIC} è negativo solamente nel 19% dei casi.

Il pannello di destra della Figura 4.2 mostra il confronto tra le sei condizioni quando l'indice di selezione dei modelli è l'$AICc$. Le considerazione che si possono fare per questo indice sono simili, se non peggiori, di quelle fatte per l'AIC.

L'ultimo indice qui considerato è il BIC. Questo indice ottiene risultati ancora peggiori rispetto all'AIC e all' $AICc$. Infatti, il Δ_{BIC} è positivo in tutte le condizioni, a significare che il BIC seleziona il modello più semplice, ovvero il BLIM, anche quando il modello generativo è il BPM o il GDM.

Le ultime analisi condotte per lo studio simulativo, riguardavano il test di Wald. In particolare è stata calcolata la proporzione delle volte in cui il test è risultato significativo, ad indicare una differenza statistica tra i parametri del BPM sopra e sotto il cutoff. Le condizioni in esame sono quelle dalla C2 alla C5, per le quali questa proporzione è pari a .65 per la C2, .94 per la C3, .99 per la C4 e .68 per la C5. Ad eccezione della C5, si osserva che al crescere della differenza tra i parametri sotto e sopra il cutoff (si ricorda che le condizioni C2, C3, C4 e C5 variano proprio per la grandezza dei parametri), cresce anche la proporzione di volte in cui il test di Wald è significativo. Invece, nella condizione C5, dove la distanza è massima, il test si comporta come nella condizione C2, dove la differenza è la più piccola. Una possibile spiegazione può riguardare la potenza del test, che è influenzata dalla distanza tra i parametri, ma anche dalla varianza delle stime dei parametri. Più alta è la distanza, più alta è la potenza del test e più è alta la varianza dei parametri

più bassa è la potenza del test. Di fatto, ciò che si osserva è spiegato proprio da quest'ultimo punto: aumentando l'intervallo di generazione dei parametri, la varianza media delle stime dei parametri aumenta: .0022 in C2, .0054 in C3, .01 in C4 e .04 in C5.

4.5 Applicazione Empirica

4.5.1 Partecipanti e metodi

La struttura di conoscenza impiegata nello studio simulativo, è stata utilizzata anche per l'applicazione empirica. I 18 item sulla teoria della probabilità sono stati presentati, in ordine casuale e attraverso una procedura computerizzata, a 209 studenti di psicologia dell'Università di Padova, iscritti all'anno accademico 2011–2012. Le risposte ai problemi sono state classificate come corrette (1) ed errate (0) e sono state utilizzate per costruire la matrice dei dati.

Sia il BLIM che il BPM sono stati applicati ai dati e i loro parametri sono stati stimati per massima verosimiglianza, usando l'algoritmo EM. La bontà di adattamento ai dati del BLIM è stata testata usando il Chi-quadro di Pearson e, a causa della matrice dei dati sparsa, il p-value è stato calcolato con una procedura bootstrap su 500 replicazioni. Per testare l'assunzione di invarianza del BLIM, questo modello è stato confrontato con il BPM utilizzando gli indici LR e AIC. Infine, nel caso del BPM è stato applicato il test di Wald per individuare differenze tra i parametri sotto e sopra il cutoff.

4.5.2 Risultati

La bontà di adattamento del BLIM è buona, infatti il p-value del Chi-quadro è pari a .47. Per essere certi però che l'assunzione di invarianza del modello non sia violata dai dati, è stato applicato anche il BPM. Dal confronto di questi due modelli emerge che l'indice $LR = 81.61$ ($df = 32$) è significativo

(p-value bootstrap = .002), a indicare che il BPM spiega i dati meglio di quanto faccia il BLIM. Questo è indice della presenza di qualche dipendenza tra i parametri d'errore e gli stati di conoscenza. In linea con l'indice LR, anche l'AIC seleziona il BPM ($AIC_{BLIM} = 3241.9$ e $AIC_{BPM} = 3224.3$). Per quanto riguarda il test di Wald, anch'esso conferma quanto già evidenziato ($W = 100.12, df = 35, p = .001$).

L'aspetto più importante da sottolineare è che, nonostante il BLIM ottenga una fit decisamente buona, è chiaro che l'assunzione di invarianza è violata dai dati. Pertanto non è appropriato concludere che i parametri d'errore degli item siano costanti attraverso gli stati di conoscenza.

4.6 Discussione

In questo capitolo è stata presentata una procedura per testare una delle assunzioni del BLIM, il modello più utilizzato nell'ambito della KST. Questa assunzione riguarda l'invarianza dei parametri d'errore β e η del modello, dagli stati di conoscenza. Da un punto di vista empirico tale assunzione stabilisce che le probabilità di distrazione (parametro careless error) e di indovinare la risposta (lucky guess) sono invarianti attraverso gli stati di conoscenza; essi sarebbero invece una proprietà intrinseca dell'item.

E' stato evidenziato come, nonostante un modello ottenga una buona fit ai dati, non va trascurata la verifica delle assunzioni su cui si basa, che possono essere, comunque, violate dai dati. Dapprima è sembrato ragionevole trasferire al caso del BLIM, i metodi sviluppati per testare lo stesso tipo di assunzione in altri ambiti (come l'IRT e i CDM). E' stato però evidenziato come questi metodi portino ad ottenere dei bias nelle stime dei parametri del modello, anche quando sia quello corretto. In particolare, le lucky guess vengono sovrastimate nel campionamento sopra a un certo cutoff e sottostimate sotto, mentre le careless error vengono sottostimate nel gruppo sopra il cutoff e sovrastimate sotto.

Per queste ragioni si è cercato un metodo alternativo che fosse più in linea con il modellamento formale, ovvero di operare un confronto tra il BLIM e un modello alternativo, più complesso, in cui l'assunzione di invarianza è esplicitamente violata. Esistono molti modelli di questo tipo, infatti si può sviluppare una gerarchia completa che va dal BLIM (assenza di dipendenze) a un modello generale di dipendenza (GDM) in cui tutti i parametri d'errore sono liberi di variare con gli stati di conoscenza. In questa ricerca si è scelto di utilizzare la famiglia di modelli più semplice in questa gerarchia, che si colloca, in termini di complessità, subito dopo al BLIM. Ciascun modello della famiglia corrisponde a una partizione della struttura di conoscenza in due classi (bipartizione): tutti gli stati la cui cardinalità è minore o uguale a un certo cutoff, e tutti i restanti. Nei modelli a bipartizione (BPMs) l'assunzione di invarianza è rispettata all'interno della classe ma non attraverso le classi. Essi si caratterizzano per un insieme di quattro parametri (anziché due come nel caso del BLIM) per ciascun item: le probabilità di careless error e lucky guess sotto il cutoff e le probabilità di careless error e lucky guess sopra il cutoff.

Dal momento che il BLIM è annidato nel BPM, un modo per testare l'assunzione di invarianza, è dato dal rapporto di verosimiglianza (LR), dove il BLIM è l'ipotesi nulla e il BPM è l'ipotesi alternativa. Nel confronto dei due modelli sono stati utilizzati anche il test di Wald e i criteri standard di selezione dei modelli come l'AIC e il BIC.

Si è voluto operare il confronto attraverso delle simulazioni, sia in un contesto realistico, attraverso un applicazione empirica. I risultati delle simulazioni hanno evidenziato come il confronto del BLIM con un BPM, sia un modo efficace per individuare le violazioni dell'assunzione di invarianza. Gli indici LR e AIC selezionano il modello corretto: quando l'assunzione è rispettata dai dati selezionano il BLIM, e quando l'assunzione è violata selezionano il BPM. Ciò che ha evidenziato l'applicazione empirica è che, nonostante il BLIM ottenga una buona fit ai dati, l'assunzione di invarianza

dei suoi parametri d'errore dagli stati di conoscenza può comunque essere violata.

Nell'insieme i risultati della ricerca suggerisco che un test generale della bontà di adattamento del BLIM non è sufficiente per conferire validità empirica a tutte le sue assunzioni. Sono necessari altri test, sviluppati specificatamente per individuare le violazioni delle assunzioni del modello, e i BPM hanno dato prova di essere uno strumento promettente nel caso dell'assunzione di invarianza.

In ricerche future sarà interessante studiare l'effetto della distribuzione di probabilità degli stati di conoscenza sull'assunzione. E' lecito infatti pensare che, in una popolazione di studenti particolarmente bravi sia meno probabile la distrazione, che, in una popolazione di studenti all'inizio di un percorso di studio potrebbe essere invece più alta.

Infine va sottolineata la potenzialità dei modelli di dipendenza come modelli in sé: non devono essere visti solamente come uno strumento facente parte di un metodo per testare una particolare assunzione, ma potrebbero essere usati, ad esempio, per stimare le probabilità di distrazione e di lucky guess di ciascun soggetto. In alcuni ambiti, come quello diagnostico, potrebbe essere un valido aiuto per individuare gli studenti con particolari disturbi o difficoltà.

Capitolo 5

Modellare i Dati Mancanti nella Knowledge Space Theory

I dati mancanti sono un problema ben noto nell'ambito dell'inferenza statistica. Questo perché, anche quando viene data massima attenzione alla fase della raccolta dei dati, le risposte mancanti a uno o più item di un test, sono piuttosto frequenti. Le ragioni sottostanti ad una risposta mancante, possono essere diverse, per esempio perché il soggetto non sa dare una risposta, o perché salta accidentalmente l'item, o, ancora, per carenza di tempo nel completare il test, o perché il test viene somministrato utilizzando una forma adattiva, e così via. In tutti questi casi, il nodo cruciale è capire come trattare le risposte mancanti nella fase delle analisi statistiche.

Nonostante il problema si presenti anche nelle applicazioni della knowledge space theory (KST), la questione non è mai stata affrontata in modo approfondito. Una ragione possibile di questa mancanza può dipende dalla tipologia dei test che si sviluppano in questo ambito, ovvero test di profitto, dove l'obiettivo è valutare le conoscenze di uno studente attraverso una serie di problemi. Così, qualora un campione presenti risposte mancanti, sembra lecito assumere che se lo studente non dà risposta ad un problema è perché non lo sa fare e per tale ragione questo tipo di risposte vengono ricodificate come risposte errate.

Nella ricerca che si presenta in questo capitolo, tratta dall'articolo di de Chiusole, Stefanutti, Anselmi, e Robusto (2014), ci si è chiesti se un'azione simile fosse lecita, o se potesse invece portare a risultati fuorvianti nel momento dell'applicazione di un modello.

5.1 Introduzione

Il BLIM, presentato nella Sezione 1.3, è un modello che non offre alcun meccanismo per ovviare al problema dei dati mancanti, ma può essere applicato solamente nel caso in cui si abbia a disposizione un'informazione completa sulle risposte degli studenti a un insieme di item. Unica soluzione a questo problema è optare per una qualche trasformazione dei dati che consenta di modificare le risposte mancanti in un dato osservato. Se ad esempio si è ragionevolmente sicuri che una risposta mancante sottenda una risposta errata, allora tutti i dati mancanti di un campione possono essere ricodificati in risposte errate, e il BLIM può essere applicato ai dati sui quali è stata fatta tale trasformazione. Comunque, in generale, assumere che una risposta non data sottenda sempre una risposta errata, non è del tutto realistico e, come si mostrerà nella Sezione 5.3, può portare a osservare dei bias nella stima dei parametri del BLIM.

Un ambito dove il problema dei dati mancanti è stato studiato a lungo è l'item response theory (Glas & Pimentel, 2008; Holman & Glas, 2005; Little & Rubin, 2002; Mislevy & Chang, 2000; Schafer, 1997). Purtroppo i metodi sviluppati in quest'ambito non possono essere applicati alla KST, per una sostanziale differenza tra i due approcci, la più importante delle quali dipende dal fatto che l'IRT è un approccio numerico, mentre la KST è una teoria non-numerica.

Non potendo fare riferimento ai metodi utilizzati in altri ambiti, si è partiti dall'analisi della letteratura tradizionale legata ai dati mancanti. Uno

dei ricercatori che per primo ha studiato questo problema è Rubin (1976), che affermò:

> *No decision should be taken about the treatment of missing data, without considering the particular underlying process that generated them.*

Secondo questa affermazione, sarebbe fuorviante trattare i dati mancanti utilizzando un qualsiasi metodo, senza fare alcuna considerazione riguardo il processo che li ha generati. Egli ha introdotto le definizioni formali, ancora in uso nella letteratura scientifica contemporanea, di differenti processi alla base delle risposte mancanti. Sia $Y_{com} = (Y_{obs}, Y_{miss})$ l'insieme dei dati completi, composti dalla parte osservata Y_{obs} e dalla parte mancante Y_{miss}. Inoltre, sia M la variabile che distingue ciò che si conosce da ciò che è mancante. Adottando la terminologia di Rubin (1976) i dati mancanti possono essere:

- *missing at random* (MAR) se la distribuzione sottesa dai dati mancanti non dipende da Y_{miss}, ovvero $P(M|Y_{com}) = P(M|Y_{obs})$;
- *missing completely at random* (MCAR), un caso speciale di MAR, se la distribuzione non dipende né da Y_{obs} nè da Y_{miss}, ovvero $P(M|Y_{com}) = P(M)$;
- *missing not at random* (MNAR), se la distribuzione dei mancanti dipende da Y_{miss}.

Un modello probabilistico che assume un processo MAR (o MCAR) alla base dei mancanti e secondo cui i parametri non siano funzionalmente dipendenti dai dati mancanti, viene chiamato *ignorable*. Invece, un modello probabilistico che assume un processo MNAR o che introduce specifiche dipendenze dei dati mancanti dai parametri del modello è chiamato *nonignorable* (Little & Rubin, 2002; Schafer, 1997; Schafer & Graham, 2002).

Tornando alla KST, a questo punto emergono due domande importanti:

1. In che modo si possono modellare i dati mancanti di tipo *ignorable* in un modello simile al BLIM?

2. Come si possono modellare i dati mancanti nel caso in cui siano di tipo *nonignorable*?

In questo capitolo si cercherà di dare risposta a questi due quesiti. Si propongono a questo scopo due estensioni del BLIM (Sezione 5.2): la prima, chiamata IMBLIM (Ignorable Missingness BLIM), assume che il processo che genera i dati mancanti sia del tutto casuale, ovvero di tipo *ignorable*; la seconda estensione, chiamata MissBLIM (Missingness BLIM), fa invece specifiche assunzioni di dipendenza dei dati mancanti dagli stati di conoscenza degli studenti, per questa ragione è di tipo *nonignorable*. Entrambi i modelli sono stati oggetto di uno studio simulativo (Sezione 5.4) e di un'applicazione empirica (sezione 5.5). I risultati mostrano che sia l'IMBLIM che il MissBLIM modellano i dati mancanti in modo soddisfacente, a seconda del processo che li genera: se i mancanti sono di tipo *ignorable*, entrambi sono adeguati, ma se sono di tipo *nonignorable*, sembra più opportuno utilizzare il MissBLIM.

5.2 Due Estensioni del BLIM per Dati Mancanti

In questa sezione si derivano le due estensioni del BLIM per dati mancanti. La prima estensione, chiamata IMBLIM si basa sull'assunzione che il processo che genera i dati mancanti sia di tipo MCAR. Nella seconda estensione, chiamata MissBLIM, i dati mancanti vengono invece parametrizzati ed incorporati nel modello, e per questo motivo il modello è appropriato quando il processo che genera i mancanti è MNAR.

In entrambe le estensioni, così come nel BLIM, si denota con $R \subseteq Q$ la collezione di tutti i problemi, nel dominio di conoscenza Q, che uno studente

risolve correttamente. In aggiunta a questa notazione, si indica con $M \subseteq Q \backslash R$ l'insieme dei problemi per i quali non si osserva risposta (dato mancante) e con W la collezione dei problemi per i quali si osserva una risposta errata. In questo modo, per ciascun soggetto, i dati sono una terna $\langle R, M, W \rangle$ di sottoinsiemi disgiunti di Q. Dal momento che l'insieme W è totalmente deducibile dagli altri due ($W = Q \setminus (R \cup M)$), verrà omesso in tutta la sezione. Per rappresentare i dati osservati si utilizzerà invece la notazione $\langle R, M \rangle$. Per ogni studente, l'insieme R si riferisce al *pattern di risposta parziale*, e M al *pattern di risposta mancante*.

Indicando con $P(R, M)$ la probabilità di osservare, in uno studente estratto casualmente, la coppia $\langle R, M \rangle$, l'equazione (1.4) del BLIM può essere riscritta, nel caso di dati mancanti, nel modo seguente

$$P(R, M) = \sum_{K \in \mathcal{K}} P(R, M|K)\pi_K, \tag{5.1}$$

dove $P(R, M|K)$ è la probabilità condizionale di osservare $\langle R, M \rangle$, dato lo stato di conoscenza K. Questa equazione costituisce il punto di partenza della derivazione di entrambe le estensione del BLIM per dati mancanti. La verosimiglianza dei modelli, per un campione finito $\mathcal{D}$ di numerosità N assume la forma generale

$$L'(\theta'|\mathcal{D}) = C \prod_{\langle R,M \rangle \in \mathcal{O}} P(R, M)^{F(R,M)} \tag{5.2}$$

dove:

- C è una costante che non dipende dai parametri del modello;
- θ' è il vettore dei parametri del modello;
- $\mathcal{O} = \{\langle R, M \rangle \in 2^Q \times 2^Q : R \cap M = \emptyset\}$ è la collezione teorica delle coppie osservabili $\langle R, M \rangle$;
- $F(R, M)$ è la frequenza osservata della coppia $\langle R, M \rangle$ nel campione.

5.2.1 Dati Mancanti *Ignorable*: l'IMBLIM

L'approccio tipico per postulare un modello probabilistico per dati mancanti consiste nel separare il *modello completo* dal *meccanismo di decimazione*. Il primo specifica le probabilità dei dati completi, mentre il secondo specifica le probabilità di un pattern di risposta mancante, dato il pattern completo. Il processo che genera la coppia $\langle R, M\rangle$ diventa quindi un processo dove si genera il pattern di risposta completo $R^* \subseteq Q$, usando il modello probabilistico per dati completi, e su questo si compie poi la decimazione, in accordo con il relativo meccanismo. Il risultato è la coppia osservabile $\langle R, M\rangle$, dove $M \subseteq Q$ è la collezione dei problemi selezionati dal meccanismo di decimazione e $R = R^* \setminus M$ è il pattern delle risposte restanti.

E' ragionevole assumere: (1) che il processo evidenziato sopra sia governato dalla probabilità $P(R^*, M|K)$ di ottenere il pattern di risposta completo R^* e il pattern M delle risposte mancanti, dato lo stato di conoscenza $K \in \mathcal{K}$; (2) che il meccanismo di decimazione sia indipendente dal processo che genera i dati completi. Questi due aspetti sono fondamentali nella derivazione di un modello per dati MCAR. Da un punto di vista formale è dunque necessario introdurre le due assunzioni che seguono.

[**A6.1**]La probabilità condizionale $P(R^*, M|K)$ si scompone come segue

$$P(R^*, M|K) = P(R^*|K)P(M),$$

dove $P(R^*|K)$ è la probabilità del pattern di risposta completo R^* dato lo stato di conoscenza K, ottenuto dall'applicazione dell'Equazione (1.6) del BLIM, e $P(M)$ è la probabilità di non osservare risposta per gli item contenuti nell'insieme M.

[**A6.2**] La probabilità $P(M)$ del pattern di risposta mancante $M \subseteq Q$ è funzionalmente indipendente dai parametri β_q e η_q degli item.

Dalle due assunzioni [**A6.1**] e [**A6.2**] appena introdotte, per le due pro-

babilità congiunte R^* e M, si ottiene il modello seguente:

$$P(R^*, M) = P(M) \sum_{K \in \mathcal{K}} P(R^*|K)\pi_K. \tag{5.3}$$

Questo modello è adatto per fare previsioni sulla coppia osservabile $\langle R^*, M\rangle$. Per fare previsioni circa i pattern incompleti $\langle R, M\rangle$, si può ottenere la probabilità marginale per $\langle R, M\rangle$, considerando la collezione

$$\langle R, M\rangle^* = \{R \cup C : C \subseteq M\}$$

di tutti i possibili pattern di risposta completi che sono *compatibili* con la coppia incompleta $\langle R, M\rangle$. Questa collezione contiene tutti i possibili modi di completare il pattern di risposta parziale R, attraverso l'unione di quest'ultimo con tutti i sottoinsiemi dei pattern di risposta mancanti M. Allora la probabilità di $\langle R, M\rangle$ corrisponde alla probabilità di osservare una qualsiasi combinazione di R nella collezione $\langle R, M\rangle^*$:

$$\begin{aligned} P(R, M) &= \sum_{R^* \in \langle R,M\rangle^*} P(R^*, M) \\ &= P(M) \sum_{K \in \mathcal{K}} \sum_{R^* \in \langle R,M\rangle^*} P(R^*|K)\pi_K. \end{aligned}$$

Ora, definendo la probabilità

$$P(\langle R, M\rangle^*|K) = \sum_{R^* \in \langle R,M\rangle^*} P(R^*|K), \tag{5.4}$$

si può dimostrare che

$$P(\langle R, M\rangle^*|K) = \left[\prod_{q \in K \setminus (R \cup M)} \beta_q\right] \left[\prod_{q \in (K \cap R) \setminus M} (1 - \beta_q)\right] \left[\prod_{q \in R \setminus (K \cup M)} \eta_q\right] \left[\prod_{q \in Q \setminus (R \cup M \cup K)} (1 - \eta_q)\right] \tag{5.5}$$

ogni volta che $R \cap M = \emptyset$, e $P(\langle R, M\rangle^*|K) = 0$ altrimenti. Si può notare che, ciò che cambia rispetto all'equazione (1.4) sono i pedici delle produttorie,

dove si aggiunge l'insieme M. Qualora $M = \emptyset$ l'Equazione (5.5) dell'IMBLIM e l'Equazione (1.4) del BLIM, darebbero lo stesso risultato.

Per fare un esempio dell'uguaglianza tra il termine di destra dell'Equazione (5.4) e quello dell'Equazione(5.5), si consideri i seguenti insiemi:

- l'insieme degli item $Q = \{a, b, c, d\}$;
- il pattern di risposta completo generato dal BLIM $R^* = \{a, b\}$, corrispondente allo stato di conoscenza $K = \{a, c\}$;
- il pattern di risposta mancante (selezionato da un qualche meccanismo di decimazione) $M = \{b, c\}$.
- il pattern di risposta osservato sarà $R = R^* \setminus M = \{a\}$.

Ora, ci sono quattro diversi pattern di risposta completi compatibili con il pattern di risposta incompleto $R = \{a\}$ e il pattern di risposta mancante $M = \{b, c\}$, che sono:

$$\langle R, M\rangle^* = \{\{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}\}.$$

Naturalmente R^* stesso appartiene a questa collezione. Allora, applicando l'Equazione (5.4),

$$P(\langle R, M\rangle^*|K) = P(\{a\}|K) + P(\{a, b\}|K) + P(\{a, c\}|K) + P(\{a, b, c\}|K)$$

e se si applica l'Equazione (1.6) a ciascun termine dell'equazione appena vista, si ottiene:

$$\begin{aligned} P(\langle R, M\rangle^*|K) = {} & (1 - \beta_a)(1 - \eta_b)\beta_c(1 - \eta_d) + (1 - \beta_a)\eta_b\beta_c(1 - \eta_d) + \\ & + (1 - \beta_a)(1 - \eta_b)(1 - \beta_c)(1 - \eta_d) + (1 - \beta_a)\eta_b(1 - \beta_c)(1 - \eta_d) \end{aligned}$$

che, semplificata diventa

$$P(\langle R, M\rangle^*|K) = (1 - \beta_a)(1 - \eta_d),$$

che non dipende dai parametri d'errore degli item appartenenti a M. Lo stesso identico risultato si ottiene se si parte dall'Equazione (5.5), anziché dall'Equazione (5.4).

Per riassumere, nel modello IMBLIM, l'Equazione (5.1) supportata dalle assunzioni [**A6.1**] e [**A6.2**] implica che la probabilità di ogni singola coppia $\langle R, M\rangle$ sia

$$P(R, M) = P(M) \sum_{K \in \mathcal{K}} P(\langle R, M\rangle^* | K)\pi_K, \tag{5.6}$$

dove $P(\langle R, M\rangle^* | K)$ è come in Equazione (5.5), e la probabilità $P(M)$ non dipende dai parametri β_q, η_q e π_K del modello.

Un'ultima nota sulla relazione tra il BLIM e l'IMBLIM. Nel caso in cui non si osservano dati mancanti, allora $M = \emptyset$. In questo caso, dall'Equazione (5.5), se $M = \emptyset$, la probabilità condizionale $P(\langle R, M\rangle^* | K)$ diventa come il termine di destra dell'Equazione (1.6). Così, quando in un campione empirico non si osservano risposte mancanti, il BLIM e l'IMBLIM fanno previsioni identiche.

5.2.2 Dati Mancanti *Nonignorable*: il MissBLIM

Si immagini ora che le risposte mancanti del pattern di risposta di uno studente dipendano dal suo stato di conoscenza. Questo è il caso, ad esempio, di uno studente che, non conoscendo la risposta a un problema decide, intenzionalmente, di saltarlo. In una situazione simile assumere che il dato mancante sia casuale è del tutto inappropriato e dunque non sarebbe opportuno applicare l'IMBLIM. In questa sezione si presenta un'estensione del BLIM che incorpora le risposta mancanti di questo tipo, piuttosto che ignorarle, permettendo di modellare i dati mancanti che non sono casuali (MNAR).

Il modello che si propone si basa sulla scomposizione del processo che genera gli esiti osservabili in due fasi:

1. la risposta ad un item q ha una certa probabilità di essere data o non data che dipende sia dall'item che dallo stato di conoscenza dello studente;

2. se la risposta viene data, allora la probabilità che sia corretta dipende dallo stato di conoscenza dello studente in un modo del tutto analogo al BLIM;

La Figura 5.1 mostra una rappresentazione grafica del processo di risposta a due fasi.

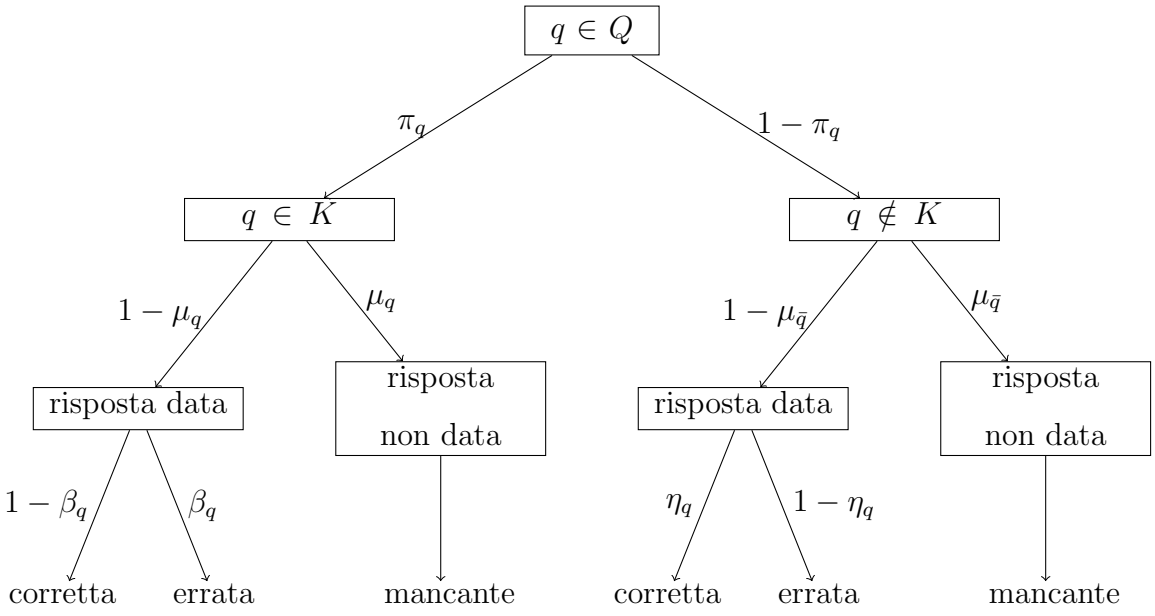

Figura 5.1: Scomposizione del processo di risposta a due fasi che produce gli esiti osservabili nel MissBLIM, per un singolo item $q \in Q$. L'item ha una probabilità π_q di appartenere allo stato di conoscenza K di uno studente (ramo sinistro dell'albero). Se l'item appartiene allo stato di conoscenza ($q \in K$), c'è una probabilità $1 - \mu_q$ che la risposta sia data, e una probabilità μ_q che la risposta sia mancante. Nel caso in cui la risposta venga data dallo studente, c'è una probabilità $1 - \beta_q$ che sia corretta e β_q che sia errata. L'interpretazione del ramo destro dell'albero è del tutto parallela.

Dati un qualunque pattern di risposta $R \subseteq Q$ e un qualunque pattern mancante $M \subseteq Q \setminus R$, se le risposte mancanti non sono indipendenti dagli stati di conoscenza, allora la probabilità congiunta della coppia $\langle R, M\rangle$ è specificata attraverso il modello a classi latenti

$$P(R, M) = \sum_{K \in \mathcal{K}} P(\langle R, M\rangle^* | K) P(M|K) \pi_K, \tag{5.7}$$

dove la probabilità $P(\langle R, M\rangle^*|K)$ è definita dall'Equazione (5.4), e $P(M|K)$ è la probabilità condizionale che uno studente nello stato di conoscenza K fornisca il pattern mancante M.

Per modellare il processo a due fasi è dunque necessario partire da specifiche assunzioni sulla probabilità $P(M|K)$. A questo scopo è conveniente avere a disposizione le seguente definizione:

***Definizione* 6.3**. Sia $\mathbf{M}$ una variabile casuale con realizzazione in 2^Q, che rappresenta il pattern di risposta mancante di uno studente. Dato un qualunque item $q \in Q$, si rappresenta con $q \in \mathbf{M}$ l'evento secondo il quale lo studente non fornisce risposta a q.

Si può ora introdurre l'assunzione principale del MissBLIM.

[**A6.4**] Per ogni item $q \in Q$ e ogni stato di conoscenza $K \in \mathcal{K}$, la probabilità condizionale $P(q \in \mathbf{M}|K)$ di non osservare una risposta a q, dato K è

$$P(q \in \mathbf{M}|K) = \begin{cases} \mu_q & \text{se } q \in K, \\ \mu_{\bar{q}} & \text{se } q \notin K. \end{cases} \tag{5.8}$$

Secondo questa assunzione si possono osservare due tipi di risposte mancanti, a seconda che:

1. lo studente non sia in grado di risolvere correttamente il problema q, e per questo decide di saltarlo (questa situazione è rappresentata da parametro $\mu_{\bar{q}}$);

2. lo studente sa rispondere correttamente all'item q ma non fornisce la risposta (questa situazione è rappresentata dal parametro μ_q).

Nel caso di un test a risposta aperta, ad esempio, il primo tipo di dato mancante può essere osservato perché lo studente non è in grado di risolvere un problema e per questa ragione non fornisce la risposta, mentre il secondo tipo di dato mancante può essere osservato perché lo studente, pur essendo in grado di risolvere il problema, non ha il tempo di svolgerlo.

Oltre a questa assunzione, relativa alla prima fase del processo, si assume anche indipendenza locale tra i dati mancanti agli item, dato lo stato di conoscenza di uno studente (Assunzione 6.5).

[**A6.5**] Le risposte mancanti agli item sono localmente indipendenti dati gli stati di conoscenza, ovvero, per ogni $M \subseteq Q$ e ogni $K \in \mathcal{K}$,

$$P(M|K) = \prod_{q \in M} P(q \in \mathbf{M}|K) \prod_{q \in Q \setminus M} [1 - P(q \in \mathbf{M}|K)].$$

Dalle due assunzioni [**A6.4**], [**A6.5**] segue che la probabilità condizionale di un pattern mancante $M \subseteq Q$, dati gli stati di conoscenza $K \in \mathcal{K}$ è

$$P(M|K) = \left[\prod_{q \in K \cap M} \mu_q\right] \left[\prod_{q \in K \setminus M} (1 - \mu_q)\right] \left[\prod_{q \in M \setminus K} \mu_{\bar{q}}\right] \left[\prod_{q \in Q \setminus (M \cup K)} (1 - \mu_{\bar{q}})\right]. \tag{5.9}$$

Un'ultima nota sulla relazione tra il BLIM e il MissBLIM. Nel caso in cui un campione sia privo di risposte mancanti, allora $\mu_q = \mu_{\bar{q}} = 0$ per tutti gli item $q \in Q$. Sotto questa uguaglianza, dall'Equazione (5.9) si ha $P(M|K) = 0$ se $M \neq \emptyset$. In questo caso l'unica coppia $\langle R, M \rangle$ che ha una probabilità diversa da zero è quella in cui M è un insieme vuoto. Sempre dall'Equazione (5.5), la probabilità $P(\langle R, \emptyset \rangle^* | K)$ si riduce al termine destro dell'Equazione (1.6) relativa al BLIM. Così, nel caso di campioni completi, i due modelli sono identici.

5.3 La trasformazione missing-as-wrong e il bias delle stime dei parametri nel BLIM

Ci sono circostanze in cui le stime dei parametri di un modello sono prive di bias, se si ottengono su un campione completo, cosa che invece non si verifica se il campione contiene dati mancanti. Questo può accadere esclusivamente a causa del metodo usato per trattare i dati mancanti (Schafer & Graham, 2002). In questa sezione si esamina, da un punto di vista formale, la situazione in cui i parametri del BLIM vengano stimati su un campione nel quale le risposte mancanti sono state ricodificate in risposte errate. In tutto il resto del capitolo, questa circostanza verrà chiamata trasformazione *missing-as-wrong*. Tale trasformazione si basa sull'assunzione, del tutto ragionevole, che uno studente salti un problema quando non sa rispondere.

Si supponga ora, che la procedura per stimare i parametri del BLIM su dati completi sia priva di bias. La domanda cruciale è: in che modo la trasformazione *missing-as-wrong* introduce dei bias nelle stime dei parametri del modello? Nelle pagine seguenti si darà risposta a questa domanda, nel caso in cui i dati siano generati a partire dal MissBLIM.

Proposizione 3. *In seguito alla trasformazione missing-as-wrong di un campione, la probabilità β_q^* che la risposta ad un item q sia codificata come* errata*, dato che l'item appartiene allo stato di conoscenza di uno studente è*

$$\beta_q^* = (1 - \mu_q)\beta_q + \mu_q. \tag{5.10}$$

Dall'altro lato, la probabilità η_q^* che la risposta ad un item q sia codificata come corretta*, dato che l'item non appartiene allo stato di conoscenza è*

$$\eta_q^* = (1 - \mu_{\bar{q}})\eta_q. \tag{5.11}$$

Si faccia riferimento all'articolo de Chiusole et al. (in press) per la dimostrazione della Proposizione 3. Quindi, se i parametri di careless error

del BLIM vengono stimati dopo aver applicato ai dati la trasformazione missing-as-wrong, si ottiene un bias pari a

$$\beta_q^* - \beta_q = \mu_q(1 - \beta_q) \geq 0. \tag{5.12}$$

Questo significa che, mediamente, il parametro β_q verrà sovrastimato e che, questa sovrastima sarà proporzionale alla probabilità μ_q. Per quanto riguarda invece il parametro di lucky guess, il bias sarà pari a

$$\eta_q^* - \eta_q = -\mu_{\bar{q}}\eta_q \leq 0 \tag{5.13}$$

In questo caso si osserva una sottostima proporzionale al parametro $\mu_{\bar{q}}$.

Ne consegue che, anche quando è del tutto lecito assumere che una risposta mancante sottenda una risposta errata, alcuni parametri del BLIM saranno comunque affetti da bias. Questa assunzione richiede che una risposta mancante per l'item q si possa verificare, solo nel caso in cui l'item q non appartenga allo stato di conoscenza di uno studente. In questo caso il valore di μ_q sarà pari a zero per tutti gli item $q \in Q$. Ma, nonostante le stime di β_q siano prive di bias, non è così per le stime del parametro η_q, il cui bias dipende dal valore del parametro $\mu_{\bar{q}}$, che in generale sarà maggiore di zero.

5.4 Studio Simulativo

L'obiettivo dello studio simulativo era quello di studiare le conseguenze dell'uso di metodi diversi nel trattamento dei dati mancanti, sulle stime dei parametri dei modelli sviluppati nella KST. A questo scopo, sono stati considerati i seguenti tre casi:

1. caso MCAR, in cui il processo che genera i dati mancanti era del tipo *ignorable*;

2. caso ks-MNAR (dall'inglese knowledge-states-MNAR), in cui il processo alla base delle risposte mancanti era del tipo *nonignorable*, ovvero era

presente una dipendenza dei dati mancanti dagli stati di conoscenza, ma in modo equiprobabile tra gli item;

3. caso iks-MNAR (dall'inglese item-and-knowledge-states-MNAR), in cui il processo alla base delle risposte mancanti era sempre del tipo *nonignorable*, ma la loro probabilità poteva variare attraverso gli item.

In tutti e tre i casi, per prima cosa, sono stati generati un certo numero di campioni in accordo con lo specifico processo, poi, su questi dati, sono stati applicati i tre modelli BLIM, IMBLIM e MissBLIM. Infine, i modelli sono stati confrontati tra di loro rispetto alla bontà di ricostruzione dei parametri veri, da un lato e sull'assegnazione corretta dello stato di conoscenza ai pattern simulati (accuratezza dell'assessment), dall'altro. L'obiettivo consisteva nell'analizzare il comportamento dei tre modelli nei tre diversi casi in cui si potevano osservare i mancanti. Naturalmente è stata data una particolare attenzione ai casi in cui un modello faceva assunzioni errate rispetto al processo che generava le risposte mancanti.

5.4.1 Disegno delle simulazioni e stima dei parametri dei modelli

Nei due casi MCAR e ks-MNAR, sono stati generati secondo il relativo processo $4 \times 100 = 400$ campioni. Ogni campioni era costituito dalla coppia $\langle K, R \rangle$, dove K era uno stato di conoscenza e R il pattern di risposta corrispondente.

Nel caso MCAR, prima sono stati generati 100 campioni completi a partire dal BLIM, ognuno composto da $1,000$ osservazioni. Dopo di che sono state considerate 4 condizioni differenti, in cui una proporzione prestabilita dei dati veniva sostituita con risposte mancanti. Nelle 4 condizioni sono state utilizzate, rispettivamente, le proporzioni .1, .2, .3 e .4. Formalmente, nel

caso MCAR i dati mancanti rispettavano la seguente condizione:

$$P(q \in \mathbf{M} | q \in \mathbf{K}) = P(q \in \mathbf{M} | q \notin \mathbf{K}) = p_{\text{MCAR}}, \tag{5.14}$$

per tutti gli item q appartenenti a Q, dove $p_{\text{MCAR}} > 0$ è costante attraverso gli item.

Nel caso ks-MNAR, è stata applicata la stessa procedura utilizzata per il caso precedente con la differenza che le risposte mancanti potevano osservarsi solo per gli item che non appartenevano allo stato di conoscenza. Formalmente, in questo secondo caso i dati mancanti rispettavano la seguente condizione:

$$P(q \in \mathbf{M} | q \in \mathbf{K}) = 0, \quad P(q \in \mathbf{M} | q \notin \mathbf{K}) = p_{\text{ks-MNAR}}, \tag{5.15}$$

per ogni item q appartenente a Q, dove $p_{\text{ks-MNAR}} > 0$ è costante attraverso gli item. Questa condizione è in linea con le assunzioni del MissBLIM, ma con le restrizioni aggiuntive

- $\mu_q = 0$ per tutti gli item $q \in Q$;
- $\mu_{\bar{q}} = p_{\text{ks-MNAR}}$ è la stessa per tutti gli item.

Da un punto di vista empirico, ciò significa che i dati mancanti sono possibili solamente quando lo studente non sa risolvere l'item. Per facilitare il confronto fra questo caso con gli altri due, la probabilità $p_{\text{ks-MNAR}}$ è stata trasformata in modo tale che le proporzioni dei dati mancanti nell'intero campione fossero uguali a quelle usate nel caso MCAR (.1, .2, .3 e .4). Per questo, le proporzioni p_{RNIM} usate erano pari a .2, .4, .6 or .8, rispettivamente.

Nel terzo e ultimo caso iks-MNAR, sono stati generati $5 \times 100 = 500$ campioni a partire dal MissBLIM, ognuno composto da $1,000$ osservazioni. In ognuna delle 5 condizioni, i parametri μ_q e $\mu_{\bar{q}}$ sono stati generati casualmente a partire da una distribuzione uniforme su un certo intervallo numerico $[a, b]$. La Tabella 5.1 mostra i cinque intervalli usati per generare i due parametri.

Tabella 5.1: Intervalli numerici usati per generare i parametri μ_q e $\mu_{\bar{q}}$ nel caso iks-MNAR. C1, C2, C3, C4 e C5 rappresentano rispettivamente le condizioni dalla 1 alla 5.

Parametro	C1	C2	C3	C4	C5
μ_q	$[.4, .5]$	$[.3, .4]$	$[.2, .3]$	$[.1, .2]$	$(0, .1]$
$\mu_{\bar{q}}$	$(0, .1]$	$[.1, .2]$	$[.2, .3]$	$[.3, .4]$	$[.4, .5]$

Naturalmente, dato che $\mu_q = P(q \in \mathbf{M} | q \in \mathbf{K})$ e che $\mu_{\bar{q}} = P(q \in \mathbf{M} | q \notin \mathbf{K})$, la generazione dei dati mancanti in questo ultimo caso rispettava la condizione:

$$P(q \in \mathbf{M} | q \in \mathbf{K}) \neq P(q \in \mathbf{M} | q \notin \mathbf{K}), \tag{5.16}$$

per ogni item q appartenente a Q. Va sottolineato che, al contrario del secondo caso: (1) le probabilità erano libere di variare attraverso gli item; (2) la tipologia di dato mancante è di tipo misto: le condizioni C1 e C2 hanno una proporzione maggiore di mancanti quando l'item appartiene allo stato ($\mu_q > \mu_{\bar{q}}$); nella condizione C3 la probabilità di osservare un mancante dato che l'item appartiene allo stato o non appartiene è uguale; nelle condizioni C4 e C5 vi è invece una proporzione maggiore di dati mancanti quando i soggetti non sanno svolgere l'item ($\mu_{\bar{q}} > \mu_q$);

Ciascuno dei $400 + 400 + 500 = 1,300$ campioni relativi ai tre casi, sono stati generati fissando la struttura di conoscenza a 500 stati e 25 item. La struttura è stata ottenuta calcolando $\{\emptyset, Q\} \cup \mathcal{P}$, dove $\mathcal{P}$ veniva generato a caso, usando un campionamento senza reinserimento sulla collezione $2^Q \setminus \{\emptyset, Q\}$. Infine, anche i parametri β_q e η_q degli item sono stati generati casualmente usando una distribuzione uniforme, ma nell'intervallo $(0, .1]$. Questi valori, così come le probabilità degli stati di conoscenza erano costanti attraverso tutte le simulazioni.

Generati i dati, i tre modelli (BLIM, IMBLIM, MissBLIM) sono stati applicati ai campioni dei tre casi, e le stime dei parametri sono state ottenute per massima verosimiglianza usando l'algoritmo EM (riferimento). I modelli

sono stati applicati ai dati massimizzando le rispettive funzioni di verosimiglianza: Equazione (1.9) per il BLIM, l'Equazione (5.2) per l'IMBLIM e il MissBLIM, dove $P(R, M)$ era specificata dall'Equazione (5.6) per l'IMBLIM e dall'Equazione (5.8) per il MissBLIM.

Si ricorda che, mentre l'IMBLIM e il MissBLIM si possono stimare su campioni contenenti risposte mancanti, per il BLIM non è possibile. Per applicare ai dati questo modello è stata applicata la trasformazione *missing-as-wrong*, ricodificando le risposte mancanti in risposte errate.

5.4.2 Confronto fra i modelli

In ognuno dei casi descritti nella sezione precedente (MCAR, ks-MNAR e iks-MNAR), sono stati applicati ai dati simulati i tre modelli. BLIM, IMBLIM e MissBLIM sono stati poi posti a confronto sia per quanto riguarda la loro capacità di ricostruire i parametri veri, ovvero quelli utilizzati per generare i dati (*parameter recovery*), sia per quanto riguarda l'accuratezza dell'assessment.

Nel caso della parameter recovery, la media delle stime dei parametri ottenute sulle 100 replicazioni, è stata posta a confronto con i valori veri usati per generare i dati, sia nel caso della careless error che in quello della lucky guess. Questo è stato fatto sia per le 4 condizioni dei casi MCAR e ks-MNAR sia per le 5 condizioni del caso iks-MNAR.

Oltre alla parameter recovery, un test importante per questi modelli riguarda l'accuratezza dell'assessment, ovvero la capacità di individuare correttamente lo stato di conoscenza di uno studente a partire dal suo pattern di risposta. A questo scopo, è stata applicata l'inferenza bayesiana per stimare lo stato di conoscenza di uno studente il cui pattern di risposta osservato è R e il cui pattern di risposta mancante è M. Il metodo si basa sul calcolo della distribuzione di probabilità a posteriori degli stati di conoscenza, data la coppia osservata $\langle R, M \rangle$. Formalmente, dato un qualsiasi stato di

conoscenza $K \in \mathcal{K}$,

$$P(K|R,M) = \frac{P(R,M|K)\pi_K}{\sum_{K' \in \mathcal{K}} P(R,M|K')\pi_{K'}}.$$

La forma specifica di questa equazione dipende da quale modello si sta considerando. L'elemento modale $\hat{K}$ della distribuzione di probabilità a posteriori rappresenta la stima dello stato di conoscenza dello studente.

Una misura dell'accuratezza dell'assessment di ciascun modello, si ottiene dal calcolo della distanza simmetrica

$$d(\hat{K}_m, K) = \left|(\hat{K}_m \setminus K) \cup (K \setminus \hat{K}_m)\right|$$

tra lo stato $\hat{K}_m$ stimato dal modello m, e lo stato di conoscenza *vero* K. Per ogni modello e per ogni coppia $\langle R, M\rangle$ nei $1,300$ campioni simulati, è stata calcolata la distanza $d(\hat{K}_m, K)$. Successivamente, è stata calcolata la media D_{scm} di queste distanze per ogni campione s, condizione c e modello m. Infine, per ogni condizione c e ogni modello m, sono state calcolate la media D_{cm} e la varianza s^2_{cm} sui 100 campioni della condizione. Queste due statistiche sono state utilizzate per confrontare i modelli sull'accuratezza del loro assessment.

5.4.3 Risulati: parameter recovery

Per quanto riguarda il caso MCAR, dove i dati mancanti erano del tipo *ignorable*, le stime per massima verosimiglianza dei parametri d'errore sono state stimate correttamente sia dall'IMBLIM che dal MissBLIM, in tutte e quattro le condizioni dello studio. Il MissBLIM ha inoltre stimato correttamente le proporzioni di dati mancanti, infatti come ci si aspettava $\hat{\mu}_q \simeq \hat{\mu}_{\bar{q}} \simeq p_{MCAR}$. Nel pannello sinistro della Tabella 5.2 sono riportate le stime ottenute per i due parametri μ_q e $\mu_{\bar{q}}$ e, come si può notare, sono molto vicine ai valori utilizzati per simulare i dati (p_{MCAR}). La Figura 5.2 mostra invece i risultati ottenuti nella stima dei parametri di careless error e lucky guess del BLIM, sui dati MCAR. Ciascuno dei quattro diagrammi si riferisce a una delle 4

Tabella 5.2: Stime dei parametri μ_q e $\mu_{\bar{q}}$ del MissBLIM, nei due casi MCAR (pannello di sinistra) e ks-MNAR (pannello di destra).

MCAR			ks-MNAR		
p_{MCAR}	$\hat{\mu}_q$	$\hat{\mu}_{\bar{q}}$	$p_{ks-MNAR}$	$\hat{\mu}_q$	$\hat{\mu}_{\bar{q}}$
.1000	.1007	.0997	.2000	.0000	.2010
.2000	.1999	.1998	.4000	.0000	.4021
.3000	.2997	.3000	.6000	.0000	.6012
.4000	.3999	.4008	.8000	.0000	.7948

proporzioni di dati mancanti utilizzate per simulare i dati. Nella figura, lungo l'asse x sono indicati i valori *veri* dei parametri, mentre sull'asse y sono indicati le stime ottenute dall'applicazione del BLIM. La linea continua è il riferimento per indicare l'uguaglianza $x = y$, mentre le due linee tratteggiate indicano il bias atteso delle stime secondo le formule teoriche (5.10) e (5.11). La linea tratteggiata più in alto si riferisce al bias del parametro di lucky guess, mentre quella più in basso indica il bias del parametro careless error. Le osservazioni principali, sono due:

1. si osserva una sovrastima evidente del parametro careless error, che aumenta all'aumentare di dati mancanti nel data set, in modo proporzionale;

2. le formule teoriche del bias, derivate nel paragrafo 5.3, ricostruiscono perfettamente il bias delle stime: sia i punti neri (riferiti alla careless error), sia i triangoli bianchi (riferiti alle lucky guess) sono allineati sulle relative linee teoriche;

Per quanto riguarda il caso ks-MNAR, dove i dati mancanti erano del tipo *nonignorable*, il MissBLIM è stato l'unico modello a ottenere stime corrette dei parametri d'errore degli item. Il bias massimo ottenuto è stato: .0001, .0001, .0003 e .0126, rispettivamente per le condizioni dalla $C1$ alla $C4$. Va comunque evidenziato che, anche se trascurabile, il bias cresce all'aumentare

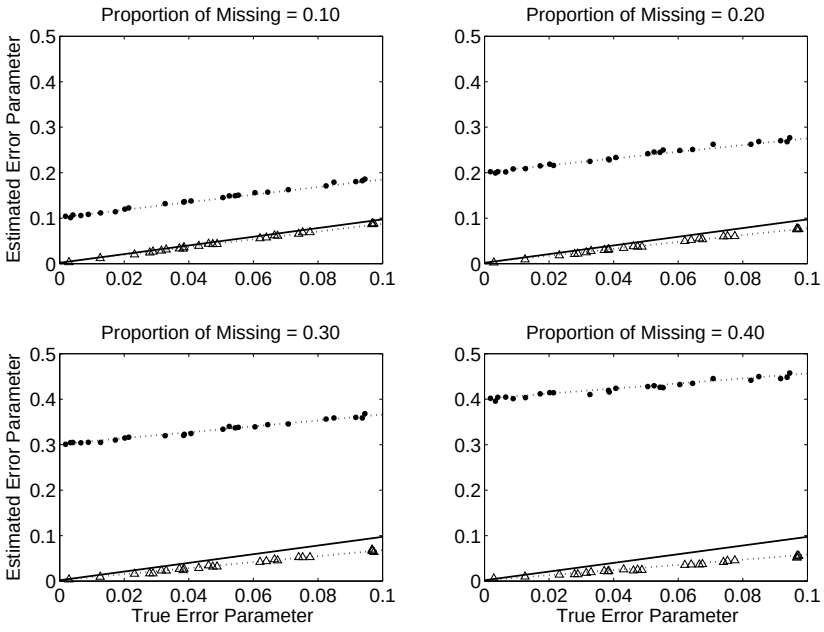

Figura 5.2: Parameter recovery dei parametri β_q e η_q del BLIM, nel caso MCAR. I punti neri rappresentano le stime di β e i triangoli bianchi rappresentano le stime di η. La linea continua è il riferimento per $x = y$ e le due linee tratteggiate sono il riferimento del bias atteso secondo le formule (5.10) e (5.11).

della proporzione di dati mancanti nel campione. Inoltre, anche in questo caso, il MissBLIM stima in modo preciso le proporzioni di dati mancanti (si faccia riferimento alla Tabella 5.2, pannello di destra). Infatti, come ci si aspettava, $\hat{\mu}_q = 0$ e $\hat{\mu}_{\bar{q}} \simeq p_{\text{ks-MNAR}}$, in ciascuna delle quattro condizioni.

Gli altri due modelli ottengono invece stime affette da bias. La Figura 5.3 illustra i risultati ottenuti dall'applicazione del BLIM, e la Figura 5.4 illustra i risultati ottenuti dall'IMBLIM. Le due figure si leggono in modo del tutto analogo alla figura precedente. Dalla Figura 5.3 risulta chiaro che, quando il processo che genera i dati mancanti è del tipo ks-MNAR, applicando il BLIM ai dati con la trasformazione *missing-as-wrong* si ottengono delle lucky guess sottostimate (triangoli bianchi in figura). I parametri di careless error (punti neri in figura) vengono invece ricostruiti correttamente, infatti si trovano tutti

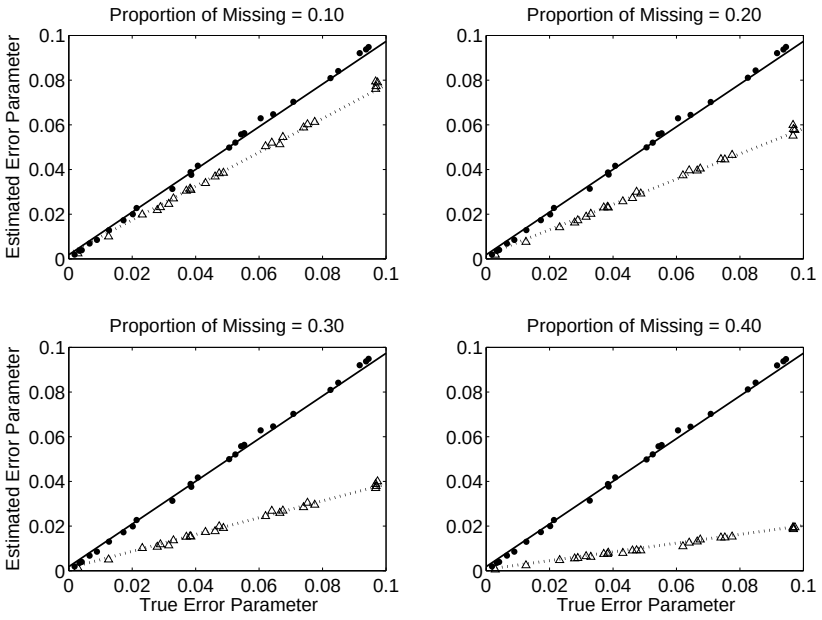

Figura 5.3: Parameter recovery del BLIM nel caso ks-MNAR. La figura si legge in modo del tutto analogo a quella precedente. Si ricorda che i punti neri rappresentano le stime di β e i triangoli bianchi rappresentano le stime di η. La linea continua e le due linee tratteggiate sono il riferimento del bias atteso secondo le formule (5.10) e (5.11).

allineati sulla linea continua nera, che indica la corrispondenza tra stime e valori *veri*. Anche questi risultati sono in accordo con i risultati teorici discussi alla fine del paragrafo 5.3.

Anche l'IMBLIM (Figure 5.4) ottiene stime affette da bias, in particolare quando la proporzione di dati mancanti nel campione supera il valore .30. In questi casi i parametri di lucky guess (triangoli bianchi) vengono sovrastimati, mentre i parametri di careless error (punti neri) vengono sottostimati. Nell'ultimo caso (iks-MNAR), dove i dati mancanti erano sempre del tipo *nonignorable*, ma la loro probabilità variava liberamente attraverso gli item, il MissBLIM ottiene stime prive di bias, l'IMBLIM ottiene stime con bias trascurabile, il BLIM è l'unico per il quale si osservano bias elevati (Figura 5.5). Anche questa figura si legge in modo del tutto analogo alle precedenti,

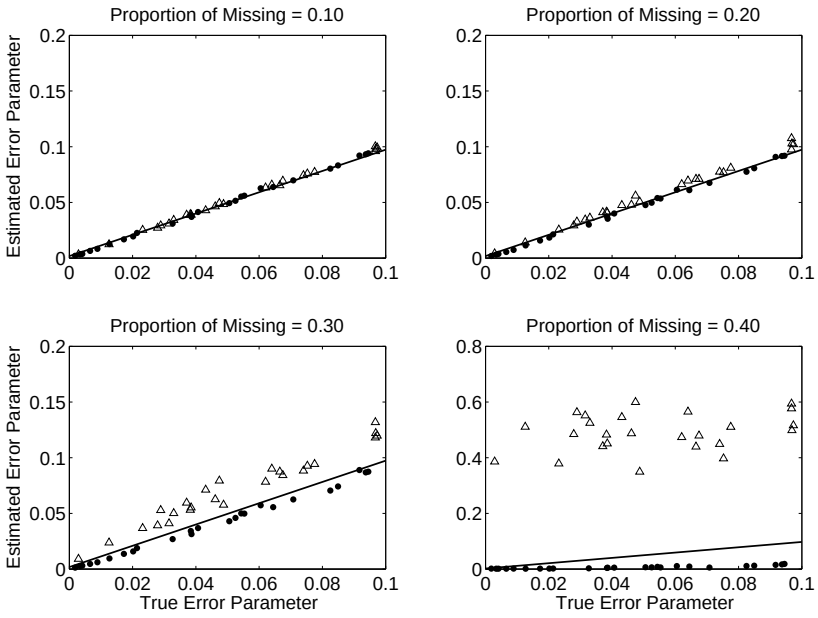

Figura 5.4: Parameter recovery dei parametri dell'IMBLIM nel caso ks-MNAR. La figura si legge in modo del tutto analogo a quella precedente. Si ricorda che i punti neri rappresentano le stime di β e i triangoli bianchi rappresentano le stime di η.

con l'unica differenza che il bias teorico è ora indicato con dei cerchietti neri, anziché con le linee tratteggiate. Questo è dovuto al fatto che, nel caso in esame, le proporzioni di mancanti μ_q e $\mu_{\bar{q}}$ variavano da item a item, rendendo impossibile interpolare il bias con una linea continua. Dalla figura è evidente che i parametri di lucky guess (triangoli bianchi) vengono sottostimati, mentre quelli di careless error (punti neri) sono sovrastimati. Per quanto riguarda le formule teoriche del bias, anche in questo caso riproducono in modo corretto i valori ottenuti per le stime. Un risultato interessante da evidenziare, è che se le sovrastime del parametro di lucky guess diminuiscono dalla condizione C5 alla C1, le sottostime del parametro di careless error aumentano. Un simile risultato sembra la conseguenza del fatto che i dati mancanti, in questo caso, sono stati generati con due diversi processi, infatti sia μ_q che $\mu_{\bar{q}}$ erano maggiori di zero, ma con una predominanza di un processo sull'altro diversa a seconda della condizione. Dunque, maggiore è

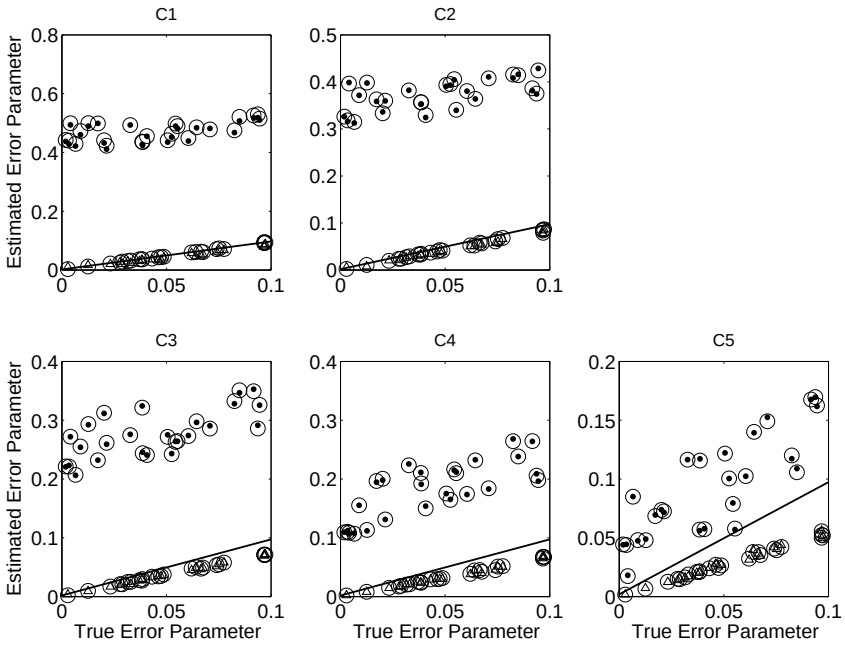

Figura 5.5: Parameter recovery dei parametri del BLIM nel caso iks-MNAR. La figura si legge in modo del tutto analogo alle precedenti. Si ricorda che i punti neri rappresentano le stime di β e i triangoli bianchi rappresentano le stime di η. I cerchietti neri sono il riferimento per il bias atteso secondo le formule (5.10) e (5.11).

la probabilità di osservare dati mancanti dato che l'item è contenuto nello stato ($\mu_q > \mu_{\bar{q}}$), maggiore sarà la sovrastima del parametro di lucky guess, mentre maggiore è la probabilità di osservare dati mancanti dato che l'item non appartiene allo stato ($\mu_{\bar{q}} > \mu_q$), maggiore è la sottostima del parametro di careless error.

Infine, la Figura 5.6 mostra le stime ottenute dall'IMBLIM. Come si può notare, a seconda della condizione, il bias è assente (C3) o del tutto trascurabile (C1, C2, C4 e C5).

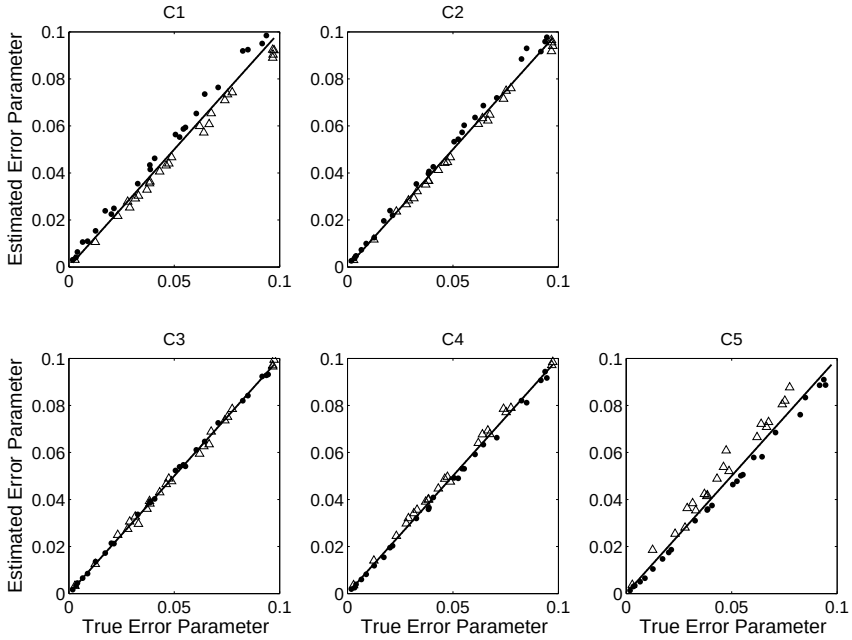

Figura 5.6: Parameter recovery dei parametri dell'IMBLIM nel caso iks-MNAR. La figura si legge in modo del tutto analogo a quella precedente. Si ricorda che i punti neri rappresentano le stime di β e i triangoli bianchi rappresentano le stime di η.

5.4.4 Risultati: accuratezza dell'assessment

I tre modelli sono stati confrontati anche in merito alla loro capacità di ricostruire correttamente lo stato di conoscenza che ha generato il pattern di risposta (*accuratezza dell'assessment*). La Figura 5.7 mostra i risultati della media della distanza D_{cm}, calcolata sui 100 campioni di ogni condizione.

Nella Figura, ciascun diagramma si riferisce a un caso specifico dello studio: MCAR, ks-MNAR e iks-MNAR, rispettivamente. Le barre nere rappresentano l'IMBLIM, le barre grigie il MissBLIM, e le barre bianche il BLIM. Le differenti condizioni delle simulazioni si trovano sull'asse delle x, mentre la distanza simmetrica media D_{cm} è collocata lungo l'asse y. I risultati degni di nota sono elencati di seguito: (1) quando il processo che genera i dati

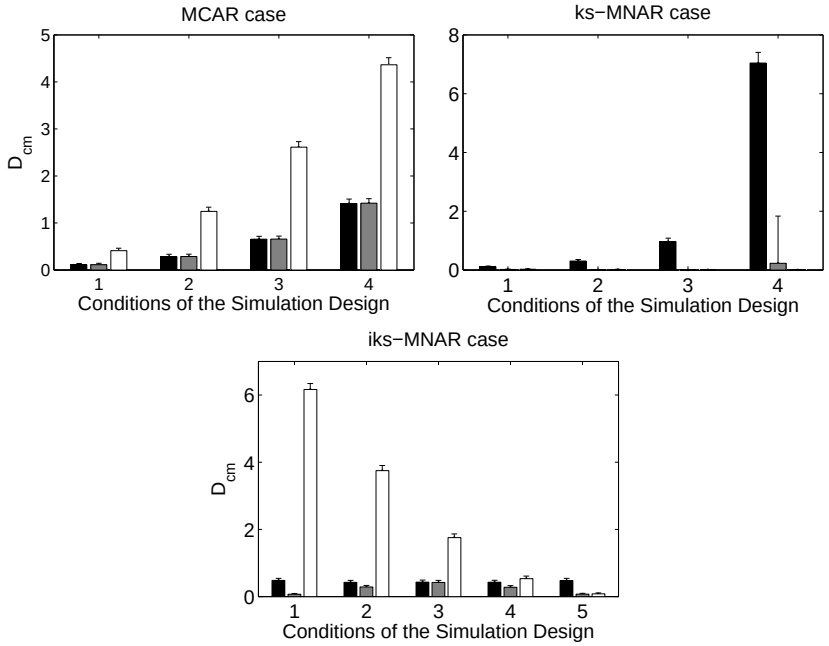

Figura 5.7: Distanze simmetriche tra lo stato di conoscenza modale e quello *vero* ottenute per i tre modelli. Ogni diagramma illustra i risultati ottenuti nei casi MCAR, ks-MNAR e iks-MNAR. Le barre nere rappresentano l'IMBLIM, le barre grigie il MissBLIM, e le barre bianche il BLIM.

mancanti è del tipo MCAR (primo diagramma), il MissBLIM e l'IMBLIM ottengono lo stesso livello di accuratezza: anche quando la proporzione di dati mancanti è piuttosto elevata (40%), la distanza simmetrica media, per entrambi i modelli, è inferiore a 2 item su un totale di 25. L'accuratezza del BLIM invece, peggiora con l'aumentare della proporzione di dati mancanti, fino ad arrivare a quasi 4 item di differenza tra lo stato di conoscenza ricostruito e quello vero; (2) quando il processo che genera i dati mancanti è del tipo ks-MNAR (secondo diagramma), l'accuratezza del MissBLIM è simile a quella del BLIM, ed entrambe sono decisamente buone. In questo caso è l'IMBLIM il modello che ottiene la prestazione peggiore in termini di accuratezza dell'assessment e peggiora all'aumentare della presenza di dati mancanti nel campione; (3) quando il processo che genera i dati è del tipo iks-MNAR,

l'accuratezza dell'assessment del MissBLIM e quella dell'IMBLIM sono equivalenti. Nel BLIM invece si osserva un comportamento diverso a seconda del tipo di dato mancante. Sembra infatti che, quando la proporzione di mancanti MNAR è maggiore a quella MCAR (condizioni C4 e C5), il modello ottenga una prestazione equiparabile a quelle degli altri due modelli. Quando invece la proporzione di mancanti MNAR è minore a quella MCAR (C1, C2 e C3), le sue prestazioni peggiorano, fino ad arrivare a 6 item di differenza tra stato vero e stato ricostruito.

Complessivamente sia i risultati della parameter recovery che i risultati dell'accuratezza dell'assessment, suggeriscono che:

- il BLIM, applicato ai dati con la trasformazione *missing-as-wrong*, è sempre inappropriato, anche quando i dati mancanti sottendono di fatto una risposta errata;

- l'IMBLIM, come ci si aspettava, è particolarmente adatto alla tipologia di dati mancanti MCAR, ma sembra reggere anche nel caso in cui i dati mancanti sono di tipo misto, ovvero sia di tipo *ignorable* che *nonignorable* (iks-MNAR). E' invece inappropriato nel caso in cui dati mancanti dipendano completamente dallo stato di conoscenza (ks-MNAR);

- il MissBLIM sembra essere un modello particolarmente flessibile, adatto alla modellazione di dati mancanti, indipendente dalla tipologia del processo che li ha generati.

5.5 Applicazione Empirica

L'IMBLIM e il MissBLIM sono stati applicati anche a dati empirici, con i seguenti obiettivi:

1. testare la loro bontà di adattamento ai dati in contesti reali;

2. confrontare i due modelli per stabile quale processo tra MCAR e MNAR è più probabile abbia generato i dati mancanti presenti nel campione.

5.5.1 Partecipanti e metodi

Il campione consisteva di 176 studenti di psicologia dell'Università di Padova, che stavano frequentando il corso di Psicometria nell'anno accademico 2012-2013. A ciascun studente è stata somministrata, attraverso una forma computerizzata, una collezione di 18 problemi sulla statistica descrittiva. I problemi erano tutti a risposta aperta, di tipo numerico. La somministrazione si caratterizzava per i seguenti aspetti: (a) i 18 problemi venivano presentati, uno per volta, sullo schermo di un computer e l'ordine di presentazione variava da studente a studente; (b) ogni studente poteva decidere di rispondere o saltare il problema, dopo di ché non era possibile tornare ai problemi precedenti; (c) c'era un limite di tempo di 2 ore per completare i problemi, al termine delle quali la procedura di assessment terminava.

In questo modo, nel campione erano presenti due tipologie diverse di dati mancanti. Un primo tipo si osservava perché lo studente saltava intenzionalmente l'item. Un secondo tipo si osservava perché il problema non veniva presentato allo studente per ragioni di tempo. A questo proposito era lecito supporre che la prima tipologia di dato mancante si manifestasse perché lo studente non era in grado di rispondere, mentre la seconda non dipendeva in alcun modo dallo stato di conoscenza dello studente.

Le risposte degli studenti sono state codificate come corrette (1), sbagliate (0) o mancanti (-1), e sono state usate per costruire la matrice dei dati. Nel caso del BLIM, le risposte mancanti sono state ricodificate come errate, mentre l'IMBLIM e il MissBLIM sono stati applicati alla matrice originale dei dati. A questo punto i tre modelli sono stati applicati ai rispettivi campioni e i loro parametri sono stati stimati per massima verosimiglianza, usando l'algoritmo EM.

La bontà di adattamento di ciascun modello è stata testata utilizzando il Chi-quadro di Pearson, definito come

$$\chi^2(\theta; \mathcal{D}, N) = \sum_{\langle R,M\rangle \in \mathcal{O}} \frac{(F(R,M) - NP_\theta(R,M))^2}{NP_\theta(R,M)},$$

dove:

- θ è il vettore delle stime dei parametri;
- $\mathcal{D}$ è il campione;
- $F(R,M)$ è la frequenza osservata della coppia $\langle R,M\rangle$;
- N è la numerosità campionaria;
- $P_\theta(R,M)$ è la probabilità marginale di $\langle R,M\rangle$ calcolata in accordo con il modello in esame (BLIM, IMBLIM o MissBLIM).

Per calcolare il p-value del Chi-quadro è stata utilizzata una procedura di bootstrap parametrico (Efron, 1979), operazione necessaria nel caso della presenza di una matrice sparsa di dati. Infatti, nella presente applicazione si avevano 173 pattern di risposta osservati su 2^{18} pattern teorici possibili. Si veda la Sezione 1.3.1, per una illustrazione della procedura bootstrap.

Il confronto dei tre modelli è stato fatto utilizzando gli indici di selezione dei modelli AIC e BIC, dal momento che non era possibile utilizzare il Chi-quadro. Infatti questa statistica è utile solo nel caso di modelli annidati e non può in alcun modo essere impiegata per il confronto fra modelli che hanno diversa complessità (si veda ad esempio, Zucchini, 2000). Gli indici AIC e BIC invece, penalizzano la verosimiglianza del modello per la sua complessità. Così, sono stati usati per stabilire quale modello tra il MissBLIM e l'IMBLIM fosse quello che meglio si approssimava ai dati.

Va sottolineato che il confronto di questi due modelli con il BLIM è del tutto fuorviante, dal momento che sia i risultati teorici che quelli delle simulazioni hanno evidenziato che applicare questo modello a un campione

trasformato, è erroneo. Per questa ragione nell'applicazione empirica si è deciso di non applicare il BLIM.

5.5.2 Risultati

Sia l'IMBLIM che il MissBLIM hanno ottenuto una bontà di adattamento piuttosto buona, raggiungendo un p-value bootstrap pari a .36 e .14, rispettivamente. Per quanto concerne i criteri di selezione dei modelli, la Tabella 5.3 riporta i risultati del confronto. Sia l'indice AIC che il BIC selezionano

Tabella 5.3: Confronto tra i modelli IMBLIM e MissBLIM stimati su dati empirici.

Modello	AIC	BIC
IMBLIM	5.28×10^5	1.36×10^6
MissBLIM	1.08×10^4	1.25×10^4

il MissBLIM come modello che meglio si approssima ai dati, suggerendo che il processo che ha generato i dati mancanti sia MNAR, e che per questo non possono essere ignorati. Questa considerazione è supportata anche dal confronto tra la media delle stime del parametro $\hat{\mu}_q$ e quella di $\hat{\mu}_{\bar{q}}$ del MissBLIM (si faccia riferimento alle ultime due colonne della Tabella 5.4). Si può infatti notare che la media di $\hat{\mu}_q$ è .08, mentre quella di $\hat{\mu}_{\bar{q}}$ è .32. Ciò sta a indicare che, mediamente, la probabilità di osservare un dato mancante dato che lo studente non sa rispondere ad un item è maggiore. Apparentemente, questo risultato sembra in favore dell'assunzione *missing-as-wrong*, ma le stime ottenute per il parametro μ_q rivelano che tale assunzione può essere sì vera per alcuni item (per i quali $\hat{\mu}_q = 0$), ma non lo è per altri (per i quali $\hat{\mu}_q > 0$).

Infine i modelli IMBLIM e MissBLIM sono stati confrontati sulle stime dei parametri di careless error e di lucky guess. La media dei parametri di careless error è .22 per l'IMBLIM, e .19 per il MissBLIM, mentre la media del parametri di lucky guess è .10 per l'IMBLIM e .08 per il MissBLIM. Me-

Tabella 5.4: Stime dei parametri dei modelli IMBLIM e MissBLIM ottenute nell'applicazione empirica

	IMBLIM		MissBLIM			
Item	$\hat{\beta}_q$	$\hat{\eta}_q$	$\hat{\beta}_q$	$\hat{\eta}_q$	$\hat{\mu}_{\bar{q}}$	$\hat{\mu}_q$
1	.00	.18	.00	.10	.66	.00
2	.18	.00	.14	.00	.26	.08
3	.14	.17	.17	.17	.18	.00
4	.07	.18	.09	.12	.21	.00
5	.18	.00	.14	.00	.45	.05
6	.34	.29	.30	.24	.05	.02
7	.11	.05	.15	.04	.78	.02
8	.55	.00	.43	.00	.51	.18
9	.15	.00	.10	.00	.16	.11
10	.35	.00	.29	.09	.28	.24
11	.38	.13	.24	.14	.45	.23
12	.00	.40	.00	.26	.29	.00
13	.00	.19	.02	.18	.44	.00
14	.59	.00	.25	.00	.00	.48
15	.31	.13	.40	.04	.16	.00
16	.20	.00	.22	.00	.21	.00
17	.22	.00	.21	.00	.17	.06
18	.12	.09	.29	.03	.44	.02
Media	.22	.10	.19	.08	.32	.08

diamente il MissBLIM ottiene stime dei parametri più basse, e considerando che un basso errore dei parametri va a supporto di una buona fit del modello (si veda, per esempio, Stefanutti & Robusto, 2009), anche questo risultato è a favore del MissBLIM.

5.6 Discussione

In questo capitolo è stato presentato uno studio sul trattamento dei dati mancanti nell'ambito della KST, dal momento che l'argomento non era mai stato trattato all'interno della teoria. In particolare, sono state derivate due

estensioni del BLIM al caso di dati mancanti: la prima, chiamata IMBLIM, assume che i dati mancanti siano di tipo MCAR, mentre la seconda, chiamata MissBLIM è adatta al caso in cui vi sono dipendenze tra dati mancanti e stati di conoscenza, ovvero quando il processo è MNAR. In quest'ultimo modello, i dati mancanti si modellano attraverso due parametri:

- μ_q, che rappresenta la probabilità condizionale di osservare una risposta mancante per un item q, dato che lo studente sa rispondere correttamente a q;
- $\mu_{\bar{q}}$, che rappresenta la probabilità condizionale di osservare una risposta mancante, dato che lo studente non sa rispondere a q.

I due modelli, insieme al BLIM, sono stati messi alla prova sia in uno studio simulativo, sia in un applicazione empirica. Nello studio simulativo sono stati generati 3 tipologie di campioni nei quali il processo che generava i dati mancanti poteva essere di tre tipi: MCAR, MNAR, con equiprobabilità di osservare una risposta mancante attraverso gli item e MNAR, con probabilità dei dati mancanti libera di variare da item a item. In quest'ultimo caso, si consideravano inoltre proporzioni miste di dati mancanti casuali e non casuali. I risultati delle simulazioni hanno evidenziato che:

- quando il processo è MCAR, sia l'IMBLIM che il MissBLIM sono adeguati alla modellazione dei dati mancanti. Nel caso del MissBLIM si osserva inoltre che i parametri μ_q e $\mu_{\bar{q}}$ sono equivalenti tra loro;
- quando il processo è MNAR, l'unico modello adatto alla modellazione dei mancanti è il MissBLIM, nel quale le stime dei parametri legate ai dati mancanti sono:
  - $\mu_q = 0$
  - $\mu_{\bar{q}} = P(M_q)$, dove M_q indica la proporzione di dati mancanti osservati nel campione.

- quando, per rendere possibile l'applicazione del BLIM, si usa la trasformazione *missing-as-wrong* (trasformazione delle risposte mancanti in risposte errate), i parametri d'errore di questo modello sono affetti da bias, indipendentemente dal processo che li ha generati.

Per quanto riguarda l'applicazione empirica, l'IMBLIM e il MissBLIM sono stati applicati a un campione che conteneva dati mancanti. L'obiettivo era testare se, attraverso il confronto dei due modelli, era possibile individuare quale fosse il processo che aveva generato i dati mancanti. I risultati evidenziano che i dati mancanti non erano indipendenti dallo stato di conoscenza degli studenti, erano quindi di tipo MNAR.

I risultati teorici ed empirici descritti in questo capitolo mostrano quindi chiari vantaggi nell'utilizzo del MissBLIM per la modellazione dei dati mancanti, ancor più se si pensa all'ambito in cui questi modelli vengono utilizzati, ovvero a quello educativo/scolastico. E' stato dimostrato come trasformare i dati mancanti in risposte errate, è del tutto inappropriato, perché porta a una errata valutazione dello stato di conoscenza degli studenti. Da questo punto di vista emerge ancor più chiaramente l'importanza di avere a disposizione modelli flessibili, come il MissBLIM, che si adattano alle più svariate situazioni in cui i dati mancanti si possono osservare.

Capitolo 6

Modellare le Dipendenze tra Abilità in Strutture di Competenza Probabilistiche

Nel Capitolo 2 sono stati introdotti i concetti teorici alla base della *competence-based* KST (CbKST), un'estensione dell'approccio comportamentale della teoria, che incorpora assunzioni di tipo psicologico sulle abilità sottostanti alla risoluzione di un insieme di problemi. Secondo questo approccio, ciascuna struttura di conoscenza $\mathcal{K}$, costruita su un insieme di problemi q appartenenti al dominio di conoscenza Q, viene messa in relazione con una struttura di competenza $\mathcal{C}$, costruita su un insieme di abilità S necessarie alla risoluzione dei problemi $q \in Q$. La struttura di competenza $\mathcal{C}$ diventa così un modello deterministico della relazione tra le abilità.

Tipicamente un ricercatore specifica un *modello deterministico* che necessita poi di una validazione empirica. Tale validazione è resa possibile da una controparte probabilistica ovvero da un *modello probabilistico*. La validazione empirica delle strutture di conoscenza, è possibile, ad esempio, grazie al modello probabilistico BLIM (Capitolo 2, Sezione 1.3). Per quanto riguarda invece la validazione empirica delle strutture di competenza, al momento

non esistono modelli probabilistici nella CbKST, che la rendono possibile. L'obiettivo della ricerca che si presenta in questo capitolo è stato quello di sviluppare e testare empiricamente un modello probabilistico per strutture di competenza. Questo capitolo riassume i risultati teorici ed empirici riportati nell'articolo (de Chiusole & Stefanutti, 2013).

6.1 Introduzione

Nella KST sono stati proposti svariati modelli deterministici con l'obiettivo di individuare le abilità possedute dagli studenti. In questi modelli la relazione tra problemi e abilità è stata studiata attraverso il concetto di skill multi-map (Doignon & Falmagne, 1999; Düntsch & Gediga, 1996; Doignon, 1994), si veda la Sezione 2.2 per una definizione formale. Questo approccio è stato poi esteso da Korossy (si veda ad esempio Korossy, 1999, 1997, 1993), che sviluppò un quadro teorico formale nel quale il livello di *performance* si differenziava dal livello *delle competenze*. Il primo si caratterizza da un insieme non vuoto Q di problemi e una struttura di conoscenza $\mathcal{K}$ su Q. Il secondo si caratterizza da un insieme non vuoto S di abilità e una collezione $\mathcal{C}$ di sottoinsiemi di S, contenenti almeno l'insieme vuoto e S. La collezione $\mathcal{C}$ è chiamata struttura di competenza e ciascuno dei suoi sottoinsiemi è chiamato stato di competenza.

Il livello performance e quello delle competenze sono collegati l'uno all'altro da due funzioni:

- la funzione $k : Q \to 2^{\mathcal{C}}$ mappa ogni problema appartenente a Q in qualche collezione non vuota di stati di competenza C;
- la funzione $p : \mathcal{C} \to 2^{Q}$ mappa ogni stato di competenza $C \in \mathcal{C}$ in qualche sottoinsieme $K \subseteq Q$ di problemi.

In particolare, va evidenziato che attraverso una struttura di competenza vengono specificate relazioni di indipendenza/dipendenza tra le abilità (così

come accade in una struttura di conoscenza tra gli item). Il problema quindi è di tradurre questa rappresentazione deterministica in una rappresentazione probabilistica, e che queste due rappresentazioni siano in accordo fra loro, circa la relazione di indipendenza/dipendenza tra abilità.

La Sezione 6.2 presenta un'estensione del BLIM al caso delle strutture di competenza. Viene poi presentata un'applicazione empirica del modello (Sezione 6.3) a un campione di bambini italiani della classe terza della scuola primaria, ai quali è stato somministrato un insieme di problemi con la sottrazione.

6.2 Un Modello per Strutture di Competenza

In questa sezione si presenta un modello probabilistico per strutture di competenza, dove le indipendenze/dipendenze teoriche e probabilistiche tra le abilità sono in accordo le une con le altre.

Le analisi che si presentano sono ristrette a strutture di competenza chiamate *spazi di competenza well-graded* (WGCS - Sezione 1.2.2) che, si ricorda, sono una classe particolare di strutture che (nel caso in cui l'insieme S sia finito) soddisfano le seguenti condizioni:

1. per ogni insieme non vuoto $C \in \mathcal{C}$ esiste un'abilità $s \in S$ tale che $C \setminus \{s\} \in \mathcal{C}$;

2. $C \cup C' \in \mathcal{C}$ per ogni C, $C' \in \mathcal{C}$.

Una proprietà molto importante di una WGCS $\mathcal{C}$ è che, data una qualsiasi abilità $s \in S$, esistono sempre almeno due stati di competenza C e D in $\mathcal{C}$ che differiscono esattamente per quell'abilità.

Detto questo, si consideri ora un modello, simile al BLIM, che assume una distribuzione di probabilità $P(C)$ sugli stati di competenza C appartenenti

alla struttura $\mathcal{C}$. Allora la probabilità di uno stato di conoscenza K si ottiene da:

$$P(K) = \sum_{C \in \mathcal{C}_K} P(C), \tag{6.1}$$

dove $\mathcal{C}_K = \{C \in \mathcal{C} : p(C) = K\}$ e p è la funzione problema. In altre parole la probabilità di uno stato di conoscenza è la somma delle probabilità di tutti gli stati di competenza C tali che $p(C) = K$. Ci si riferirà a questa versione del BLIM come *skill based* BLIM (sbBLIM). Va evidenziato che, questo modello, non pone nessun tipo di vincolo circa la relazione tra le abilità $s \in S \setminus C$. Quello che si può verificare dunque è che le relazioni specificate a livello deterministico non siano di fatto incontrate a livello probabilistico.

Nella pratica è naturalmente possibile considerare un insieme di abilità indipendenti tra loro, ma è altrettanto vero che, tra le abilità necessarie alla risoluzione di un insieme di problemi appartenenti allo stesso dominio di conoscenza, ce ne siano alcune che si trovino in una relazione di dipendenza. Per fare un esempio, quando un bambino impara a svolgere le sottrazioni in colonna con più cifre decimali, prima di svolgere le operazione che richiedono l'abilità di manipolare due o più prestiti, è necessario che impari a svolgere quelle che ne richiedono uno solo. In questo caso dunque, l'abilità che chiamiamo *manipolare due o più prestiti* ha come prerequisito l'abilità *manipolare un prestito*. Il modello sbBLIM sopra definito, necessita di alcune restrizioni.

Per ogni stato di competenza non vuoto $C \in \mathcal{C}$ e per ogni abilità s non appartenente a C ($s \in S \setminus C$), si denoti con C_s l'unione di C con $\{s\}$ ($C_s = C \cup \{s\}$) e si consideri il rapporto incrociato (o odds):

$$\theta(C, s) = \frac{P(C_s)}{P(C)}.$$

Con l'obiettivo di ottenere una rappresentazione probabilistica delle dipendenze, che sia in accordo con quella deterministica, è necessario che l'odds $\theta(C, s)$ soddisfi alcuni vincoli. Per fare un esempio, sia $S = \{a, b, c, d\}$ un'insieme di 4 abilità e si consideri la struttura di competenza WGCS

$\mathcal{C} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}, S\}$, il cui grafo è illustrato in Figura 6.1. In questa struttura, se si sa che uno studente non possiede

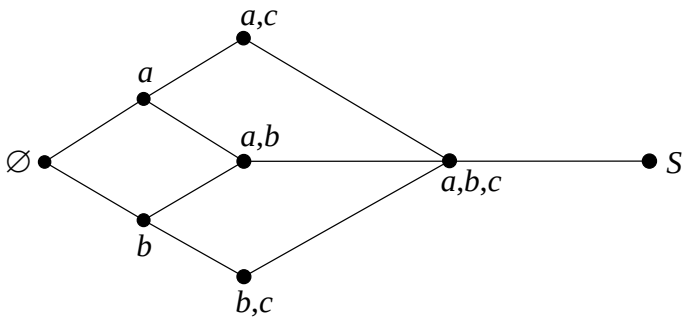

Figura 6.1: Grafo della struttura di competenza *well-graded* $\mathcal{C}$, sull'insieme $S = \{a, b, c, d\}$.

l'abilità c allora, condizionalmente a questo evento, le abilità a e b sono indipendenti. Questo è vero perché, sapendo che c non appartiene allo stato di competenza dello studente, gli unici stati osservabili sono $\emptyset, \{a\}, \{b\}, \{a, b\}$. In altre parole, ogni combinazione rispetto alle altre due abilità è possibile.

Il problema, dunque, è capire come la relazione tra le tre abilità a, b e c si possa rappresentare da un punto di vista probabilistico. L'associazione condizionale tra le abilità a e b, dato che l'abilità c non è posseduta dallo studente, si misura dal rapporto incrociato condizionale

$$\rho_{ab|\bar{c}} = \frac{P(\emptyset)P(\{a, b\})}{P(\{a\})P(\{b\})},$$

che può essere riscritto in una delle due forme equivalenti

$$\rho_{ab|\bar{c}} = \frac{\theta(\{b\}, a)}{\theta(\emptyset, a)}, \quad \rho_{ab|\bar{c}} = \frac{\theta(\{a\}, b)}{\theta(\emptyset, b)}.$$

Allora, l'indipendenza stocastica condizionale tra a e b, dato che c non appartiene allo stato, è vera se

$$\rho_{ab|\bar{c}} = 1,$$

cioè, quando le seguenti due equazioni sono soddisfatte:

$$\begin{cases} \theta(\emptyset, a) = \theta(\{b\}, a), \\ \theta(\emptyset, b) = \theta(\{a\}, b). \end{cases}$$

Questo sistema a due equazioni può essere visto come la rappresentazione probabilistica dell'indipendenza condizionale algebrica tra a e b.

La struttura di competenza $\mathcal{C}$ contiene altre due indipendenze condizionali. Per quanto riguarda la prima, si supponga di sapere che l'abilità a appartiene allo stato di competenza dello studente, mentre l'abilità d no. In questo caso gli unici stati che soddisfano questa condizione sono $\{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}$, pertanto il relativo rapporto condizionale sarà

$$\rho_{bc|a\bar{d}} = \frac{\theta(\{a, c\}, b)}{\theta(\{a\}, b)} = \frac{\theta(\{a, b\}, c)}{\theta(\{a\}, c)}$$

L'indipendenza probabilistica condizionale tra b e c, dato che a appartiene allo stato mentre d no, richiede che le seguenti due equazioni siano vere:

$$\begin{cases} \theta(\{a\}, c) = \theta(\{a, b\}, c); \\ \theta(\{a\}, b) = \theta(\{a, c\}, b). \end{cases}$$

L'ultimo caso di indipendenza condizionale nella struttura di competenza $\mathcal{C}$, si nota se si considera il caso in cui sia noto che l'abilità b appartiene allo stato dello studente, mentre d no. Gli unici stati possibili sono dunque $\{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}$, e i relativi rapporti incrociati saranno

$$\rho_{ac|b\bar{d}} = \frac{\theta(\{b, c\}, a)}{\theta(\{b\}, a)} = \frac{\theta(\{a, b\}, c)}{\theta(\{b\}, c)}.$$

La condizione $\rho_{ac|b\bar{d}} = 1$ è vera se le seguenti due equazioni sono vere

- $\theta(\{b\}, c) = \theta(\{a, b\}, c)$;
- $\theta(\{b\}, a) = \theta(\{b, c\}, a)$.

Ora, se si considerano contemporaneamente tutti e sei i vincoli e le relative equazioni, si ottiene il seguente sistema di equazioni

$$\begin{cases} \theta(\emptyset, a) = \theta(\{b\}, a) = \theta(\{b, c\}, a), \\ \theta(\emptyset, b) = \theta(\{a\}, b) = \theta(\{a, c\}, b), \\ \theta(\{a\}, c) = \theta(\{b\}, c) = \theta(\{a, b\}, c). \end{cases} \tag{6.2}$$

Si può notare che nel sistema in (6.2) il rapporto incrociato $\theta(C, s)$ dipende solamente dalle abilità s, mentre rimane costante attraverso tutti gli stati $C \in \mathcal{C}$. Si nota inoltre che la struttura $\mathcal{C}$ contiene esattamente dieci coppie (C, s), con $C, C_s \in \mathcal{C}$. Parallelamente, ci sono dieci rapporti incrociati $\theta(C, s)$ e nove di essi sono vincolati dalle equazioni del sistema in 6.2.

L'esempio serve ad illustrare come, dati due stati di competenza $C, D \in \mathcal{C}$ tali per cui sia C_s che D_s appartengono a $\mathcal{C}$ per qualche abilità $s \in S$, allora l'equazione

$$\theta(C, s) = \theta(D, s) = \theta(s)$$

deve essere vera. Questo vincolo è soddisfatto da una distribuzione di probabilità sugli stati di competenza del tipo

$$P(\mathbf{C} = C) = \frac{\prod_{s \in C} \theta(s)}{\sum_{D \in \mathcal{C}} \prod_{s \in D} \theta(s)}, \tag{6.3}$$

per tutti gli stati di competenza $C \in \mathcal{C}$. Ci si riferirà a questo modello come al *Dependence* BLIM, abbreviato con DBLIM.

6.3 Applicazione Empirica

6.3.1 Metodi

Con l'obiettivo di testare un modello che cattura la dipendenza/indipendenza tra un insieme di abilità, il primo passo è stato quello di considerare un modello di abilità in cui questa distinzione fosse ben chiara. E' stato scelto il dominio di conoscenza relativo all'operazione aritmetica della sottrazione. Apprendere a svolgere le sottrazioni in colonna, richiede infatti una serie di abilità procedurali, tra cui:

- *manipolare decine e centinaia* (f);
- *manipolare le migliaia* (g);
- *manipolare i prestiti* (h);

- *manipolare la vicinanza dei prestiti* (i);
- *manipolare la presenza dello zero*(j);
- *calcolo mentale* (k).

Alcune di queste abilità possono essere considerate indipendenti tra loro, come ad esempio possedere l'abilità *manipolare le migliaia* è indipendente dal possedere o meno l'abilità relativa al *calcolo mentale*. La relazione di dipendenza tra altre è invece piuttosto chiara, come ad esempio possedere l'abilità *manipolare le migliaia* richiede necessariamente l'utilizzo anche dell'abilità *manipolare decine e centinaia*, e così via. In questo modello cognitivo, l'abilità *manipolare i prestiti* è stata successivamente scomposta in tre distinte *sotto-abilità*, ovvero: (h_1) *un prestito*; (h_2) *due prestiti*; (h_3) *tre prestiti*. Anche tra queste tre sotto-abilità esiste una chiara relazione di dipendenza.

In generale, la relazione di dipendenza tra queste abilità si può specificare nella forma di relazioni di *quasi-ordini*, utilizzando il simbolo $\prec$. Formalmente, date due abilità $x, y \in S$, $x \prec y$ sta a significare che non possedere l'abilità x implica non possedere l'abilità y. I quasi ordini ipotizzati nel modello cognitivo descritto sopra sono:

- $f \prec g$;
- $h_1 \prec h_2 \prec h_3$;
- $h_1 \prec j$.

In accordo con la corrispondenza nota tra i quasi-ordini e gli spazi di conoscenza (strutture chiuse sia all'unione che all'intersezione, si veda Sezione 1.2.2), (Falmagne et al., 1990; Birkhoff, 1999), è stato possibile derivare la struttura di competenza corrispondente alla relazione $\prec$. Inoltre Falmagne et al. (1990) hanno dimostrato che ogni struttura finita, chiusa sia all'unione che all'intersezione è necessariamente well graded.

A questo punto, è stato costruito un test composto da 8 problemi sulla sottrazione, in cui ciascun problema era in relazione con le abilità descritte sopra, secondo la skill function illustrata nella Tabella 6.1. E' stato quindi

Problemi	Stati di Competenza	Problemi	Stati di Competenza
(1) 73 - 58	$\{h_1\}$	(5) 253 - 178	$\{f, h_1, h_2, h_3, i\}$
(2) 317 - 94	$\{f, h_1\}$	(6) 2245 - 418	$\{f, g, h_1, h_2\}$
(3) 784 - 693	$\{f, h_1, i\}$	(7) 156 - 68	$\{f, h_1, h_2, h_3, i\}$
(4) 507 - 49	$\{f, h_1, h_2, h_3, j\}, \{k\}$	(8) 3642 - 753	$\{f, g, h_1, h_2, h_3\}$

Tabella 6.1: Skill function della relazione tra i problemi e le abilità del modello cognitivo costruito nell'applicazione empirica (funzione p di Korossy).

costruito un test carta e matita a risposte aperte, composto da 8 problemi, che è stato somministrato a un campione di 298 bambini (148 femmine) italiani iscritti alla classe terza della scuola primaria. Le risposte di ciascun bambino a ciascun item sono state classificate come corrette o errate, ottenendo per ciascun bambino il relativo pattern di risposta R (si ricorda che un pattern di risposta è l'insieme delle risposte corrette).

Il DBLIM, rappresentato dall'Equazione (6.3), è stato quindi applicato all'insieme dei pattern di risposta dei 298 partecipanti, e i suoi parametri sono stati stimati per massima verosimiglianza. La bontà di adattamento del modello è stata testata utilizzando la statistica di Chi-quadro, e a causa della matrice sparsa dei dati, il suo p-value è stato ottenuto attraverso una procedura di bootstrap parametrico (si veda la Sezione 1.3.1 per i dettagli), con 1000 replicazioni.

Con l'intenzione di fare un confronto tra il DBLIM e il modello sbBLIM (6.1), anche quest'ultimo è stato applicato agli stessi dati, con la stessa skill function. L'interesse era rivolto all'individuazione di eventuali violazioni probabilistiche del modello sbBLIM. A questo scopo, sono state calcolate le probabilità congiunte e quelle marginali delle abilità in entrambi i modelli e

confrontate le une alle altre. Per fare un esempio, siano $x, y \in S$ due abilità, allora la probabilità congiunta di x e y si calcola secondo l'equazione seguente:

$$P(x, y) = \sum_{C \in \mathcal{C}_{x,y}} P(\mathcal{C}),$$

dove $\mathcal{C}_{x,y} = \{C \in \mathcal{C} : \{x, y\} \subseteq \mathcal{C}\}$. Per quanto riguarda le probabilità marginali di x e y, si calcolano invece, secondo l'equazione:

$$P(x) = \sum_{C \in \mathcal{C}_x} P(\mathcal{C}),$$

dove $\mathcal{C}_x = \{C \in \mathcal{C} : \{x\} \subseteq \mathcal{C}\}$. In accordo con la corrispondenza richiesta, se x e y sono indipendenti in senso insiemistico, allora devono esserlo anche in senso probabilistico. Per cui l'uguaglianza

$$P(x, y) = P(x)P(y)$$

deve essere vera. Probabilità congiunta e probabilità marginale sono state calcolate per ogni coppia di abilità definita nel modello cognitivo ipotizzato, separatamente per il modello sbBLIM e per il DBLIM. Naturalmente per quest'ultimo modello le violazioni di indipendenza non si osserveranno mai, al contrario di quello che potrebbe accadere con il sbBLIM.

6.3.2 Risultati

La Tabella 6.2 illustra le statistiche di fit riassuntive. Come si può notare il modello DBLIM ottiene un Chi-quadro pari a 249.6 ($df = 231$), mentre il modello sbBLIM ottiene un Chi-quadro pari a 244.1 ($df = 222$). In entrambi i modelli il p-value è risultato maggiore a .10, ad indicare un adattamento ai dati piuttosto buono. Dato che il modello delle dipendenze è una restrizione del modello sbBLIM, i due modelli sono annidati l'uno all'altro. La differenza tra i loro Chi-quadro approssimerà pertanto, anch'essa, a una distribuzione di Chi-quadro. La differenza è pari a $249.6 - 244.1 = 5.5$ che con $231 - 222 = 9$ gradi di libertà non è significativa. Questo porta dunque ad un rifiuto del

Modello	Chi-quadro	df	p-value
DBLIM	249.6	231	$> .10$
sbBLIM	244.1	222	$> .10$

Tabella 6.2: Statistiche di fit riassuntive relative all'applicazione ai dati dei due modelli probabilistici sbBLIM e DBLIM.

modello più complesso, ovvero del sbBLIM, e alla conclusione che il modello di dipendenza spiega meglio i dati di quanto non faccia l'altro.

Una delle ragioni possibili di questo risultato può dipendere proprio dalle violazioni di indipendenza del modello sbBLIM. Per verificare ciò, le probabilità congiunte e quelle marginali sono state calcolate per ogni coppia di abilità del modello cognitivo, sulla base delle stime delle probabilità degli stati di competenza ottenute dal modello.

La Tabella 6.3 illustra le probabilità marginali e quelle congiunte delle coppie di abilità per le quali ci si aspettava indipendenza probabilistica. L'ultima riga (rispettivamente l'ultima colonna) della tabella, illustra le sti-

	f	g	h_1	h_2	h_3	i	j	k	
g									.895
h_1	.959	.895							.970
h_2	.933	.895							.933
h_3	.906	.868							.906
i	.928	.879	.928	.914	.887				.928
j	.621	.604				.607			.621
k	.621	.604	.621	.621	.621	.607	.621		.621
	.959	.895	.970	.933	.906	.928	.621	.621	

Tabella 6.3: Violazioni dell'indipendenza tra abilità nel modello sbBLIM

me delle probabilità marginali delle 8 abilità del modello cognitivo ipotizzato. Ciascuna cella della tabella contiene un valore solamente se la corrispondente coppia di abilità è indipendente, secondo le attese del modello cognitivo. In

caso affermativo, il valore nella cella dovrebbe uguagliare il prodotto delle probabilità marginali delle due abilità. Per fare un esempio, se si considera la coppia di abilità (h_1, g), si ottiene che il prodotto delle corrispondenti probabilità marginali $.959 \times .970 = .930$ è diverso dalla loro probabilità congiunta .959. In generale, ciò che si osserva dalla Tabella 6.3 è che queste uguaglianze non sono mai rispettate, a supporto dell'ipotesi che l'indipendenza probabilistica è violata dal modello sbBLIM e non corrisponde dunque all'indipendenza specificata a livello deterministico dalla struttura di competenza $\mathcal{C}$.

6.4 Conclusioni

Nella letteratura scientifica relativa alla *competence-based* KST, è stata data molta attenzione agli aspetti deterministici della teoria, mentre quelli probabilistici sono stati trascurati. Non esistevano infatti modelli probabilistici per la validazione empirica delle strutture di competenza. In questo capitolo si è presentata una ricerca che aveva l'obiettivo di sviluppare e testare un modello probabilistico per strutture di competenza.

Il modo più immediato per derivare un modello adatto a questo scopo, era quello di estendere il BLIM, calcolando la probabilità degli stati di competenza a partire dagli stati di conoscenza. Questo modello, chiamato *skill-based* BLIM (sbBLIM), però, non fa alcun tipo di assunzione circa l'indipendenza/dipendenza tra le abilità di un modello cognitivo, cosa che invece è ben chiara nella struttura di competenza. Quello che si può verificare, è una mancata corrispondenza tra il modello deterministico, specificato dalla struttura, e il modello probabilistico.

E' stato dunque sviluppato un modello probabilistico per strutture di competenza, chiamato *Dependence* BLIM (DBLIM), che permette di specificare le relazioni di indipendenza/dipendenza tra le abilità. Alla base del

modello c'è il requisito della corrispondenza tra l'indipendenza insiemistica e l'indipendenza probabilistica.

Il modello è stato applicato a un campione raccolto con 298 bambini italiani della terza classe della scuola primaria, ai quali è stato somministrato un test composto da 8 problemi sulla sottrazione. Questi 8 problemi sono stati messi in relazione a un insieme di 8 abilità, ipotizzate alla base della loro risoluzione. Entrambi i modelli, sbBLIM e DBLIM, sono stati applicati ai dati, e confrontati tra loro. I risultati mostrano chi entrambi i modelli ottengono un adattamento ai dati piuttosto buono, anche se da un confronto dei due, emerge che il DBLIM spiega meglio i dati. Questo risultato sta a indicare che, anche se maggiormente restrittivo, quest'ultimo modello permette di catturare piuttosto bene le caratteristiche essenziali dei dati empirici, cosa che il modello sbBLIM non fa. Quest'ultimo infatti, non facendo alcun tipo di assunzione sulla relazione tra le abilità del modello cognitivo, non modella esplicitamente le dipendenze, che sono quindi fuori dal controllo del ricercatore.

Capitolo 7

La Valutazione Efficiente delle Abilità

La KST è stata sviluppata, fin dalle sue origini, con l'obiettivo di costruire un intelligent tutoring system per la valutazione della conoscenza degli individui. Nella Sezione 1.4 si è illustrato come il formalismo della teoria possa essere utilizzato per la costruzione di sistemi di questo tipo. Il sistema ALEKS è sicuramente quello che ha conseguito, in quest'ottica, il maggior successo. Va però evidenziato che ALEKS, così come la maggior parte delle applicazioni computerizzate della teoria, conseguono la valutazione ad un livello che, nella Sezione 2.1, abbiamo chiamato *performance*.

Nel Capitolo 2 è stato presentato un approccio della teoria, chiamato *competence-based* KST (CbKST), che sposta l'attenzione sugli aspetti cognitivi legati alla risoluzione di un insieme di problemi, considerando dunque il livello chiamato *delle competenze*. L'obiettivo è capire *come* i problemi vengano risolti dagli individui, o in altre parole, individuare quali abilità sono coinvolte nella risoluzione di un insieme di item. In quest'ottica vi è dunque la possibilità di costruire un intelligent tutoring system basato sulla CbKST, con l'obiettivo di individuare lo stato di competenza, e quindi le abilità che possiede uno studente.

Volendo perseguire questo scopo, si incontrano però delle problematiche

legate alla mancanza di una corrispondenza biunivoca tra il livello performance e il livello delle competenze, sia per quanto riguarda gli aspetti deterministici che per quelli probabilistici della CbKST. Nel capitolo precedente si è cercato di risolvere il problema legato alla sfera probabilistica. In questo capitolo ci si occupa invece della sfera deterministica, presentando i risultati salienti riportati nell'articolo di Stefanutti e de Chiusole (2014).

7.1 Introduzione

Come visto nella Sezione 2.2, il livello performance e il livello delle competenze sono connessi tra loro tramite due funzioni, chiamate *skill function* e *problem function*. Con l'assegnazione delle abilità ai problemi, la skill function va dal livello performance a quello delle competenze. La problem function invece va nella direzione opposta: dato uno stato di competenza C, essa specifica il sottoinsieme $K \subseteq Q$ di problemi che possono essere risolti da C.

L'obiettivo di un assessment con la CbKST è quello di inferire lo stato di competenza di un individuo dalle sue risposte osservate (codificate come errate o corrette) a un insieme di problemi del dominio Q. Essenzialmente, l'assessment coinvolge due passaggi separati:

1. inferire lo stato di conoscenza K dalle risposte ai problemi;

2. derivare lo stato di competenza C da K.

Nel primo passaggio l'inferenza è di tipo probabilistico dal momento che la risposta a un problema può essere il risultato di un errore di distrazione o di una lucky guess (Falmagne & Doignon, 1988a, 1988b). Il secondo passaggio invece, è di tipo deterministico e si basa sulla problem function. Dal momento che questa è una funzione che va dagli stati di competenza agli stati di conoscenza, se fosse una funzione biunivoca la sua inversa potrebbe

essere utilizzata per inferire lo stato di competenza che corrisponde agli stati di conoscenza individuati con il passaggio (1).

Tuttavia, Heller, Stefanutti, Anselmi, e Robusto (2014) hanno evidenziato che la problem function potrebbe non essere biunivoca e per questo motivo, non ammettere una funzione inversa. E' proprio questa condizione che pone un certo numero di problematiche legate alla valutazione della conoscenza e dell'apprendimento degli studenti nell'ambito della CbKST. Il problema diviene a maggior ragione evidente, nel caso in cui si voglia monitorare l'apprendimento di uno studente attraverso sessioni ripetute di valutazione. E' questo il caso di un intelligent tutoring system che alterna in continuazione sessioni di valutazione a sessioni di apprendimento. Quello che potrebbe succedere è che un cambiamento a livello dello stato di competenza non si rifletta nel cambiamento di uno stato di performance.

Il lavoro che si presenta in questo capitolo si sviluppa sulla nozione di *abilità efficace*. Una volta appresa dallo studente che si trova nello stato di competenza C, l'abilità efficace produce sempre un cambiamento omologo nel suo stato di conoscenza. Se questo tipo speciale di abilità si potesse individuare per ogni stato di competenza, allora la valutazione degli individui con la CbKST diverrebbe possibile.

Nella Sezione 7.2 si presenta nello specifico il problema di dedurre lo stato di competenza di un individuo a partire dal suo stato di conoscenza. I contributi teorici principali della ricerca vengono poi illustrati, sotto forma di teoremi, nelle Sezioni 7.3 e 7.4. A questo proposito si vuole sottolineare che i teoremi si presentano senza dimostrazione (si faccia riferimento a Stefanutti e de Chiusole (2014) per le dimostrazioni). Segue una breve conclusione (Sezione 7.5).

7.2 Floor e Ceiling di uno Stato di Competenza

Il materiale riassunto in questa sezione si basa sul lavoro di Heller et al. (2014). I risultati teorici presi da questo articolo (Proposizioni 4, 5 e 6) si presentano senza le dimostrazioni.

Si ricorda che μ denota una skill function congiuntiva per gli insiemi finiti e non vuoti Q, per gli item e S, per le abilità. La problem function indotta da μ si denota invece con la lettera p. Come hanno dimostrato Heller et al. (2014) la problem function p non è necessariamente biunivoca. Si consideri l'esempio seguente.

Esempio 1. Con $Q = \{a, b, c, d\}$ e $S = \{s, t, u\}$, si consideri la skill function congiuntiva

$$\mu(a) = \{\{s\}\}, \quad \mu(b) = \{\{s, t\}\}, \quad \mu(c) = \{\{s, u\}\}, \quad \mu(d) = \{\{t, u\}\}.$$

La problem function che corrisponde a μ delinea la seguente struttura di conoscenza:

$$\begin{array}{llll} p(\emptyset) = \emptyset & p(\{s\}) = \{a\} & p(\{t\}) = \emptyset & p(\{u\}) = \emptyset \\ p(\{s, t\}) = \{a, b\} & p(\{s, u\}) = \{a, c\} & p(\{t, u\}) = \{d\} & p(S) = Q. \end{array}$$

Questa problem function non è biunivoca perché, per esempio, tutti e tre gli stati di competenza $\emptyset$, $\{t\}$, e $\{u\}$ delineano lo stesso stato di conoscenza $\emptyset$.

Questa mancanza di corrispondenza biunivoca, induce una relazione di equivalenza $\sim_p$ sugli stati di competenza $C \subseteq S$. Dati due stati $C, C' \subseteq S$

$$C \sim_p C' \text{ se e solo se } p(C) = p(C').$$

Parallelamente dato uno stato di competenza $C \subseteq S$, si denota con

$$[C]_p := \{C' \subseteq S : C' \sim_p C\}$$

la classe di equivalenza dello stato di competenza C, indotto da p. La classe $[C]_p$ è quindi la collezione di tutti i sottoinsiemi di S che delineano lo stesso stato di competenza C. Tuttavia se la skill function è congiuntiva (come si assume in questa sezione), allora questa classe di equivalenza si caratterizza per una proprietà interessante.

Proposizione 4. *La classe di equivalenza $C \subseteq S$ è chiusa sotto intersezione e quindi contiene il suo limite inferiore. Formalmente:*

$$\bigcap [C]_p \in [C]_p.$$

Si può dimostrare che il limite inferiore di $[C]_p$ è la collezione di tutte le abilità minimamente sufficienti per delineare lo stato di conoscenza $K = p(C)$. Questa collezione è chiamata *floor* di C, e si denota con

$$\lfloor C \rfloor_p = \bigcap [C]_p.$$

Se la skill function è congiuntiva allora ogni stato di competenza $C \subseteq S$ ammette un floor e, come formalizzato dalla proposizione che segue, la collezione di tutti i floor è chiusa all'unione.

Proposizione 5. *La collezione $\mathcal{C}_p := \{\lfloor C \rfloor_p : C \subseteq S\}$ è chiusa all'unione e forma dunque uno spazio di competenza sull'insieme $\bigcup \mathcal{C}_p \subseteq S$.*

Nelle pagine che seguono ci si riferisce a $\mathcal{C}_p$ come allo *spazio di competenza indotto dalla problem function* p, e ad ogni $C \in \mathcal{C}_p$ come allo *stato di competenza minimo*, a significare che ogni stato di competenza minimo è il floor di qualche stato di competenza $C \subseteq S$. Il risultato interessante è che, mentre lo stato di conoscenza e lo stato di competenza non sono necessariamente in una relazione biunivoca, per ogni stato di conoscenza esiste esattamente uno stato di competenza minimo. Più precisamente, esiste un isomorfismo d'ordine fra lo spazio di competenza minimale e la struttura di conoscenza.

Proposizione 6. *Sia $\mathcal{K}_p = \{p(C) : C \subseteq S\}$ la struttura di conoscenza delineata da p. Allora sotto la restrizione $p^* : \mathcal{C}_p \to \mathcal{K}_p$ la funzione p è un isomorfismo d'ordine da $\mathcal{C}_p$ a $\mathcal{K}_p$.*

Grazie a questa corrispondenza biunivoca, uno stato di conoscenza $K \in \mathcal{K}_p$ veicola un'informazione non ambigua circa il sottoinsieme minimo di abilità che uno studente *padroneggia*. Non solo: lo stato di competenza dello studente fornisce informazioni non ambigue anche a proposito del sottoinsieme minimo di abilità che *non padroneggia*. Sia $C \subseteq S$ uno dei possibili stati di competenza che delineano lo stato di conoscenza $K = p(C)$. Allora, tutte le abilità contenute nel complemento di $\bigcup[C]_p$, sono abilità che lo studente non padroneggia. Se si definisce il *ceiling* di C come

$$\lceil C \rceil_p := \bigcup [C]_p,$$

allora $S \setminus \lceil C \rceil_p$ è l'insieme minimo di abilità che uno studente nello stato di conoscenza $K = p(C)$ non padroneggia.

Per riassumere, ogni stato di competenza $C \subseteq S$ (o, equivalentemente, ogni stato di competenza $K \in \mathcal{K}_p$) suddivide l'insieme totale S di abilità , in tre sottoinsiemi:

- la collezione minima $\lfloor C \rfloor_p$ di abilità che uno studente possiede,
- la collezione minima $S \setminus \lceil C \rceil_p$ di abilità che uno studente non possiede,
- la collezione $\lceil C \rceil_p \setminus \lfloor C \rfloor_p$ di tutte le abilità la cui classificazione è sconosciuta.

Questi risultati rendono chiaro che, in generale, esiste solo un'informazione parziale circa lo stato di competenza sottostante a uno stato di conoscenza. Più precisamente, la valutazione dello stato di competenza di uno studente non può andare oltre a un'approssimazione. Nelle pagine seguenti, tale approssimazione viene indicata con la coppia $\langle \lfloor C \rfloor_p, \lceil C \rceil_p \rangle$.

7.3 Le Frange di uno Stato di Competenza

Nell'ottica dell'apprendimento è importante identificare tutte le abilità che sono immediatamente accessibili per uno studente a partire dal suo stato di

competenza. Nel Capitolo 1 si è visto che nell'approccio tradizionale della KST questo è possibile nel caso in cui le strutture di conoscenza siano degli spazi di conoscenza well-graded (Sezione 1.2.2). In questo caso uno stato di conoscenza ammette sempre una frangia esterna. La nozione di frangia esterna e di spazio well-graded può essere facilmente estesa all'approccio della CbKST.

Definizione 8. Data una struttura di competenza $(S, \mathcal{C})$, la frangia esterna di uno stato di competenza $C \in \mathcal{C}$ è

$$C^{\circ} = \{s \in S : C \cup \{s\} \in \mathcal{C}\}.$$

Inoltre, se $\mathcal{C}$ è uno spazio di competenza per l'insieme finito di abilità S, allora è well-graded se per ogni insieme non vuoto $C \in \mathcal{C}$ esiste $s \in C$ tale che $C \setminus \{s\} \in \mathcal{C}$.

In questo modo, se un'abilità s appartiene alla frangia esterna di uno stato di competenza C allora quell'abilità può essere appresa da C senza il bisogno di apprendere altre abilità. Se uno stato di competenza ha una frangia esterna non vuota, allora le abilità in S possono essere apprese una alla volta, procedendo gradualmente verso stati di conoscenza sempre più grandi, fino a che lo studente giunge a padroneggiare l'insieme totale S.

In questa sezione si userà un concetto leggermente differente rispetto a quello di frangia esterna di uno stato di competenza. Si utilizzerà infatti il concetto di frangia esterna relativa al floor di uno stato di competenza.

Definizione 9. Dato uno stato di competenza $C \subseteq S$ appartenente allo spazio di competenza $\mathcal{C}_p$, la *frangia esterna minimale* di C è

$$C^{*} = \{s \in S \setminus \lfloor C \rfloor_p : \lfloor C \rfloor_p \cup \{s\} \in \mathcal{C}_p\}.$$

Dal momento che l'analisi è ristretta al caso speciale in cui la struttura di competenza è l'insieme potenza di S, è evidente che la frangia esterna di ogni stato $C \in 2^S$ non è mai vuota e che ogni abilità può essere appresa da

ogni stato di competenza. Ciò nonostante anche in questo caso, è fondamentale osservare che non tutte la abilità nella frangia esterna di uno stato di competenza giocano lo stesso ruolo. A questo scopo, si consideri l'esempio che segue.

Esempio 2. Si consideri la skill function dell'Esempio 1 e si supponga che uno studente sia nello stato di competenza $C = \emptyset$. La frangia esterna di questo stato è $C^\circ = \{s, t, u\}$. Apprendendo l'abilità s questo studente si sposterà dall'insieme vuoto all'insieme $\{s\}$, e il suo stato di conoscenza cambierà da $\emptyset$ a $\{a\}$. In questo caso il cambiamento a livello delle competenza corrisponde a un cambiamento a livello performance. Si supponga ora che anziché imparare l'abilità s, lo studente apprenda l'abilità t. Allora, il suo stato di competenza si sposterà dall'insieme vuoto a $\{t\}$. In questo caso però il cambiamento a livello delle competenze non si riflette a livello performance, infatti lo stato di conoscenza delineato da $\{t\}$ è ancora l'insieme vuoto. Si può dire che l'abilità s è *efficace* nel trasformare lo stato di conoscenza dello studente, mentre l'abilità t non lo è.

In generale, ciò che si è voluto evidenziare con l'Esempio 2 è che la frangia esterna di uno stato di competenza può essere suddivisa in due sottoinsiemi:

- uno contenente tutte le abilità che sono *efficaci* nel determinare un cambiamento "osservabile"[1] sullo stato di conoscenza delineato da C;
- e l'altro contenente tutte quelle abilità che non producono cambiamenti osservabili a livello performance.

A questo proposito, si consideri la seguente definizione:

Definizione 10. La *frangia efficace* di uno stato di competenza C è la collezione:

$$C^+ = \{s \in C^\circ : p(C) \subset p(C \cup \{s\})\},$$

[1]Nelle analisi qui presentate si considera la situazione ideale in cui siano esclusi possibili errori di distrazione e di lucky guess. In questo caso dunque il pattern osservato delle risposte corrette e sbagliate di uno studente, corrisponde al suo stato di conoscenza

mentre la *frangia inefficace* è il complemento $C^- = C^\circ \setminus C^+$.

Si supponga ora che un tutor voglia individuare l'abilità da far apprendere ad uno studente che si trova nello stato di competenza C. In condizioni ideali (ovvero assumendo zero rumore) ciò che il tutor conosce è lo stato di conoscenza K, e sulla base di questa informazione vuole individuare l'abilità che, una volta appresa dallo studente, trasformi il suo stato di conoscenza K in uno più grande. Il tutor dovrà quindi individuare l'abilità che appartiene alla frangia esterna efficace C^+ dello stato di competenza attuale. Ma, per individuare C^+ è necessario conoscere lo stato C, di cui però conosce solamente un'approssimazione, rappresentata dalla coppia $\langle \lfloor C \rfloor_p, \lceil C \rceil_p \rangle$. Ci si è chiesti allora se la frangia efficace di C potesse essere almeno parzialmente ricostruita dalla sua approssimazione. Il seguente teorema risponde a questa domanda.

Teorema 1. *La frangia esterna minimale di uno stato di competenza $C \subseteq S$ è inclusa nella sua frangia efficace, cioè $C^* \subseteq C^+$.*

Quindi, se la frangia esterna minimale C^* non è vuota, e il tutor seleziona una delle abilità in C^*, allora sta selezionando un'abilità efficace. Oltre a questo è importante considerare un altro risultato che chiarisce la relazione tra frangia esterna minimale e ceiling di uno stato di competenza.

Teorema 2. *La frangia esterna minimale di uno stato di competenza $C \subseteq S$ è il complemento del suo ceiling, ovvero l'uguaglianza $C^* = S \setminus \lceil C \rceil_p$ è vera per ogni $C \subseteq S$.*

Richiamando l'esempio, grazie ai Teoremi 1 e 2, il tutor saprà che ogni abilità appartenente al complemento del ceiling di C è efficace e, nel caso in cui l'abilità sarà appresa dallo studente, si determinerà un cambiamento osservabile del suo stato di conoscenza.

A questo punto emergono due osservazioni: (1) dal momento che C^* è un sottoinsieme di C^+, la frangia esterna minimale fornisce al tutor una lista

incompleta delle abilità efficaci; (2) nel peggiore dei casi si può verificare che questa frangia sia vuota, lasciando il tutor senza alcuna scelta. L'esempio che segue, illustra questa situazione.

Esempio 3. Si consideri la skill function congiuntiva sugli insiemi $S = \{s, t, u\}$ e $Q = \{a, b, c\}$:

$$\mu(a) = \{\{s\}\}, \qquad \mu(b) = \{\{s, t, u\}\}, \qquad \mu(c) = \{\{t, u\}\},$$

La corrispondente problem function delinea i seguenti stati di conoscenza:

$$p(\emptyset) = \emptyset, \quad p(\{s\}) = \{a\}, \quad p(\{t\}) = \emptyset, \quad p(\{u\}) = \emptyset,$$
$$p(\{s, t\}) = \{a\}, \quad p(\{s, u\}) = \{a\}, \quad p(\{t, u\}) = \{c\}, \quad p(S) = Q,$$

e il seguente spazio di competenza $\mathcal{C}_p = \{\emptyset, \{s\}, \{t, u\}, S\}$. Si supponga che il tutor sappia che lo stato di uno studente è $p(C) = \{a\}$. Questo stato è delineato dai 3 differenti stati di competenza $\{s\}, \{s, t\}, \{s, u\}$. Lo stato di competenza minimale è $\lfloor C \rfloor_p = \{s\}$, e la frangia esterna minimale è $C^* = \emptyset$. Tuttavia, a seconda dello stato attuale C dello studente, la frangia efficace di C potrebbe essere:

1. $C^+ = \emptyset$, se $C = \{s\}$;
2. $C^+ = \{t\}$, se $C = \{s, u\}$;
3. $C^+ = \{u\}$, se $C = \{s, t\}$.

Nel caso 1, non c'è alcun modo di osservare un cambiamento nello stato di conoscenza, apprendendo esattamente un'abilità. Nel caso 2, tutto procederà liscio se il tutor sceglie di insegnare l'abilità t allo studente, e questa venga effettivamente appresa. Il nuovo stato di competenza dello studente sarà $\{s, t, u\}$, che delinea lo stato di conoscenza $\{a, b, c\}$. Il cambiamento a livello delle competenze si rifletterà pertanto in un cambiamento a livello performance. Nel caso 3, se il tutor decidesse di insegnare l'abilità t, lo studente rimarrebbe nel suo stato di competenza $C = \{s, t\}$ e quindi nello stato di

conoscenza $\{a\}$. Il problema in tutti questi casi è che, a livello performance non si osserverebbe nessun cambiamento sia nel caso appena descritto, che nel caso in cui lo studente sia nello stato $\{s, u\}$. In questi due casi lo stato di performance è $\{a\}$. Il tutor non avrebbe idea di come stia proseguendo l'apprendimento dello studente.

L'esempio rende chiaro che, in un intelligent tutoring system efficiente, sarebbe desiderabile avere una frangia esterna minimale non vuota. In particolare, la cosa migliore corrisponderebbe ad avere tutte le abilità efficaci nella frangia esterna minimale dello stato. Dai Teoremi 1 e 2 la frangia esterna minimale di uno stato di competenza C fornisce un'informazione parziale circa la frangia efficace di C. Il prossimo teorema stabilisce le condizioni sotto le quali questa informazione è completa.

Teorema 3. *Se lo spazio di competenza $\mathcal{C}_p$ è well-graded allora, per ogni stato di competenza $C \subseteq S$, la frangia esterna minimale di C e la sua frangia efficace sono uguali: $C^* = C^+$.*

E' interessante notare che la condizione well-graded di uno spazio di competenza $\mathcal{C}_p$ assicura non solo che tutti gli stati di competenza abbiano una frangia esterna minimale non vuota, ma anche che ogni stato di competenza abbia una frangia esterna efficace. L'effetto di questo risultato è che uno studente possa fare progressi osservabili, apprendendo esattamente una abilità per volta.

Va infine evidenziato che, sebbene la condizione well-graded di uno spazio di competenza sia sufficiente per avere una frangia efficace non vuota, non è una condizione necessaria. Si consideri l'esempio seguente

Esempio 4. Sia $Q = \{1, 2, 3, 4, 5\}$ un insieme di item, $S = \{a, b, c, d\}$ un insieme di abilità, e μ la skill function congiuntiva tale per cui:

$$\mu(1) = \{\{a\}\}, \qquad \mu(2) = \{\{b\}\}, \qquad \mu(3) = \{\{a, c\}\},$$
$$\mu(4) = \{\{b, c, d\}\}, \qquad \mu(5) = \{\{a, c, d\}\}.$$

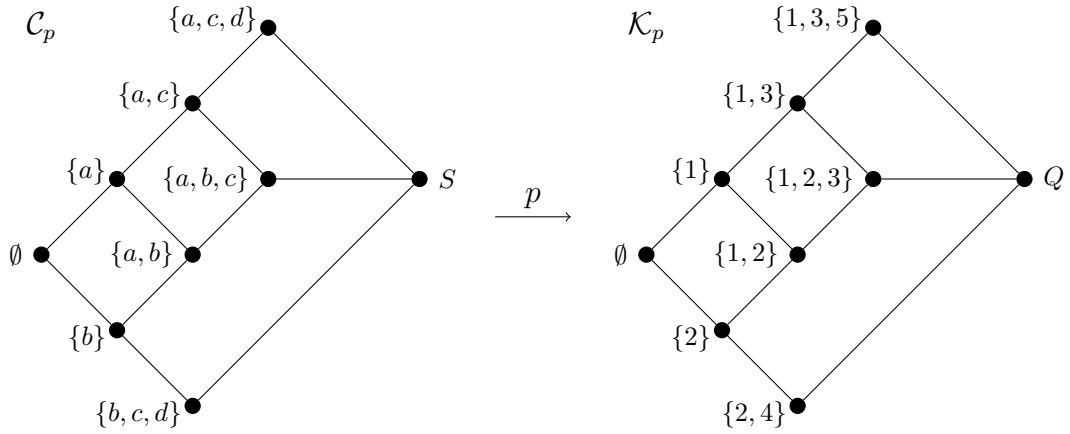

Figura 7.1: Diagramma di Hasse dello stato di competenza $\mathcal{C}_p$ indotto dalla skill function dell'Esempio 4 (digramma di sinistra), e la corrispondente struttura di conoscenza $\mathcal{K}_p$ (diagramma di destra) ottenuta applicando la problem function p a $\mathcal{C}_p$.

La funzione μ delinea la struttura di conoscenza $\mathcal{K}_p$ riportata nel diagramma destro della Figura 7.1, e induce lo spazio di competenza $\mathcal{C}_p$, rappresentato nella parte sinistra della stessa figura. Lo spazio di competenza rappresentato in figura non è well-graded perché non c'è nessuna abilità $s \in S$ tale per cui $\{b,c,d\} \setminus \{s\}$ sia uno stato di competenza. Oltre a ciò, è possibile notare che ogni stato di competenza ammette una frangia esterna minimale non vuota e quindi anche una frangia efficace non vuota. Tuttavia, dato che $\mathcal{C}_p$ non è well-graded, la frangia esterna minimale di qualche stato di competenza ricostruisce solo parzialmente la corrispondente frangia effettiva. Ad esempio, nel caso dello stato $C = \{b,c\}$ la frangia esterna minimale è $\{a\}$, mentre la frangia efficace è $\{a,d\}$.

7.4 Skill Function Esclusive

All'inizio della sezione precedente, si è fatto presente che uno studente può fare progressi osservabili a livello performance, apprendendo esattamente una

abilità per volta, se lo spazio di competenza $\mathcal{C}_p$ è well graded. In questa sezione si analizza come questa proprietà sullo spazio di competenza si possa ottenere sulla base di specifiche caratteristiche della skill function. In altre parole, si sono cercate le proprietà di una skill function congiuntiva μ che rendono lo spazio di competenza $\mathcal{C}_p$ well-graded.

Dal momento che le analisi sono ristrette al caso di skill function congiuntive, in seguito si utilizzerà la notazione $\tau : Q \to 2^S$ di Falmagne et al. (1990), per rappresentare una skill function congiuntiva.

Esempio 5. La skill function congiuntiva (Q, S, μ) dell'Esempio 4 corrisponde alla skill-multimap (Q, S, τ), dove:

$$\tau(1) = \{a\}, \qquad \tau(2) = \{b\}, \qquad \tau(3) = \{a, c\},$$
$$\tau(4) = \{b, c, d\}, \qquad \tau(5) = \{a, c, d\}.$$

Si può notare che la problem function p corrispondente alla skill function congiuntiva (Q, S, τ) si ricava sulla versione semplificata

$$p(C) = \{q \in Q : \tau(q) \subseteq C\}. \tag{7.1}$$

Per continuare sono necessarie altri due concetti: (1) uno spazio di competenza su un insieme finito S ammette sempre una *base*. La base dello spazio $\mathcal{C}$ è la famiglia minimale $\mathcal{B} \subseteq \mathcal{C}$ la cui chiusura all'unione è $\mathcal{C}$ (Doignon & Falmagne, 1999); (2) data una qualsiasi abilità $s \in S$, uno stato di competenza $C \in \mathcal{C}$ è un *atomo in* s se esiste un sottoinsieme minimale in $\mathcal{C}$ contenete s. Allora, la base $\mathcal{B}$ dello spazio $\mathcal{C}$ è la collezione di tutti gli atomi (Teorema 1.26 in Doignon & Falmagne, 1999).

Alcuni item in Q hanno un ruolo particolare non solo nel determinare lo spazio di competenza $\mathcal{C}_p$, ma anche nel determinare se $\mathcal{C}_p$ sia o meno well graded. Si introduce una definizione relativa a questi item.

Definizione 11. In una skill function congiuntiva (Q, S, τ), un item $q \in Q$ è *atomico* per l'abilità $s \in S$ se:

(1) $s \in \tau(q)$,

(2) $\tau(q') \subset \tau(q)$ implica $s \notin \tau(q')$ per ogni $q' \in Q$.

Definendo inoltre la funzione $\alpha : S \to 2^Q$ tale che, per ogni $s \in S$,

$$\alpha(s) := \{q \in Q : q \text{ è atomico per } s\},$$

e sia

$$A = \bigcup \{\alpha(s) : s \in S\}$$

la collezione di tutti gli item atomici.

Come detto sopra, solo gli item atomici hanno un ruolo nel determinare uno spazio di competenza.

Proposizione 7. *Lo spazio di competenza delineato dalla skill function congiuntiva* (Q, S, τ) *ha la base:* $\mathcal{B}_p = \{\tau(q) : q \in A\}$.

A questo punto è possibile introdurre una definizione e un teorema:

Definizione 12. Una skill function congiuntiva (Q, S, τ) è *esclusiva* se nessuna abilità appartenente a S ha lo stesso item come atomico. Formalmente, l'implicazione

$$s \neq t \implies \alpha(s) \cap \alpha(t) = \emptyset.$$

è vera per ogni coppia $s, t \in S$.

Teorema 4. *Lo spazio di competenza* $\mathcal{C}_p$ *è well graded se e solo se la skill function congiuntiva* (Q, S, τ) *è esclusiva.*

Per comprendere la Definizione 12 e il Teorema 4, si consideri il seguente esempio.

Esempio 6. L'applicazione della Definizione 11 alla skill function dell'Esempio 4 delinea la struttura di conoscenza:

$$\alpha(a) = \{1\}, \qquad \alpha(b) = \{2\}, \qquad \alpha(c) = \{3, 4\}, \qquad \alpha(d) = \{4, 5\}.$$

Dal momento che $\alpha(c) \cap \alpha(d) = \{4\} \neq \emptyset$, si conclude che quella skill function non è esclusiva. In particolare, l'item 4 è un atomico sia per l'abilità c che per l'abilità d.

La skill function τ può essere ricondotta alla condizione di esclusività, aggiungendo un item che richieda o l'abilità c o la d, ma non entrambe. Per esempio, definendo una nuova skill function (Q', S, τ') tale che $Q' = Q \cup \{6\}$, $\tau'(q) = \tau(q)$ per ogni $q \in Q$, e $\tau'(6) = \{b, c\}$, si ottiene:

$$\alpha(a) = \{1\}, \qquad \alpha(b) = \{2\}, \qquad \alpha(c) = \{3, 6\}, \qquad \alpha(d) = \{4, 5\}.$$

Dal momento che non ci sono abilità che condividano item atomici, la skill function τ' è esclusiva, e il corrispondente spazio di competenza, rappresentato in Figura 7.2, è well graded.

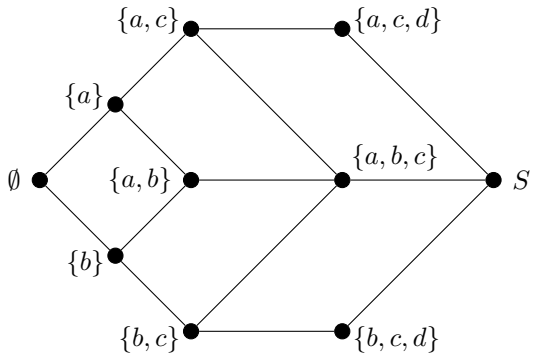

Figura 7.2: Lo spazio di competenza delineato dalla skill function congiuntiva τ' dell'Esempio 6 è well graded.

Riassumendo, l'esclusività di una skill function congiuntiva (Q, S, τ) può essere testata attraverso i seguenti tre passaggi:

(i) per ogni abilità $s \in S$, si definisca l'insieme $\alpha(s)$ controllando, per ogni item $q \in Q$, le condizioni (1) e (2) della Definizione 11. Gli elementi di $\alpha(s)$ sono gli item in Q che soddisfano entrambe le condizioni;

(ii) per ogni coppia di abilità $s, t \in S$ si verifichi la condizione di esclusività (Definizione 12).

(iii) se la condizione di esclusività è vera per ogni coppia di abilità, allora la skill function (Q, S, τ) è esclusiva, e delinea uno spazio di competenza well graded.

7.5 Conclusioni

L'obiettivo della valutazione nella CbKST è di inferire lo stato di competenza di uno studente, sulla base delle risposte osservate a un insieme di problemi. In questa direzione, il problema maggiore dipende dal fatto che il livello performance e il livello delle competenze non stanno in relazione biunivoca tra loro. Nella Sezione 7.2, è stato evidenziato che è comunque possibile condurre una valutazione, sebbene parziale. Si conoscono infatti due sottoinsiemi: l'insieme minimale di abilità che uno studente possiede e l'insieme minimale delle abilità che non possiede.

Questa problematica diviene piuttosto limitante se si voglia costruire un intelligent tutoring system basato sulla CbKST. Infatti, come è stato evidenziato nella Sezione 7.3, i cambiamenti dell'apprendimento che avvengono a livello delle competenze non sempre si riflettono a livello performance. La conseguenza è che un tutor non sarebbe in grado di stabilire se l'apprendimento di specifiche abilità sia avvenuto o meno.

Nella ricerca presentata in questo capitolo è stato evidenziato che l'impasse può essere risolto se si considera una classe speciale di skill function, chiamate *congiuntive*. Esse delineano strutture di competenza chiamate *spazi di competenza*, nelle quali diventa possibile monitorare i progressi di uno studente, che possono avvenire gradualmente, abilità per abilità, fino all'insieme totale. L'aspetto chiave è che questi cambiamenti divengono osservabili anche a livello performance. Se inoltre, uno spazio di competenza è well graded si viene a soddisfare una particolare proprietà della skill function che è stata chiamata *esclusività*. Nella Sezione 7.4 è stato presentato un test per verificare questa proprietà.

A questo proposito vi sono tre considerazioni importanti da fare su quelli che sono i possibili sviluppi di questa ricerca:

- in primo luogo occorre approfondire se i risultati ottenuti per le skill function esclusive si possano in qualche modo generalizzare alle skill function non esclusive;

- inoltre i risultati ottenuti in questo lavoro si applicano solamente al caso di skill function congiuntive. C'è da chiedersi se sia possibile estendere i risultati anche al caso delle skill function in generale.

- infine, un altro limite riguarda il fatto che sono state considerate strutture di conoscenza che corrispondono all'insieme potenza delle abilità. Di fatto, con il Capitolo 6 è stato visto come questa situazione possa essere poco realistica. Sarebbe pertanto interessante prendere in considerazione anche strutture di competenza che sono un sottoinsieme di 2^S.

Capitolo 8

Discussione Generale

Le ricerche che sono state presentate in questa tesi si sviluppano entro la knowledge space theory (KST, Falmagne et al., 1990; Falmagne & Doignon, 2011), una teoria matematica recente che fornisce un importante quadro di riferimento formale per lo sviluppo di sistemi computerizzati web-based che abbiano l'obiettivo di *valutare la conoscenza e l'apprendimento degli individui*. Il suo approccio differisce notevolmente da quelli psicometrici tradizionali, sviluppati allo stesso scopo. Infatti, l'obiettivo della valutazione nella KST è *descrivere* quello che uno studente sa e non sa in un particolare dominio di conoscenza, senza per questo attribuire un punteggio numerico per quantificare la conoscenza o l'apprendimento di uno studente.

La nozione al centro dell'intera teoria è quella di *stato di conoscenza*, cioè l'insieme dei problemi q che uno studente è capace di risolvere, fra quelli che si possono formulare in un certo dominio di conoscenza Q. La collezione di tutti gli stati di conoscenza formulabili in una popolazione di studenti costituisce una *struttura di conoscenza* $\mathcal{K}$. Uno degli obiettivi principali della KST è di specificare la struttura di conoscenza su Q, stabilendo delle regole per separare i sottoinsiemi di Q che sono stati della struttura da quelli che non lo sono. Questo compito permette di individuare una serie di relazioni (come ad esempio *relazioni di prerequisito*) tra gli item $q \in Q$, nonché una relazione di inclusione tra gli stati della struttura. In questo modo una

struttura di conoscenza, non solo stabilisce quali sono gli stati che caratterizzano una particolare popolazione di studenti, ma stabilisce anche quali sono i percorsi d'apprendimento che uno studente può seguire per passare da uno stato K a uno stato K' più grande, e questo fino a giungere all'acquisizione di tutti i problemi del dominio. Se si considera una classe speciale di strutture di conoscenza, chiamate *spazi di conoscenza* (strutture chiuse all'unione), ciascuno stato di conoscenza (ad eccezione dell'insieme totale) si caratterizza per avere due tipi di frange: una *frangia interna*, che rappresenta l'insieme dei problemi da ultimi appresi e una *frangia esterna*, che rappresenta l'insieme dei problemi che lo studente è pronto ad apprendere. Un ultimo aspetto che si vuole ricordare è che la KST, dato il suo rigore formale, è facilmente trasferibile al linguaggio informatico. In particolare è possibile sviluppare algoritmi computerizzati che consentono di operare una valutazione della conoscenza di tipo adattivo. In questo contesto, la valutazione di uno studente è unica, dal momento che i problemi da somministrare vengono scelti sulla base delle risposte precedenti dello studente. Riassumendo, la valutazione nell'ottica della KST si caratterizza per essere:

- *qualitativa*, preferisce descrivere la conoscenza di uno studente piuttosto che quantificarla;
- *formativa*, è orientata ad individuare ciò che uno studente è pronto ad apprendere;
- *adattiva*, è personalizzata ovvero si costruisce sulla base del particolare studente in esame.

Le strutture di conoscenza sono, di fatto, un modello deterministico teorico dell'organizzazione della conoscenza all'interno di un particolare dominio. La loro validazione empirica è resa possibile grazie alla verifica probabilistica della loro plausibilità. Il *basic local independence model* (BLIM) è un modello probabilistico che è stato sviluppato a questo scopo. Esso si caratterizza

per tre tipi di parametri: una probabilità β_q di distrazione e una probabilità η_q di lucky guess per ciascun item $q \in Q$, e una probabilità π_q per ciascuno stato K appartenente alla struttura di conoscenza $\mathcal{K}$. Nonostante sia il modello più utilizzato nella KST, alcuni problemi circa la sua applicabilità rimanevano ancora aperti. L'obiettivo generale della presente tesi è stato quello di rispondere a queste domande per conferire una maggiore validità alle applicazioni empiriche del modello.

Un primo quesito al quale si è dato risposta è stato quello di derivare la matrice di covarianza delle stime dei parametri del BLIM, che ancora non si conosceva. Questa matrice è particolarmente importante per calcolare la varianza dei suoi parametri e dunque ottenere gli intervalli di confidenza. Oltre a ciò, avere a disposizione le formule analitiche per il calcolo della varianza consente di studiare il comportamento asintotico di questa statistica e individuare eventuali errori sistematici nella stima dei parametri del modello, sotto determinate condizioni. Seguendo le indicazioni della letteratura scientifica sulla derivazione della matrice di covarianza dei modelli multinomiali, è stata per prima derivata la matrice d'informazione di Fischer del BLIM (Proposizione 1), per poi derivare la sua matrice di covarianza. Questo risultato teorico è stato utilizzato per sviluppare un tool MATLAB per il calcolo della varianza delle stime dei parametri del BLIM ed è stato applicato sia in uno studio simulativo che in uno empirico. Lo studio simulativo ha permesso di individuare una serie di condizioni critiche per la stima dei parametri del BLIM. In particolare si attende un'incertezza elevata della probabilità di lucky guess, nel caso di item molto facili e di careless error nel caso di item molto difficili. Per quanto riguarda i risultati dell'applicazione empirica, essi suggeriscono che gli intervalli di confidenza hanno un valore diagnostico nell'individuazione di specificazioni errate del modello: la stima puntuale particolarmente elevata del parametro di un item associata a un intervallo di confidenza piccolo, è indice di una possibile specificazione errata del modello per quel particolare item. Questa ricerca fornisce dunque gli

strumenti per operare una interpretazione corretta delle stime dei parametri che si ottengono dall'applicazione del BLIM a dati empirici.

Un secondo quesito a cui si è data risposta riguardava l'assunzione di invarianza dei parametri del BLIM. Da un punto di vista empirico tale assunzione stabilisce che le probabilità di distrazione (parametro careless error) e di indovinare la risposta (lucky guess) sono invarianti attraverso gli stati di conoscenza. Tali parametri sarebbero dunque una proprietà intrinseca dell'item. Essendo questa un'assunzione implicita del modello è stato evidenziato come possa essere violata dai dati. Per questa ragione è stata sviluppata una procedura che consente di testare tale assunzione. Tale procedura consiste nel confrontare il BLIM con un modello alternativo, più complesso, in cui l'assunzione di invarianza è esplicitamente violata. In questo modello alternativo, chiamato modello a bipartizione (BPM), si basa su una partizione della struttura di conoscenza in due classi: tutti gli stati la cui cardinalità è minore o uguale a un certo cutoff, e tutti i restanti. Esso si caratterizza per un insieme di quattro parametri (anziché due come nel caso del BLIM) per ciascun item: le probabilità di careless error e lucky guess sotto il cutoff e le probabilità di careless error e lucky guess sopra il cutoff. La procedura consiste dunque nel confrontare la fit dei due modelli applicati allo stesso dataset, e a seconda del modello scelto concludere se l'assunzione sia o meno violata dai dati. Dopo aver derivato le formule per le stime dei parametri di questo modello, sono stati condotti uno studio simulativo e uno studio empirico. I risultati delle simulazioni hanno evidenziato come il confronto del BLIM con il BPM (dove il cutoff è rappresentato dalla mediana), sia un modo efficace per individuare le violazioni dell'assunzione di invarianza. Sono in particolare il rapporto di verosimiglianza e l'indice AIC che selezionano il modello corretto: quando l'assunzione è rispettata dai dati selezionano il BLIM, e quando l'assunzione è violata selezionano il BPM. Ciò che ha evidenziato l'applicazione empirica invece è che, nonostante il BLIM ottenga una buona fit ai dati, l'assunzione di invarianza dei suoi parametri d'errore dagli stati di

conoscenza può comunque essere violata. Da questa ricerca emerge dunque l'importanza di testare l'assunzione di invarianza dei parametri del BLIM prima di fare qualsiasi interpretazione sui risultati che si ottengono dalle sue applicazioni empiriche, e fornisce gli strumenti per farlo.

L'ultima ricerca della tesi basata sul BLIM, riguardava il trattamento dei dati mancanti. Il BLIM infatti, è un modello che si applica solamente a data set completi e non fornisce alcuna indicazione su come operare nel caso di dati mancanti. Per rendere possibile l'applicazione del BLIM, quello che si può fare è usare la trasformazione *missing-as-wrong* (MAW), ovvero trasformare le risposte mancanti in risposte errate. Questa operazione implica chiaramente che le risposte mancanti siano sempre determinate dal fatto che lo studente non sappia rispondere, ma questa è un'assunzione piuttosto forte che non poteva essere verificata in alcun modo. Sono state quindi derivate due estensioni di questo modello: la prima, chiamata IMBLIM, assume che i dati mancanti siano di tipo missing-completely-at-random (MCAR), mentre la seconda, chiamata MissBLIM, è adatta al caso in cui vi sono dipendenze tra dati mancanti e stati di conoscenza, ovvero quando il processo che genera i mancanti è missing-not-at-random (MNAR). I due modelli, insieme al BLIM, sono stati messi alla prova sia in uno studio simulativo, sia in un applicazione empirica. Nello studio simulativo sono stati generati dei campione nei quali il processo che generava i dati mancanti poteva essere di MCAR oppure MNAR. I risultati delle simulazioni hanno evidenziato che: (a) quando il processo è MCAR, sia l'IMBLIM che il MissBLIM sono adeguati alla modellazione dei dati mancanti; (b) quando il processo è MNAR, l'unico modello adatto alla modellazione dei mancanti è il MissBLIM; (c) quando si usa la trasformazione MAW i parametri d'errore del BLIM sono affetti da bias, indipendentemente dal processo che li ha generati. Per quanto riguarda l'applicazione empirica, l'IMBLIM e il MissBLIM sono stati applicati a un campione che conteneva dati mancanti. L'obiettivo era testare se, attraverso il confronto dei due modelli, fosse possibile individuare il processo che aveva

generato i dati mancanti. I risultati evidenziano che i dati mancanti non erano indipendenti dallo stato di conoscenza degli studenti, erano quindi di tipo MNAR. E' stato quindi dimostrato come trasformare i dati mancanti in risposte errate, è del tutto inappropriato, perché porta a una errata valutazione dello stato di conoscenza degli studenti. Da questo punto di vista emerge ancor più chiaramente l'importanza di avere a disposizione modelli flessibili, come il MissBLIM, che si adattano alle più svariate situazioni in cui i dati mancanti si possono osservare.

Nella KST, la nozione di stato di conoscenza è puramente comportamentale, nel senso che non ha alcun tipo di interpretazione psicologica o cognitiva. Invece, nella *competence-based* KST (CbKST) l'obiettivo principale della valutazione diviene quello di fornire un'interpretazione cognitiva dello stato di conoscenza, attraverso la definizione di uno *stato di competenza*, ovvero l'insieme delle abilità che possiede uno studente. La conoscenza di un individuo sarebbe dunque caratterizzata da uno stato di conoscenza, se si considera il *livello performance* e da uno stato di competenza, se si considera il *livello delle competenze*. Le altre due ricerche che sono state presentate in questa tesi, si collocano all'interno di questo quadro teorico. Esse avevano l'obiettivo di rispondere ad alcune mancanze che erano ancora presenti nella letteratura scientifica della CbKST, una di tipo probabilistico e l'altra di tipo deterministico.

La quarta ricerca che è stata presentata nella tesi aveva l'obiettivo di sviluppare e testare un modello probabilistico per strutture di competenza. Infatti, se a livello performance esistono modelli probabilistici, come il BLIM, per validare le strutture di conoscenza, questo non era vero per il livello delle competenze. Il modo più immediato per derivare un modello adatto a questo scopo, era quello di estendere il BLIM, calcolando la probabilità degli stati di competenza a partire dagli stati di conoscenza. Questo modello, chiamato *skill-based* BLIM (sbBLIM), però, non fa alcun tipo di assunzione circa l'indipendenza/dipendenza tra le abilità di un modello co-

gnitivo, cosa che invece è ben chiara nella struttura di competenza. Quello che si può verificare, è una mancata corrispondenza tra il modello deterministico, specificato dalla struttura, e il modello probabilistico. E' stato dunque sviluppato un modello probabilistico per strutture di competenza, chiamato *Dependence* BLIM (DBLIM), che permette di specificare le relazioni di indipendenza/dipendenza tra le abilità. Alla base del modello c'è il requisito della corrispondenza tra l'indipendenza insiemistica e l'indipendenza probabilistica. Entrambi i modelli sbBLIM e DBLIM, sono stati applicati a un data set reale e confrontati tra loro. I risultati mostrano una buona fit di entrambi, anche se da un confronto dei due, emerge che, anche se è maggiormente restrittivo, il DBLIM spiega meglio i dati. Grazie ai risultati ottenuti in questa ricerca, la validazione di strutture di competenza è ora possibile, sia nel caso in cui le abilità sottostanti un insieme di problemi siano indipendenti, sia nel caso in cui non lo siano.

La quinta e ultima ricerca presentata nella tesi, aveva l'obiettivo di studiare la corrispondenza tra il livello performance e il livello delle competenze, e fornire una soluzione al problema che non stanno in corrispondenza biunivoca tra loro. Questo aspetto diviene fortemente problematico se l'obiettivo della valutazione è quello di inferire lo stato di competenza di uno studente, sulla base delle risposte osservate a un insieme di problemi. Infatti quello che può accadere è che, individuato lo stato di conoscenza di uno studente non si sappia individuare in modo univoco quale sia il corrispondente stato di competenza. Nella ricerca presentata è stato dimostrato che questo impasse può essere risolto se si considera una classe speciale di skill function, chiamate *congiuntive*. Esse delineano strutture di competenza chiamati *spazi di competenza*, nelle quali diventa possibile monitorare i progressi di uno studente, che possono avvenire gradualmente, abilità per abilità, fino all'insieme totale. L'aspetto chiave è che questi cambiamenti divengono osservabili anche a livello performance. I risultati ottenuti in questa ricerca aprono la strada allo sviluppo di un intelligent tutoring system basato sulla CbKST, che ha

l'obiettivo di individuare lo stato di competenza di uno studente, e quindi le abilità che possiede.

Si vuole concludere la discussione generale, evidenziando alcune prospettive di ricerca. Esse si collocano su due binari separati ma paralleli. Da un lato ci si riferisce a ricerche di tipo teorico e dall'altra a ricerche puramente di carattere applicativo. Da un punto di vista teorico sono ancora molti i progressi da fare sia nell'ambito della KST sia in quello della CbKST, di seguito se ne discutono alcuni.

Nel caso della KST sarebbe interessante sviluppare un *algoritmo adattivo* che, tenendo conto delle risposte mancanti di un studente, si basi sul MissBLIM anziché sul BLIM. Questo perché, nonostante le procedure computerizzate adattive consentono allo studente di saltare la risposta, trattano questo esito come se fosse una risposta errata. Conoscendo le conseguenze di questa scelta (discusse sopra), sembra necessario estendere l'algoritmo non deterministico continuo di Falmagne e Doignon (1988a), al caso in cui uno studente decida di non rispondere ad un item. Questo consentirebbe di migliorare l'efficienza della procedura, dal momento che quando uno studente non risponde ad un item per una motivazione diversa rispetto al non sapere la risposta, la procedura verrebbe portata fuori strada e giungerebbe ad un aggiornamento errato delle probabilità degli stati.

Una strada altrettanto importante da seguire nell'ambito della valutazione adattiva della conoscenza, riguarda la *costruzione delle strutture di conoscenza*. Individuare correttamente quali sono gli stati di conoscenza che caratterizzano una popolazione di studenti è fondamentale per una valutazione accurata della loro conoscenza. Sebbene ad oggi esistano svariati metodi di costruzione delle strutture di tipo *theory-driven*, ovvero che tengono conto delle relazioni teoriche che esistono tra un insieme di item, sono poche e non convincenti quelle di tipo *data-driven*, ovvero procedure che costruiscono le strutture "interrogando" i dati. Nell'ambito dell'intelligenza artificiale sono state sviluppate con successo procedure di tipo data-driven che si basano, ti-

picamente, su metodi di classificazione. Sarebbe interessante adattare queste tecniche all'ambito della KST.

Vi è poi un altro obiettivo di ricerca, estremamente urgente, sia nella KST che nella CbKST. Ci si riferisce alla necessità di sviluppare *modelli di apprendimento* che consentano di fare una previsione sul percorso di apprendimento di uno studente nel tempo continuo. L'approccio che si può seguire è quello dei processi markoviani bivariati (BMP; Mark & Ephraim, 2012), una classe di modelli piuttosto flessibili e potenti che si basano sull'assunzione che il processo modellato è di tipo markoviano, e avviene nel tempo continuo. Il processo viene chiamato bivariato perché avviene sia ad un livello osservabile che ad un livello latente. Si intravede la possibilità di estendere questo approccio a quello della KST e della CbKST.

Si vuole infine concludere con una prospettiva di ricerca puramente applicativa, ma molto ambiziosa. Alcune delle ricerche che sono state presentate in questa tesi aprono la possibilità allo sviluppo di un intelligent tutoring system basato sulla CbKST. Un *intelligent tutoring system* (ITS), in generale, è un ambiente di apprendimento computerizzato che si basa su modelli matematici derivanti da ambiti di ricerca come quelli delle scienze cognitive, della psicologia dell'apprendimento, dell'intelligenza artificiale e della matematica. L'obiettivo di un ITS è quello di descrivere in modo estremamente accurato la conoscenza di uno studente, per individuare il percorso d'apprendimento più opportuno per lui e seguirlo in questo percorso fornendogli i contenuti didattici di cui necessita. I vantaggi principali degli ITS derivano dal fatto che l'insegnamento avviene in un contesto uno-a-uno, il computer e lo studente, e può dunque essere visto come uno strumento di supporto alla didattica, soprattutto (ma non solo) per gli studenti maggiormente in difficoltà. Nell'ambito della KST è stato sviluppato `Aleks`, che è stato considerato l'ITS che, ad oggi, ha avuto più successo (Desmarais & Baker, 2012). Si possono comunque trovare due punti deboli di `Aleks`: (1) non fornisce un'interpretazione cognitiva dello stato di conoscenza di uno studente, la valutazione si

colloca infatti a livello performance; e (2) fin da quando è stato sviluppato è stato concepito per la valutazione della conoscenza e non per l'apprendimento, anche se recentemente gli sviluppatori abbiano iniziato a muoversi anche in questa direzione.

Si intravede dunque la possibilità sviluppare un ITS basato sulla CbK-ST, che fornisca una valutazione a livello delle competenze, individuando la abilità che possiede uno studente e quelle che è pronto ad apprendere. Sulla base di questa valutazione l'ITS dovrebbe poi essere in grado di individuare il percorso d'apprendimento più opportuno per quello studente, fornendogli il materiale didattico che risponde ai suoi bisogni educativi. In questi anni si sta sviluppando uno strumento di questo tipo, chiamato `KnowLab` (www.knowlab.org), che fino ad oggi è stato utilizzato per la valutazione delle abilità di centinaia di studenti di psicologia dell'Università di Padova, nel dominio di conoscenza della Psicometria. `KnowLab`, al momento, è un prototipo che è oggetto di sperimentazione. Per la valutazione dell'apprendimento, questo sistema implementa e applica alcuni dei risultati teorici descritti in questa tesi.

Bibliografia

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Academiai Kiado.

Albert, D., & Lukas, J. (1999). *Knowledge spaces: Theories, empirical research, and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, *38*(1), 123–140.

Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *The journal of wildlife management*, 912–923.

Bamber, D., & van Santen, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, *29*, 443-473.

Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, *44*, 20-40.

Birkhoff, G. (1999). *Lattice theory.* Providence, R.I.: American Mathematical Society.

Bolt, D. (2007). The present and future of IRT-based cognitive diagnostic models (ICDMs) and related methods. *Journal of Educational Measurement*, *44*(4), 377–383.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333-353.

de la Torre, J., & Lee, Y.-S. (2010). A note on the invariance of the dina model parameters. *Journal of Educational Measurement*, *47*(1), 115–127.

de Chiusole, D., & Stefanutti, L. (2013). Modeling skill dependence in probabilistic competence structures. *Electronic Notes in Discrete Mathematics*, *42*, 41–48.

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2013). Assessing parameter invariance in the BLIM: Bipartition Models. *Psychometrika*, *78*(4), 710–724.

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2014). Modeling missing data in knowledge space theory. *Manuscript under revision*.

de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (in press). Naïve tests of blim's invariance. *The Spanish Journal of Psychology*.

Degreef, E., Doignon, J., Ducamp, A., & Falmagne, J. (1986). Languages for the assessment of knowledge. *Journal of Mathematical Psychology*, *30*, 243-256.

Dempster, A., Laird, N., & Rubin, D. (1997). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, *22*(1-2), 9–38.

DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overwiew: Irt-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, *44*(4), 285–291.

Doignon, J. P. (1994). Knowledge spaces and skill assignments. In G. H. F. e D. Laming (Ed.), *Contributions to mathematical psy-*

chology, psychometrics, and methodology (pp. 111–121). New York: Springer-Verlag.

Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (p. 111-121). New York: Springer-Verlag.

Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*, 175-196.

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge Spaces*. Berlin, Heidelberg, and New York: Springer-Verlag.

Dowling, C. E., & Hockemeyer, C. (2001). Automata for the assessment of knowledge. *Knowledge and Data Engineering, IEEE Transactions on*, *13*(3), 451–461.

Düntsch, I., & Gediga, G. (1995). Skills and knowledege structures. *British Journal of Mathematical and Statistical Psychology*, *48*, 9–27.

Düntsch, I., & Gediga, G. (1995). Skills and knowledge structures. *British Journal of Mathematical and Statistical Psychology*, *48*, 9-27.

Düntsch, I., & Gediga, G. (1996). On query procedures to build knowledge structures. *Journal of Mathematical Psychology*, *40*(2), 160–168.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1–26.

Falmagne, J.-C., Albert, D., Doble, C., Eppstein, D., & Hu, X. (2013). *Knowledge spaces: Applications in education*. Springer Science & Business.

Falmagne, J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, *41*, 1-23.

Falmagne, J.-C., & Doignon, J.-P. (1988b). A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*,

32, 232-258.

Falmagne, J.-C., & Doignon, J.-P. (2011). *Learning spaces.* New York: Springer.

Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johanessen, L. (1990). Introduction to knowledge spaces: how to build, test and search them. *Psychological Review*, *97*, 204–224.

Gediga, G., & Düntsch, I. (2002). Skill set analysis in knowledge structures. *British Journal of Mathematical and Statistical Psychology*, *55*, 361-384.

Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political Research Quarterly*, *52*(3), 647–674.

Glas, C., & Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educational and Psychological Measurement*, *68*(6), 907–922.

Glas, C., & Verhelst, N. (1995). Testing the Rasch model. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications.* New York: Springer.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215–231.

Heller, J., & Repitsch, C. (2012). Exploiting prior information in stochastic knowledge assessment. *Methodology*, *8*, 12-22.

Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2014). Cognitive diagnostic models and knowledge space theory: The non-missing link. *Manuscript under revision.*

Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (under revision). On the link between cognitive diagnostic models and knowledge space theroy. *Psychometrika.*

Heller, J., Ünlü, A., & Albert, D. (2013). Skills, competencies and knowledge structures. In *Knowledge spaces* (pp. 229–242). Springer.

Heller, J., & Wickelmaier, F. (2011). Parameter estimation in probabilistic

knowledge structures. *Manuscript in preparation*.

Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, *42*(4), 49–56.

Hockemeyer, C. (2002). A comparison of non-deterministic procedures for the adaptive assessment of knowledge. *Psychologische Beiträg*, *44*, 495–503.

Holman, R., & Glas, C. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology*, *58*(1), 1–17.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.

Kambouri, M., Koppen, M., Villano, M., & Falmagne, J.-C. (1994). Knowledge assessment: Tapping human expertise by the query routine. *International Journal of Human-Computer Studies*, *40*(1), 119–151.

Kelvin, W. T. (1889). *Popular lectures and adresses (in 3 volumes). (vol. 1: Constitution of matter, chater electrical units of measurement.)*. London: MacMillan.

Koppen, M., & Doignon, J.-P. (1990). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, *34*(3), 311–331.

Korossy, K. (1993). *Modellierung von wissen als kompetenz und performance. eine erweiterung der wissensstruktur-theorie von doignon und falmagne* [Modelling knowledge as competence and performance. an extension of the theory of knowledge structures by doignon and falmagne]. Unpublished doctoral dissertation, University of Heidelberg, Heidelberg.

Korossy, K. (1997). Extending the theory of knowledge spaces: A competence-performance approach. *Zeitschrift für Psychologie*, *205*, 53-82.

Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, applications* (p. 103-132). Mahwah, NJ: Lawrence Erlbaum Associates.

Krantz, D., Luce, D., Suppes, P., & Tversky, A. (1971). Foundations of measurement: Vol. i.

Lehmann, E. L. (1999). *Elements of large-sample theory*. New York: Springer-Verlag.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation, second edition*. New York: Springer-Verlag.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

Mislevy, R. J., & Chang, H.-H. (2000). Does adaptive testing violate local independence? *Psychometrika*, *65*(2), 149–156.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, *5*(2), 241.

Pastore, M. (2009). I limiti dell'approccio nhst e l'alternativa bayesiana. *Giornale italiano di Psicologia*, *36*, 925–938.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing (3rd ed.)*. New York: Cambridge University Press.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Schrepp, M. (1999). Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, *52*, 213-224.

Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire. *Methods of Psychological Research Online*, *19*, 43–79.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Spoto, A., Stefanutti, L., & Vidotto, G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *Journal of Mathematical Psychology*, *56*(4), 248-255.

Stefanutti, L., & de Chiusole, D. (2014). Effective skill assessment in competence based knowledge sace theory. *Manuscript under revision*.

Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing the local identifiability of probabilistic knowledge structures. *Behaviour Research Methods*, *44*, 1197-1211.

Stefanutti, L., & Koppen, M. (2003). A procedure for the incremental construction of a knowledge space. *Journal of Mathematical Psychology*, *47*(3), 265–277.

Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*, 83-96.

Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, *51*, 337-350.

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (p. 453-488). Hillsdale: Lawrence Erlbaum Associates.

Vygotsky, L. S. (1978). *Mind in society: The development of higher*

psychological processes. Harvard University Press, Cambridge, MA.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic bulletin & review*, *14*(5), 779–804.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*, 41–61.