

ĐẠI HỌC SƯ PHẠM HÀ NỘI
KHOA CÔNG NGHỆ THÔNG TIN



ĐỖ ĐỨC QUANG PHƯƠNG

**CÁC PHƯƠNG PHÁP HỌC MÁY TRONG LÝ THUYẾT
HỒI ĐÁP CÂU HỎI IRT**

KHÓA LUẬN TỐT NGHIỆP

Ngành: Công nghệ thông tin

HÀ NỘI, 05 - 2025

ĐẠI HỌC SƯ PHẠM HÀ NỘI
KHOA CÔNG NGHỆ THÔNG TIN



**CÁC PHƯƠNG PHÁP HỌC MÁY TRONG LÝ THUYẾT
HỒI ĐÁP CÂU HỎI IRT**

KHÓA LUẬN TỐT NGHIỆP

Ngành: Công nghệ thông tin

Cán bộ hướng dẫn: PGS. TS. Phạm Thọ Hoàn

Sinh viên thực hiện: Đỗ Đức Quang Phương

Mã sinh viên: 715105184

HÀ NỘI, 05 - 2025

LỜI CẢM ƠN

Trước tiên, tôi xin bày tỏ lòng biết ơn sâu sắc đến PGS. TS. Phạm Thọ Hoàn, người đã tận tình hướng dẫn, góp ý và giúp tôi hoàn thiện khóa luận này. Những kiến thức quý báu, sự động viên và tinh thần trách nhiệm của thầy đã giúp tôi vượt qua những khó khăn trong quá trình nghiên cứu.

Tôi cũng xin chân thành cảm ơn Khoa Công Nghệ Thông Tin, Trường Đại Học Sư Phạm Hà Nội cùng các giảng viên đã trang bị cho tôi nền tảng kiến thức và kỹ năng để thực hiện đề tài này.

Cuối cùng, tôi xin gửi lời cảm ơn đến gia đình, bạn bè đã luôn đồng hành, động viên tôi trong suốt quá trình học tập và hoàn thành khóa luận.

Mặc dù đã cố gắng, khóa luận không tránh khỏi những thiếu sót. Tôi rất mong nhận được sự góp ý từ quý thầy cô.

Hà Nội, ngày tháng 05 năm 2025

Tác giả luận văn

Đỗ Đức Quang Phương

LỜI CAM ĐOAN

Tôi xin cam đoan bản luận văn này là kết quả do sự tìm hiểu và nghiên cứu của bản thân của tôi dưới sự hướng dẫn của PGS. TS. Phạm Thọ Hoàn. Tất cả số liệu, kết quả nghiên cứu và nội dung trình bày trong luận văn là trung thực. Mọi trích dẫn, tham khảo từ các nguồn tài liệu đều được ghi rõ nguồn gốc.

Tôi hoàn toàn chịu trách nhiệm về tính chính xác và tính trung thực của những nội dung cam kết trên.

Hà Nội, ngày tháng 05 năm 2025

Tác giả luận văn

Đỗ Đức Quang Phương

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Hà Nội, ngày tháng 05 năm 2025

Giảng viên hướng dẫn

Phạm Thọ Hoàn

NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Hà Nội, ngày tháng 05 năm 2025

Giảng viên phản biện

Nguyễn Thị Thanh Huyền

MỤC LỤC

DANH MỤC HÌNH VẼ	iv
DANH SÁCH BẢNG BIỂU	v
DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT	vi
Mở đầu	1
1. Lý do chọn đề tài	1
2. Mục đích và nhiệm vụ nghiên cứu	3
3. Đối tượng và phạm vi nghiên cứu	4
4. Phương pháp nghiên cứu	4
5. Cấu trúc của luận văn	4
1 Tổng quan về lý thuyết hồi đáp câu hỏi IRT	5
1.1 Các quan niệm về IRT	5
1.2 Các giả định trong IRT	7
1.2.1 Tính đơn chiều	7
1.2.2 Tính độc lập cục bộ	8
1.3 Đường cong đặc trưng câu hỏi	9
1.4 Các mô hình IRT	11
1.4.1 Mô hình IRT 1 tham số	11
1.4.2 Mô hình IRT 2 tham số	15
1.4.3 Mô hình IRT 3 tham số	17
2 Các phương pháp ước lượng tham số trong mô hình IRT	20
2.1 Phương pháp hợp lý cực đại	21
2.1.1 Ước lượng năng lực của thí sinh	23
2.1.2 Ước lượng các tham số của câu hỏi	34
2.1.3 Ước lượng đồng thời tham số câu hỏi và năng lực	55
2.2 Phương pháp ước lượng hợp lý cực đại biên	57
2.2.1 Giải pháp của Bock và Lieberman	59
2.2.2 Giải pháp của Bock và Aitkin	64

2.3	Phương pháp ước lượng hậu nghiệm cực đại	69
2.3.1	Ước lượng năng lực của thí sinh	71
2.3.2	Ước lượng các tham số của câu hỏi	76
2.3.3	Cách chọn các tham số trong phân phối tiên nghiệm	87
2.4	Ước lượng kỳ vọng hậu nghiệm	89
3	Thực nghiệm và đánh giá kết quả	91
3.1	Phương pháp MLE và MAP	91
3.1.1	Dữ liệu thực nghiệm	91
3.1.2	Ước lượng các tham số của câu hỏi	94
3.1.3	Ước lượng năng lực của thí sinh	100
3.2	Phương pháp MMLE và EAP	101
3.2.1	Mô tả dữ liệu	101
3.2.2	Ước lượng các tham số trong mô hình	102
	Kết luận và hướng phát triển trong tương lai	108
	Tài liệu tham khảo	109

DANH MỤC HÌNH VẼ

Hình 1.1	Đường đặc trưng câu hỏi	10
Hình 1.2	Họ đường cong đặc trưng câu hỏi một tham số với các tham số b khác nhau.	13
Hình 1.3	Đường cong đặc trưng câu hỏi hai tham số với $b = 0$ và các giá trị a khác nhau.	16
Hình 1.4	Họ các đường cong đặc trưng câu hỏi ba tham số với $a = 1.0, b = 0.0$ và các giá trị c khác nhau.	19
Hình 3.1	Dữ liệu phản hồi của 1 thí sinh với nhiều câu hỏi	94
Hình 3.2	Tỷ lệ quan sát trả lời đúng tại mỗi mức năng lực theo mô hình 1pl	94
Hình 3.3	Đường cong ĐTCH 1 tham số với tỷ lệ quan sát được dùng MLE	95
Hình 3.4	Tỷ lệ quan sát trả lời đúng tại mỗi mức năng lực theo mô hình 2pl	95
Hình 3.5	Đường cong ĐTCH 2 tham số với tỷ lệ quan sát được dùng MLE	96
Hình 3.6	Tỷ lệ quan sát trả lời đúng tại mỗi mức năng lực theo mô hình 3pl	97
Hình 3.7	Đường cong ĐTCH 3 tham số với tỷ lệ quan sát được dùng MLE	97
Hình 3.8	Đường cong ĐTCH 1 tham số với tỷ lệ quan sát được dùng MAP	98
Hình 3.9	Đường cong ĐTCH 2 tham số với tỷ lệ quan sát được dùng MAP	99
Hình 3.10	Dữ liệu phản hồi của thí sinh khi làm bài kiểm tra trong tập dữ liệu Dicht_Data2.csv	101
Hình 3.11	Biểu đồ Histogram cho tham số độ khó của mô hình IRT 1 tham số sau khi ước lượng bằng MMLE	102
Hình 3.12	Biểu đồ Histogram cho năng lực thí sinh của mô hình IRT 1 tham số sau khi ước lượng bằng EAP	103

Hình 3.13	Biểu đồ Histogram cho tham số độ khó của mô hình IRT 2 tham số sau khi ước lượng bằng MMLE	104
Hình 3.14	Biểu đồ Histogram cho tham số độ phân biệt của mô hình IRT 2 tham số sau khi ước lượng bằng MMLE	105
Hình 3.15	Biểu đồ Histogram cho năng lực thí sinh của mô hình IRT 2 tham số sau khi ước lượng bằng EAP	105
Hình 3.16	Biểu đồ Histogram cho tham số độ khó của mô hình IRT 3 tham số sau khi ước lượng bằng MMLE	106
Hình 3.17	Biểu đồ Histogram cho tham số độ phân biệt của mô hình IRT 3 tham số sau khi ước lượng bằng MMLE	107
Hình 3.18	Biểu đồ Histogram cho năng lực thí sinh của mô hình IRT 3 tham số sau khi ước lượng bằng EAP	107

DANH MỤC BẢNG BIỂU

Bảng 2.1	Bảng giá trị của phân phối χ -bình phương [22]	38
Bảng 2.2	Bảng giá trị X_k và $A(X_k)$ [19]	69
Bảng 2.3	Bảng phân vị của phân phối chuẩn tắc[22]	88
Bảng 3.1	Dữ liệu ước lượng đặc trưng câu hỏi cho mô hình IRT 1PL	92
Bảng 3.2	Dữ liệu ước lượng đặc trưng câu hỏi cho mô hình IRT 2PL	93
Bảng 3.3	Dữ liệu ước lượng đặc trưng câu hỏi cho mô hình IRT 3PL	93
Bảng 3.4	So sánh kết quả ước lượng năng lực theo các mô hình với phương pháp MLE và MAP	100
Bảng 3.5	Tham số độ khó trong mô hình 1pl sau khi ước lượng bằng MMLE	102
Bảng 3.6	Các tham số trong mô hình IRT 2 tham số sau khi ước lượng bằng MMLE	104
Bảng 3.7	Các tham số độ phân biệt và độ khó trong mô hình IRT 3 tham số sau khi ước lượng bằng MMLE	106

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Viết tắt	Ý nghĩa
IRT	Lý thuyết hồi đáp câu hỏi
1PL	Mô hình IRT một tham số
2PL	Mô hình IRT hai tham số
3PL	Mô hình IRT ba tham số
MLE	Ước lượng hợp lý tối đa
MAP	Ước lượng hậu nghiệm cực đại
MMLE	Ước lượng hợp lý tối đa biên
ICC	Đường cong đặc trưng câu hỏi
ICF	Hàm đặc trưng câu hỏi

MỞ ĐẦU

1. Lý do chọn đề tài

Sau gần 12 năm thực hiện nghị quyết số 29-NQ/TW tại hội nghị trung ương 8 khóa XI “*về đổi mới căn bản, toàn diện giáo dục và đào tạo*”[4]. Một trong những nhiệm vụ, giải quyết của nghị quyết là: “*Tiếp tục đổi mới mạnh mẽ và đồng bộ các yếu tố cơ bản của giáo dục, đào tạo theo hướng coi trọng phát triển phẩm chất, năng lực của người học*”[4]. Do đó ở đây phải chú trọng năng lực của người học.

Theo chương trình giáo dục phổ 2018 năng lực được định nghĩa như sau: “*Năng lực là thuộc tính cá nhân được hình thành, phát triển nhờ tố chất sẵn có và quá trình học tập, rèn luyện, cho phép con người huy động tổng hợp các kiến thức, kỹ năng và các thuộc tính cá nhân khác như hứng thú, niềm tin, ý chí,... thực hiện thành công một loại hoạt động nhất định, đạt kết quả mong muốn trong những điều kiện cụ thể*”[14].

Để đánh giá kết quả học tập theo định hướng tiếp cận năng lực theo Nghị định 29-NQ/TW và phù hợp với khái niệm năng lực, có nhiều phương pháp và công cụ đánh giá khác nhau. Những phương pháp này có thể được chia thành 11 nhóm chính: “(1) Đặt câu hỏi; (2) Đối thoại trên lớp; (3) Phản hồi thường xuyên; (4) Phản ánh; (5) Đánh giá đồng đẳng và tự đánh giá; (6) Sử dụng thang năng lực; (7) Sử dụng bảng kiểm danh sách các hành vi; (8) Đánh giá tình huống; (9) Phương pháp trắc nghiệm; (10) Hồ sơ học tập; (11) Đánh giá thực tiễn”[13]. Trong đó phương pháp trắc nghiệm được sử dụng phổ biến.

Hình thức thi trắc nghiệm khách quan đã trở nên phổ biến trong hệ thống giáo dục Việt Nam, từ cấp phổ thông đến đại học trong những năm gần đây. Điều này được thể hiện rõ qua việc Bộ Giáo dục và Đào tạo áp dụng hình thức thi trắc nghiệm trong kỳ thi trung học phổ thông Quốc gia từ năm 2017. Theo đó, các môn thi như toán, tiếng Anh, khoa học tự nhiên (vật lý, hóa học, sinh học) và khoa học xã hội (lịch sử, địa lý, giáo dục công dân) đều được chuyển sang hình thức trắc nghiệm [9]. Việc áp dụng này nhằm mục đích nâng cao tính khách quan, minh bạch trong đánh giá, đồng thời giảm thiểu sai sót do yếu tố chủ quan từ người chấm thi.

Ở cấp đại học, hình thức thi trắc nghiệm cũng được sử dụng rộng rãi trong các kỳ thi tuyển sinh và đánh giá kết quả học tập. Nhiều trường đại học lớn như Đại học Quốc gia Hà Nội, Đại học Quốc gia Thành phố Hồ Chí Minh, Trường Đại học Sư Phạm Hà Nội,... đã áp dụng hình thức này để đánh giá năng lực sinh viên một cách nhanh chóng và hiệu quả.

Hiện nay, nhiều trường đại học ở Việt Nam tổ chức các kỳ thi đầu vào riêng để xét tuyển sinh viên, bên cạnh việc sử dụng kết quả từ kỳ thi Trung học Phổ thông Quốc gia. Tiêu biểu như kỳ thi đánh giá tư duy - TSA của Đại học Bách khoa Hà Nội, kỳ thi đánh giá năng lực - HSA của Đại học Quốc gia Hà Nội, kỳ thi đánh giá năng lực - SPT của Trường Đại học Sư Phạm Hà Nội.

Tuy nhiên, việc thiết kế và sử dụng các câu hỏi trắc nghiệm và một số dạng câu hỏi khác để xây dựng ngân hàng câu hỏi đòi hỏi phải tuân thủ các nguyên tắc khoa học, đảm bảo tính chính xác và độ tin cậy của bài thi. Điều này đặt ra yêu cầu cao đối với công tác biên soạn đề thi và phân tích kết quả. Một trong những thách thức lớn nhất là việc đảm bảo chất lượng câu hỏi trắc nghiệm, đặc biệt là khả năng phân loại thí sinh và đo lường chính xác năng lực của người học. Để giải quyết vấn đề này, các mô hình đo lường giáo dục tiên tiến như lý thuyết hồi đáp câu hỏi (Item Response Theory - IRT) đang được nghiên cứu và ứng dụng ngày càng nhiều trong việc phân tích và cải thiện chất lượng đề thi trắc nghiệm.

Trên thế giới, việc nghiên cứu và áp dụng irt vào trong nhiều lĩnh vực đặc biệt là trong giáo dục. Có thể kể đến như cuốn sách “Probabilistic Models for Some Intelligence and Attainment Tests” của Georg Rasch là nền tảng cho sự phát triển của IRT, giới thiệu mô hình Rasch - một trong những mô hình IRT đơn giản và phổ biến nhất [16]. Quyển “Applications of Item Response Theory to Practical Testing Problems” được xuất bản lần đầu tiên vào năm 1980 của tác giả Lord [25] cung cấp cái nhìn tổng quan về lý thuyết và ứng dụng của IRT trong thực tiễn kiểm tra và đánh giá. Hay cuốn “Fundamentals of Item Response Theory” (1991) [18] được coi là “kinh thánh” về IRT cung cấp kiến thức cơ bản và nâng cao về các mô hình IRT, cách ước lượng tham số và ứng dụng trong thực tiễn. Bài báo “BILOG: Maximum Likelihood Item Analysis and Test Scoring with Binary Logistic Models”(1981) giới thiệu phần mềm BILOG, một công cụ phổ biến để ước lượng tham số IRT sử dụng phương pháp MLE [7]. Chalmers (2012) đã có bài báo “mirt: A Multidimensional Item Response Theory Package for the

R Environmen” giới thiệu gói phần mềm mirt trong R, một công cụ mạnh mẽ để ước lượng tham số IRT sử dụng các phương pháp MLE và Bayesian [10].

Có thể thấy trên thế giới nghiên cứu về IRT đã có những thành công đột phá, trong khi đó ở Việt Nam trong 2 thập niên gần đây thì IRT đang dần được quan tâm và áp dụng trong nhiều lĩnh vực, đặc biệt là giáo dục và tâm lý học, nhằm nâng cao chất lượng đánh giá năng lực người học. Nhóm tác giả Minh và Thiện (2006) có bài báo trình bày về phương pháp thi trắc nghiệm khách quan và ứng dụng mô hình Rasch trong đánh giá giáo dục, đặc biệt là phương pháp PROX để đo lường và đánh giá tự động trên máy tính [26]. Nhóm tác giả Lê Thị Bích Hằng và cộng sự năm (2022) công bố bài báo giới thiệu giải pháp đánh giá câu hỏi trắc nghiệm khách quan trên hệ thống Moodle sử dụng IRT, giúp phát hiện và chỉnh sửa các câu hỏi có vấn đề để cải thiện chất lượng ngân hàng bài kiểm tra và hiệu quả đánh giá [21]. Nhóm tác giả Thanh và Thảo (2023) đã viết bài báo trình bày việc sử dụng câu hỏi trắc nghiệm khách quan và lý thuyết IRT để đánh giá năng lực vật lý của học sinh trong chương Sóng ánh sáng Vật lý 12, với sự hỗ trợ của phần mềm Quest/Conquest nhằm phân tích độ khó, độ phân biệt và độ tin cậy của đề thi [33]. Quyển sách “Đo lường trong giáo học lý thuyết và ứng dụng” của tác giả Lâm Quang Thiệp (2010) đã cho ta một cách nhìn tổng quan về mô hình IRT 1, 2 và 3 tham số và trình bày phương pháp MLE để ước lượng tham số trong mô hình [34]. Bài báo của nhóm tác giả Bình và Duy (2016), trình bày đánh giá khả năng học sinh dựa trên mô hình irt 2 tham số, dùng phương pháp MLE và K-Means với dữ liệu kiểm tra của kỳ thi tuyển sinh vào trường THPT Chuyên - Đại học Sư phạm Hà Nội, bao gồm các câu trả lời kiểm tra tiếng Anh của 1111 thí sinh thi vào lớp 10 năm 2014 [5].

Từ những ưu điểm nổi bật của của IRT và mong muốn tìm hiểu chuyên sâu về IRT, tôi đã chọn đề tài: ***“Các phương pháp học máy trong Lý thuyết hồi đáp câu hỏi IRT”***.

2. Mục đích và nhiệm vụ nghiên cứu

Nghiên cứu tập trung vào các mô hình IRT, bao gồm mô hình 1PL, 2PL và 3PL, nhằm đo lường năng lực thí sinh qua bài kiểm tra trắc nghiệm.

Nghiên cứu áp dụng các nguyên lý lý thuyết từ các tài liệu của Baker (2001, 2017) để xây dựng quy trình ước lượng tham số câu hỏi trong các mô hình IRT

cho phương pháp MLE và MAP. Và nghiên cứu các phương pháp MMLE, EAP.

Triển khai các phương pháp ước lượng tham số gồm tự triển khai bằng Python với MLE, MAP và MMLE, EAP được thực hiện qua gói TAM trong R.

3. Đối tượng và phạm vi nghiên cứu

Các mô hình IRT nhị phân: mô hình 1 tham số, 2 tham số và 3 tham số.

Bài toán ước lượng Maximum Likelihood estimation, Maximum A Posteriori estimation, Marginal Maximum Likelihood estimation và Expected A Posteriori.

4. Phương pháp nghiên cứu

4.1. Nghiên cứu lý thuyết

Nghiên cứu lý thuyết IRT tập trung vào các mô hình IRT nhị phân, bao gồm mô hình 1PL, 2PL và 3PL, cùng với các phương pháp ước lượng tham số như MLE, MAP, MMLE và EAP.

Nghiên cứu các công thức lý thuyết và cách thức áp dụng các phương pháp này để ước lượng các tham số trong mô hình, đồng thời đánh giá các yếu tố ảnh hưởng đến độ chính xác và hiệu quả tính toán của từng phương pháp.

4.2. Thực nghiệm

Nghiên cứu thực nghiệm triển khai các phương pháp ước lượng tham số trong các mô hình IRT bằng cách sử dụng các công cụ lập trình Python và R. Tự triển khai code và các gói phần mềm trong Python được sử dụng để triển khai các phương pháp MLE và MAP, trong khi phương pháp MMLE và EAP được thực hiện qua gói TAM trong R.

Thực nghiệm được thực hiện trên các tập dữ liệu tự tạo và dữ liệu thực tế (tập dữ liệu Dicht_Data2.csv từ Kaggle), nhằm kiểm tra độ chính xác và hiệu quả tính toán của các phương pháp ước lượng.

5. Cấu trúc của luận văn

Ngoài phần mở đầu, hướng đề xuất, kết luận và mục lục, luận án gồm 3 chương:

- Chương 1: Tổng quan về lý thuyết hồi đáp câu hỏi IRT.
- Chương 2: Các phương pháp ước lượng tham số trong mô hình IRT.
- Chương 3: Thực nghiệm và đánh giá kết quả.

CHƯƠNG 1. TỔNG QUAN VỀ LÝ THUYẾT HỒI ĐÁP CÂU HỎI IRT

1.1. Các quan niệm về IRT

Lý thuyết hồi đáp câu hỏi (Item Response Theory - IRT) là một trong những lý thuyết đo lường tiên tiến trong lĩnh vực giáo dục và tâm lý học, ra đời vào nửa sau thế kỷ 20 và tiếp tục phát triển mạnh mẽ cho đến ngày nay. IRT được xây dựng nhằm khắc phục những hạn chế của Lý thuyết trắc nghiệm cổ điển (Classical Test Theory - CTT), vốn đã hình thành từ cuối thế kỷ 19 và đạt đến sự hoàn thiện vào những năm 1970. Mặc dù CTT đã có những đóng góp quan trọng trong việc đánh giá giáo dục, nó vẫn tồn tại một số điểm yếu, chẳng hạn như việc phụ thuộc vào mẫu thử nghiệm và thiếu tính linh hoạt trong việc phân tích các câu hỏi riêng lẻ.

Để giải quyết những hạn chế này, các nhà tâm trắc học đã phát triển IRT dựa trên nền tảng toán học vững chắc, sử dụng các mô hình phức tạp để ước lượng khả năng của thí sinh và đặc điểm của câu hỏi một cách độc lập. Sự phát triển của IRT đã được hỗ trợ đáng kể nhờ những tiến bộ vượt bậc trong công nghệ tính toán, đặc biệt là sự ra đời và phổ biến của máy tính điện tử vào cuối thế kỷ 20 và đầu thế kỷ 21. Nhờ đó, IRT không chỉ trở thành một công cụ mạnh mẽ trong nghiên cứu và ứng dụng thực tiễn mà còn mở ra nhiều hướng phát triển mới trong lĩnh vực đo lường giáo dục và tâm lý.

Với khả năng phân tích chi tiết và linh hoạt, IRT đã mang lại những đóng góp quan trọng trong việc thiết kế bài kiểm tra, đánh giá năng lực cá nhân, và cải thiện chất lượng đo lường trong giáo dục. Sự kết hợp giữa lý thuyết toán học và công nghệ hiện đại đã giúp IRT trở thành một trong những nền tảng đo lường tiên tiến nhất hiện nay.

Theo Embretson & Reise (2013), IRT có nguồn gốc từ hai dòng phát triển chính, tạo nền tảng cho các ứng dụng hiện đại. Dòng phát triển tại Hoa Kỳ: IRT bắt đầu được hình thành từ những năm 1950 với các nghiên cứu của Lord (1953) và Birnbaum (1957, 1958a, 1958b). Cuốn sách *Statistical Theories of Mental Test Scores* của Lord và Novick (1968) là cột mốc quan trọng, giới thiệu IRT một cách hệ thống và chặt chẽ. Các nhà nghiên cứu như R. Darrell Bock và các cộng sự đã phát triển các thuật toán và phần mềm (như BILOG, MULTILOG) để ước lượng

tham số IRT, đặc biệt là phương pháp ước lượng khả năng cực đại biên. Dòng phát triển tại châu Âu: Georg Rasch (1960) là người tiên phong với mô hình Rasch, tập trung vào tính khách quan cụ thể và khả năng tách biệt tham số người làm bài và câu hỏi. Rasch truyền cảm hứng cho các nhà nghiên cứu như Gerhard Fischer và Benjamin Wright, những người đã mở rộng và phổ biến mô hình Rasch trong giáo dục và ứng dụng thực tiễn [15].

Theo Lord(1952), IRT là một phương pháp thống kê tiên tiến dùng để đo lường năng lực hoặc các đặc điểm tâm lý của thí sinh thông qua phản ứng của họ với các câu hỏi trong bài kiểm tra. Khác với cách tính điểm truyền thống sử dụng tổng điểm các câu hỏi, IRT hướng đến việc xác định một thước đo năng lực bất biến, không phụ thuộc vào đặc điểm cụ thể của bài kiểm tra. IRT sử dụng các mô hình toán học như hàm logistic hoặc ogive chuẩn để mô tả mối quan hệ giữa xác suất trả lời đúng và năng lực của thí sinh, đồng thời giải quyết các vấn đề như sai số đo lường, độ khó câu hỏi và khả năng phân biệt của bài kiểm tra. Nhờ đó, IRT cung cấp một công cụ chính xác và linh hoạt để đánh giá năng lực trong giáo dục và tâm lý học [24].

Theo Hambleton & Jones (1993), IRT là một lý thuyết thống kê tiên tiến nhằm phân tích hiệu suất của thí sinh trên từng câu hỏi và toàn bộ bài kiểm tra, đồng thời xem xét mối quan hệ giữa hiệu suất đó với năng lực được đo lường thông qua các câu hỏi trong bài kiểm tra [17].

Theo Wu & Adams (2007), trọng tâm của IRT là sử dụng các mô hình toán học để dự đoán xác suất trả lời đúng một câu hỏi dựa trên hai yếu tố chính: năng lực của thí sinh và độ khó của câu hỏi. Xác suất này được biểu diễn thông qua một đường cong đặc trưng gọi là đường cong đặc trưng câu hỏi (Item Characteristic Curve - ICC). Đường cong này thể hiện mối quan hệ giữa năng lực của thí sinh và xác suất trả lời đúng, giúp phân tích và đánh giá hiệu quả của từng câu hỏi trong bài thi [20].

Từ các quan điểm trên thì IRT là một phương pháp thống kê tiên tiến, sử dụng các mô hình toán học để đo lường năng lực hoặc đặc điểm tâm lý của thí sinh thông qua phản ứng với các câu hỏi. Khác với cách tính điểm truyền thống, IRT hướng đến việc tạo ra một thước đo năng lực bất biến, không phụ thuộc vào đặc điểm cụ thể của bài kiểm tra. Thông qua đường cong đặc trưng câu hỏi, IRT

mô tả mối quan hệ giữa năng lực của thí sinh và xác suất trả lời đúng, từ đó giải quyết các vấn đề như sai số đo lường, độ khó câu hỏi và khả năng phân biệt. Nhờ vậy, IRT trở thành công cụ chính xác và linh hoạt trong đánh giá giáo dục và tâm lý học.

1.2. Các giả định trong IRT

Theo Hambleton, Swaminathan & Rogers (1991) [18], các mô hình trong IRT dựa trên một số giả định quan trọng. Đầu tiên, IRT giả định rằng chỉ có một năng lực được đo lường bởi các câu hỏi trong bài kiểm tra. Điều này liên quan đến tính độc lập cục bộ, nghĩa là phản ứng của thí sinh với một câu hỏi không ảnh hưởng đến phản ứng với các câu hỏi khác. Thứ hai, IRT giả định rằng hàm đặc trưng câu hỏi phản ánh chính xác mối quan hệ giữa năng lực tiềm ẩn và phản ứng quan sát được. Cuối cùng, các mô hình IRT khác nhau dựa trên số lượng và loại đặc điểm câu hỏi được giả định là ảnh hưởng đến hiệu suất của thí sinh. Các giả định này cần được kiểm tra thông qua đánh giá mức độ phù hợp của mô hình với dữ liệu thực tế.

1.2.1. Tính đơn chiều

Như đã cập ở trên, một giả định cơ bản trong các mô hình IRT là chỉ có một năng lực chính được đo lường bởi các câu hỏi trong bài kiểm tra. Tuy nhiên, trên thực tế, hiệu suất làm bài của thí sinh không chỉ phụ thuộc vào năng lực này mà còn chịu ảnh hưởng bởi nhiều yếu tố khác như động lực, lo lắng, tốc độ làm bài, xu hướng đoán câu trả lời, và các kỹ năng nhận thức khác. Do đó, giả định về tính đơn chiều không thể được đáp ứng một cách tuyệt đối. Thay vào đó, để giả định này được coi là hợp lý, cần có một yếu tố chi phối đóng vai trò chính trong việc giải thích hiệu suất làm bài. Yếu tố này được gọi là năng lực tiềm ẩn được đo bởi bài kiểm tra, mặc dù năng lực này không nhất thiết phải cố định mà có thể thay đổi theo thời gian do học tập, quên lãng, hoặc các yếu tố khác.

Các mô hình IRT giả định rằng chỉ một năng lực chi phối đủ để giải thích hiệu suất của thí sinh được gọi là mô hình đơn chiều. Trong khi đó, các mô hình yêu cầu nhiều hơn một năng lực để giải thích hiệu suất làm bài được gọi là mô hình đa chiều. Các mô hình đa chiều thường phức tạp hơn và hiện vẫn chưa được phát triển đầy đủ [18].

1.2.2. Tính độc lập cục bộ

Tính độc lập cục bộ có nghĩa là khi các năng lực ảnh hưởng đến hiệu suất làm bài được giữ cố định, phản ứng của thí sinh đối với bất kỳ cặp câu hỏi nào là độc lập về mặt thống kê. Nói cách khác, sau khi đã tính đến năng lực của thí sinh, không có mối quan hệ nào tồn tại giữa phản ứng của thí sinh đối với các câu hỏi khác nhau. Đơn giản hơn, điều này có nghĩa là các năng lực được xác định trong mô hình là các yếu tố duy nhất ảnh hưởng đến phản ứng của thí sinh đối với các câu hỏi. Tập hợp các năng lực này đại diện cho không gian tiềm ẩn đầy đủ. Khi giả định về tính đơn chiều được đáp ứng, không gian tiềm ẩn đầy đủ chỉ bao gồm một năng lực duy nhất [18].

Tính độc lập cục bộ còn được gọi là tính độc lập có điều kiện. Điều này có nghĩa là khi các yếu tố chung (như năng lực) được loại bỏ hoặc giữ cố định, các phản ứng của thí sinh đối với các câu hỏi trở nên độc lập về mặt thống kê. Trong IRT, tính độc lập cục bộ đảm bảo rằng các phản ứng của thí sinh chỉ phụ thuộc vào năng lực được đo lường và không bị ảnh hưởng bởi các yếu tố khác [18].

Khi giả định về tính đơn chiều được đáp ứng, tức là chỉ có một năng lực duy nhất ảnh hưởng đến hiệu suất làm bài, tính độc lập cục bộ sẽ tự động đạt được. Trong trường hợp này, tất cả các phản ứng của thí sinh đối với các câu hỏi đều được giải thích bởi một năng lực chung. Tuy nhiên, tính độc lập cục bộ cũng có thể đạt được trong các mô hình đa chiều (multidimensional models), miễn là tất cả các chiều năng lực ảnh hưởng đến hiệu suất làm bài đã được xác định và tính đến. Nếu bất kỳ chiều năng lực nào bị bỏ sót, tính độc lập cục bộ sẽ không còn được đảm bảo [18].

Để minh họa rõ hơn về tính độc lập cục bộ và mối quan hệ của nó với tính đơn chiều, chúng ta có thể xem xét các tình huống thực tế trong các bài kiểm tra khác nhau. Những ví dụ dưới đây sẽ cho thấy cách các yếu tố ngoài năng lực chính có thể ảnh hưởng đến phản ứng của thí sinh, từ đó làm nổi bật tầm quan trọng của việc xác định đầy đủ các chiều năng lực trong mô hình IRT:

- Ví Dụ 1: Bài kiểm tra toán học yêu cầu kỹ năng đọc hiểu. Trong một bài kiểm tra toán học, giả sử có một câu hỏi yêu cầu thí sinh đọc và hiểu một đoạn văn dài trước khi giải quyết vấn đề toán học. Những thí sinh có kỹ năng đọc hiểu kém sẽ gặp khó khăn trong việc hiểu đề bài, dẫn đến việc

không thể trả lời đúng câu hỏi, ngay cả khi họ có năng lực toán học tốt. Trong trường hợp này, kỹ năng đọc hiểu trở thành một chiều năng lực khác ngoài năng lực toán học. Nếu chỉ sử dụng mô hình đơn chiều (chỉ đo lường năng lực toán học), tính độc lập cục bộ sẽ bị vi phạm vì phản ứng của thí sinh không chỉ phụ thuộc vào năng lực toán học mà còn bị ảnh hưởng bởi kỹ năng đọc.

- Ví Dụ 2: Câu hỏi chứa manh mối trong bài kiểm tra tiếng Anh. Trong một bài kiểm tra tiếng Anh, giả sử câu hỏi A chứa một từ vựng hoặc cấu trúc ngữ pháp giúp thí sinh trả lời câu hỏi B. Những thí sinh phát hiện ra manh mối này sẽ có lợi thế hơn so với những thí sinh không nhận ra. Điều này làm cho phản ứng của thí sinh đối với câu hỏi B không chỉ phụ thuộc vào năng lực tiếng Anh mà còn bị ảnh hưởng bởi khả năng phát hiện manh mối. Nếu chỉ sử dụng mô hình đơn chiều (chỉ đo lường năng lực tiếng Anh), tính độc lập cục bộ sẽ bị vi phạm vì phản ứng của thí sinh còn phụ thuộc vào một yếu tố khác ngoài năng lực chính.
- Ví Dụ 3: Bài kiểm tra khoa học yêu cầu tư duy logic. Trong một bài kiểm tra khoa học, giả sử có một câu hỏi yêu cầu thí sinh áp dụng tư duy logic để giải quyết vấn đề. Những thí sinh có kỹ năng tư duy logic kém sẽ không thể trả lời đúng câu hỏi, ngay cả khi họ có kiến thức khoa học tốt. Trong trường hợp này, kỹ năng tư duy logic trở thành một chiều năng lực khác ngoài kiến thức khoa học. Nếu chỉ sử dụng mô hình đơn chiều (chỉ đo lường kiến thức khoa học), tính độc lập cục bộ sẽ bị vi phạm vì phản ứng của thí sinh không chỉ phụ thuộc vào kiến thức khoa học mà còn bị ảnh hưởng bởi kỹ năng tư duy logic.

Các ví dụ trên cho thấy rằng, trong thực tế, các bài kiểm tra thường yêu cầu nhiều hơn một năng lực để giải quyết các câu hỏi. Nếu chỉ sử dụng mô hình đơn chiều, tính độc lập cục bộ có thể bị vi phạm do các yếu tố ngoài năng lực chính ảnh hưởng đến phản ứng của thí sinh. Do đó, việc xác định và tính đến tất cả các chiều năng lực ảnh hưởng là rất quan trọng để đảm bảo tính hợp lệ của mô hình IRT.

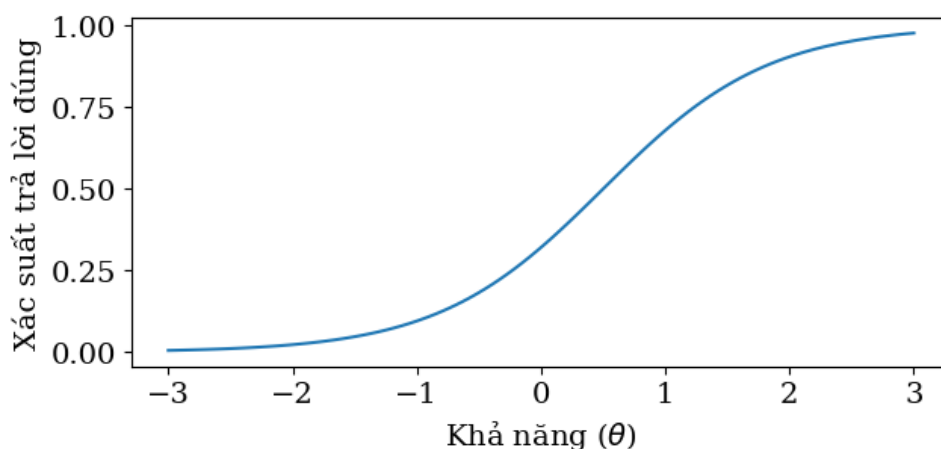
1.3. Đường cong đặc trưng câu hỏi

Sau khi đã xem xét các giả định cơ bản của IRT, bao gồm tính đơn chiều và tính độc lập cục bộ, chúng ta chuyển sang một khái niệm trung tâm khác của IRT:

Đường cong đặc trưng câu hỏi (Item Characteristic Curve - ICC). ICC là công cụ chính để mô tả mối quan hệ giữa năng lực của thí sinh và xác suất trả lời đúng một câu hỏi. Đây là nền tảng để hiểu cách IRT đo lường và phân tích các đặc điểm tiềm ẩn.

Trong đo lường giáo dục và tâm lý, các đặc điểm tiềm ẩn như trí thông minh hay năng lực học tập không thể quan sát trực tiếp mà phải đo lường gián tiếp. Để làm điều này, năng lực được biểu diễn trên một trục liên tục từ $-\infty$ đến $+\infty$, với điểm trung bình của phân bố năng lực được chọn làm điểm 0 và độ lệch chuẩn ($\sigma = 1$) làm đơn vị đo. Trong IRT, thang đo năng lực thường có phạm vi từ -3 đến +3 để thuận tiện cho việc phân tích, mặc dù giá trị lý thuyết có thể vượt ra ngoài khoảng này. Cách tiếp cận này giúp so sánh và đánh giá năng lực thí sinh một cách chính xác và nhất quán [3].

ICC là một đường cong mô tả mối quan hệ giữa năng lực của thí sinh θ và xác suất trả lời đúng một câu hỏi. Đường cong này có dạng tăng đơn điệu, nghĩa là khi năng lực của thí sinh tăng, xác suất trả lời đúng cũng tăng theo. ICC thường được biểu diễn bằng các hàm toán học như hàm logistic hoặc hàm ogive chuẩn. Mỗi thí sinh có một mức năng lực tiềm ẩn, được biểu diễn bằng giá trị θ trên thang đo năng lực. Ở mỗi mức năng lực, xác suất trả lời đúng $P(\theta)$ thấp đối với thí sinh có năng lực thấp và cao đối với thí sinh có năng lực cao. Khi vẽ $P(\theta)$ theo năng lực, ta thu được một đường cong hình chữ S. Mỗi câu hỏi trong bài kiểm tra có một ICC riêng và đây là nền tảng cơ bản của IRT, giúp phân tích và đánh giá hiệu quả của từng câu hỏi.



Hình 1.1: Đường đặc trưng câu hỏi

Để mô tả ICC, có hai đặc tính kỹ thuật quan trọng:

- Độ khó: Mô tả vị trí câu hỏi trên thang đo khả năng, từ câu hỏi dễ (dành cho người có khả năng thấp) đến câu hỏi khó (dành cho người có khả năng cao).
- Khả năng phân biệt: Phản ánh khả năng của câu hỏi trong việc phân biệt giữa người có khả năng thấp và cao. Câu hỏi càng dốc, khả năng phân biệt càng tốt.

Hai thuộc tính này chỉ mô tả hình dạng của ICC và không liên quan đến việc câu hỏi có thực sự đo lường năng lực tiềm ẩn hay không. Đây là vấn đề về độ giá trị, cần được xem xét riêng [3].

1.4. Các mô hình IRT

Lý thuyết hồi đáp câu hỏi dựa trên hai tiên đề cơ bản:

- Xác suất trả lời đúng của một thí sinh trên một câu hỏi được dự đoán bởi các năng lực tiềm ẩn.
- Mỗi quan hệ giữa xác suất trả lời đúng và năng lực tiềm ẩn được mô tả bằng hàm đặc trưng câu hỏi (Item Characteristic Function- ICF), một hàm tăng đơn điệu: càng có năng lực cao, xác suất trả lời đúng càng tăng [18].

1.4.1. Mô hình IRT 1 tham số

Từ giả thuyết trên, mô hình được công bố lần đầu tiên bởi nhà toán học người Đan Mạch Georg Rasch vào những năm 1960. Rasch tiếp cận việc phân tích dữ liệu kiểm tra từ góc độ lý thuyết xác suất. Mặc dù ông bắt đầu từ một khuôn khổ tham chiếu rất khác, nhưng mô hình đường cong đặc trưng câu hỏi mà ông phát triển lại là một mô hình logistic. Để phân tích mối quan hệ giữa thí sinh và câu hỏi trong quá trình ứng đáp, Rasch gán cho mỗi thí sinh một tham số năng lực, trong khi đối với mỗi câu hỏi, ông chỉ sử dụng một tham số duy nhất là độ khó độ khó của câu hỏi. Tiên đề mà Rasch đưa ra làm cơ sở cho mô hình của ông có thể được phát biểu như sau: *“Một người có năng lực cao hơn một người khác thì xác suất để người đó trả lời đúng một câu hỏi bất kỳ phải lớn hơn xác suất của người sau, cũng tương tự như vậy, một câu hỏi khó hơn một câu hỏi khác có nghĩa là xác suất để một người bất kỳ trả lời đúng câu hỏi đó phải bé hơn xác suất để trả lời đúng câu hỏi sau”* [34].

Với tiền Rasch, xác suất để trả lời đúng câu hỏi ta có thể biểu diễn như sau:

$$f(P) = \frac{\Theta}{\beta}, \quad (1.1)$$

trong đó:

- f là một hàm nào đó của xác suất trả lời đúng câu hỏi,
- P là xác suất để trả lời đúng câu hỏi của thí sinh phụ thuộc vào mối tương quan giữa Θ và β ,
- Θ biểu diễn năng lực của thí sinh,
- β biểu diễn độ khó của câu hỏi.

Để biểu diễn hợp lý hàm $f(P)$, đầu tiên Rasch đã thực hiện lấy logarit tự nhiên của (1.1), ta được:

$$\begin{aligned} \ln f(P) &= \ln \frac{\Theta}{\beta} \\ &= \ln \Theta - \ln \beta \\ &= \theta - b. \end{aligned}$$

Tiếp theo, để đơn giản hóa mô hình, khi xem xét mô hình trắc nghiệm nhị phân, Rasch đã chọn hàm f là tỷ lệ odds (mức được thua), được biểu diễn bằng công thức $[P/(1 - P)]$. Tỷ lệ này phản ánh tỷ số giữa xác suất xảy ra sự kiện khẳng định (trả lời đúng) so với xác suất xảy ra sự kiện phủ định (trả lời sai) [34]. Cụ thể, nếu P là xác suất trả lời đúng câu hỏi, thì $(1 - P)$ là xác suất trả lời sai. Khi đó, tỷ lệ odds $[P/(1 - P)]$ cho biết khả năng thí sinh trả lời đúng câu hỏi cao hơn bao nhiêu lần so với trả lời sai. Cách tiếp cận này không chỉ đơn giản mà còn mang tính trực quan, giúp mô hình Rasch trở nên dễ hiểu và dễ áp dụng trong thực tiễn trắc nghiệm. Như vậy, ta có thể biến đổi như sau:

$$\begin{aligned} \ln \frac{P}{1 - P} &= \theta - b \\ \frac{P}{1 - P} &= e^{\theta - b} \\ P &= (1 - P)e^{\theta - b} \\ (1 + e^{\theta - b})P &= e^{\theta - b} \end{aligned}$$

$$P = \frac{e^{\theta-b}}{1+e^{\theta-b}}.$$

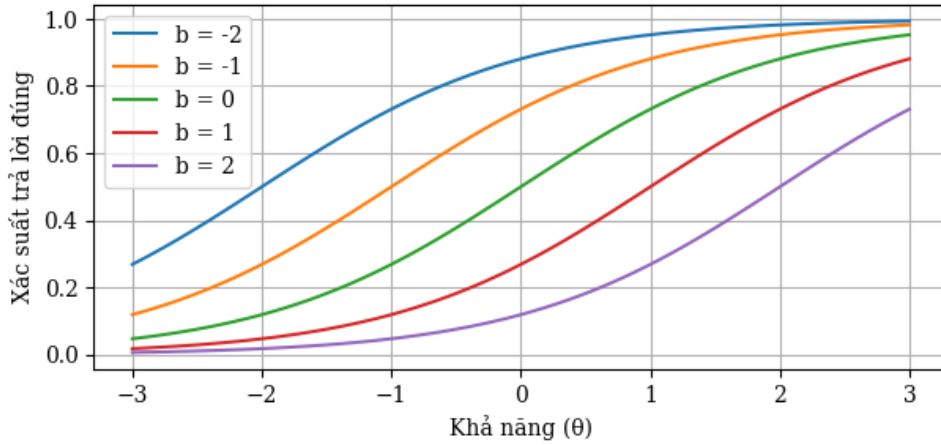
Do đó, ta thu được mô hình Rasch như sau:

$$P(U_{ij} = 1 | \theta_i, b_j) = \frac{1}{1+e^{-(\theta_i-b_j)}}, \quad (1.2)$$

trong đó:

- $P(\theta)$ là xác suất mà một thí sinh được chọn ngẫu nhiên với năng lực θ trả lời đúng câu hỏi thứ j .
- θ_i là năng lực của thí sinh thứ i ,
- b_j là độ khó của câu hỏi thứ j ,
- U_{ij} là ứng đáp của thí sinh thứ i với câu hỏi thứ j :
 - + $U_{ij} = 1$ nếu thí sinh trả lời đúng câu hỏi và
 - + $U_{ij} = 0$ khi thí sinh trả lời sai.

Biểu thức (1.2) có dạng logistic và có hình dạng như hình vẽ dưới đây:



Hình 1.2: Họ đường cong đặc trưng câu hỏi một tham số với các tham số b khác nhau.

Khi biểu diễn tất cả các đường cong đặc trưng của câu hỏi trên cùng một hệ trục với hoành độ θ (năng lực thí sinh), ta thu được một họ các đường cong có dạng giống nhau nhưng được tịnh tiến dọc theo trục θ như được biểu diễn ở hình 1.2. Mỗi đường cong có gốc tại hoành độ $\theta = b_j$, nơi mà xác suất thí sinh có năng lực $\theta_i = b_j$ trả lời đúng câu hỏi đạt 0.5. Trong mô hình Rasch, tham số b_j này

chính là độ khó của câu hỏi, hay còn gọi là ngưỡng câu hỏi, vì đó là mức năng lực tại đó xác suất trả lời đúng đạt 0.5. Khi năng lực thí sinh tăng lên vô hạn, xác suất trả lời đúng sẽ tiệm cận 1, ngược lại, khi năng lực giảm xuống rất thấp, xác suất trả lời đúng gần về 0. Một đặc điểm quan trọng của các đường cong đặc trưng của câu hỏi trong mô hình này là chúng không cắt nhau, thể hiện tính nhất quán trong cách thức câu hỏi phân biệt thí sinh dựa trên mức năng lực của họ. Baker (2001) cho rằng *“điểm nổi bật trong mô hình Rasch là nó mô tả được mối liên hệ giữa năng lực của mỗi thí sinh đối với các tham số đặc trưng của các câu hỏi thông qua sự ứng đáp của mỗi thí sinh khi trả lời các câu hỏi trong đề thi”* [11].

Khi các giá trị năng lực của một nhóm thí sinh được chuẩn hóa sao cho giá trị trung bình là 0 và độ lệch chuẩn là 1, các giá trị của tham số độ khó b_j thường nằm trong khoảng từ -3.0 đến +3.0. Cụ thể, những câu hỏi có giá trị b_j gần -3.0 được xem là rất dễ, trong khi những câu hỏi có giá trị b_j gần 3.0 được coi là rất khó đối với nhóm thí sinh đó. Theo Baker (2001), độ khó của câu hỏi có thể được phân thành 5 mức:

- Rất dễ: $b < -2.0$,
- Dễ: $-2.0 \leq b < -0.5$,
- Trung bình: $-0.5 \leq b < 0.5$,
- Khó: $0.5 \leq b < 2.0$,
- Rất khó: $b \geq 2.0$.

Những câu hỏi có giá trị độ khó quá thấp (dưới -3.0) hoặc quá cao (trên 3.0) thường không mang lại nhiều ý nghĩa trong việc đánh giá năng lực, vì chúng không đủ khả năng phân biệt giữa các thí sinh có trình độ khác nhau [2].

Và với năng lực của thí sinh, khi năng lực của thí sinh tăng lên thì xác suất trả lời đúng phải tăng lên, nên ở đây chúng tôi đề xuất chia năng lực của thí sinh có thể chia thành 5 mức sau:

- Yếu: $\theta < -2.0$,
- Trung bình: $-2.0 \leq \theta < -0.5$,
- Khá: $-0.5 \leq \theta < 0.5$,
- Giỏi: $0.5 \leq \theta < 2.0$,

- Xuất sắc: $\theta \geq 2.0$.

1.4.2. Mô hình IRT 2 tham số

Lord (1952) là người tiên phong trong việc phát triển mô hình ứng đáp câu hỏi hai tham số, dựa trên hàm phân phối chuẩn tích lũy (normal ogive). Sau đó, Birnbaum (1968) đã cải tiến mô hình này bằng cách thay thế hàm normal ogive bằng hàm logistic hai tham số, mang lại nhiều ưu điểm hơn trong tính toán và ứng dụng. Hàm logistic không chỉ đơn giản hóa các phép toán mà còn dễ dàng xử lý về mặt toán học so với hàm normal ogive, vốn đòi hỏi phải thực hiện tích phân phức tạp [18]. Mô hình logistic hai tham số của Birnbaum được biểu diễn bằng phương trình:

$$P(U_{ij} = 1 \mid \theta_i, a_j, b_j) = \frac{1}{1 + e^{-Da_j(\theta_i - b_j)}}, \quad (1.3)$$

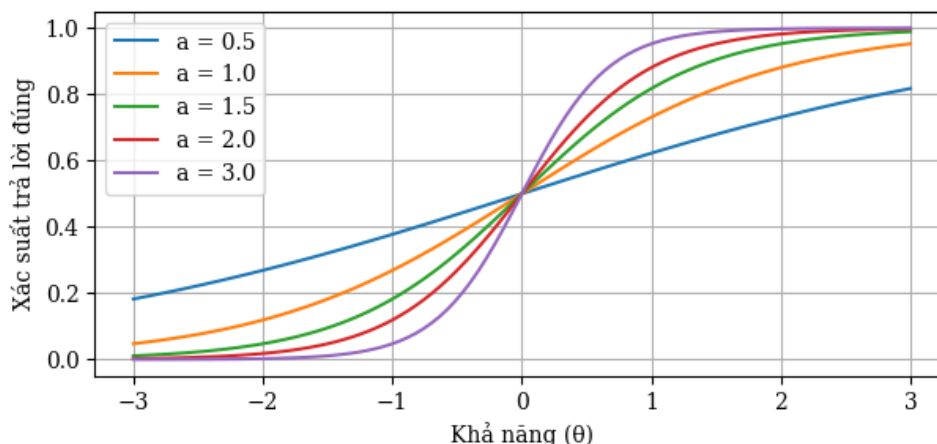
trong đó:

- $P(\theta)$: Xác suất thí sinh có năng lực θ trả lời đúng câu hỏi j ,
- D : Hệ số tỷ lệ ($D = 1.7$ hoặc $D = 1.702$), được sử dụng để làm cho hàm logistic gần giống với hàm normal ogive nhất có thể.
- θ_i là năng lực của thí sinh thứ i ,
- b_j là độ khó của câu hỏi thứ j ,
- a_j là độ phân biệt của câu hỏi thứ j , phản ánh khả năng phân biệt giữa các thí sinh có năng lực khác nhau.
- U_{ij} là ứng đáp của thí sinh thứ i với câu hỏi thứ j :
 - + $U_{ij} = 1$ nếu thí sinh trả lời đúng câu hỏi và
 - + $U_{ij} = 0$ khi thí sinh trả lời sai.

Trong phần lớn tài liệu IRT, giá trị logistic của tham số a thường được chia cho 1.7 hoặc 1.702 để tương thích với mô hình normal ogive. Tuy nhiên, ở đây, việc chia tỷ lệ không được áp dụng để tránh tạo ra hai hệ quy chiếu khác nhau và tất cả tham số đều được diễn giải dựa trên hàm logistic. Do đó, ở đây mô hình có thể viết đơn giản như sau:

$$P(U_{ij} = 1 \mid \theta_i, a_j, b_j) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}. \quad (1.4)$$

ICC của mô hình IRT 2 tham số được biểu diễn như sau:



Hình 1.3: Đường cong đặc trưng câu hỏi hai tham số với $b = 0$ và các giá trị a khác nhau.

Trong mô hình Rasch (mô hình một tham số), tham số phân biệt a luôn được cố định bằng 1, trong khi mô hình hai tham số cho phép a thay đổi để phản ánh mức độ phân biệt của từng câu hỏi. Về lý thuyết, giá trị của a nằm trong khoảng $(-\infty, +\infty)$, nhưng trong thực tế, nó thường dao động từ -2.80 đến 2.80. Theo Baker (2001), độ phân biệt a có thể được chia thành 5 mức:

- Rất kém: Khi giá trị a nhỏ hơn 0.35,
- Kém: Khi giá trị a từ 0.35 đến dưới 0.65,
- Trung bình: Khi giá trị a từ 0.65 đến dưới 1.35,
- Tốt: Khi giá trị a từ 1.35 đến dưới 1.70,
- Rất tốt: Khi giá trị a từ 1.70 trở lên [2].

Tuy nhiên, độ phân biệt của câu hỏi trong đề thi nên nằm trong khoảng từ 0.5 đến dưới 2.0. Những câu hỏi có giá trị a quá thấp hoặc quá cao thường không có ý nghĩa hoặc có ý nghĩa rất thấp trong việc đo lường và phân loại năng lực của thí sinh.

Tham số a_j tỷ lệ thuận với độ dốc của ICC tại điểm b_j trên thang đo năng lực. Các câu hỏi có độ dốc lớn hơn sẽ hữu ích hơn trong việc phân biệt thí sinh thành các nhóm năng lực khác nhau so với các câu hỏi có độ dốc thấp. Cụ thể, mức độ hữu ích của một câu hỏi trong việc phân biệt thí sinh gần một mức năng lực θ (tách biệt thí sinh có năng lực $\leq \theta$ với thí sinh có năng lực $> \theta$) tỷ lệ thuận

với độ dốc của ICC tại điểm θ . Các câu hỏi có độ phân biệt âm thường bị loại bỏ khỏi bài kiểm tra vì chúng cho thấy có vấn đề (ví dụ như đáp án sai) khi xác suất trả lời đúng giảm dần khi năng lực thí sinh tăng. Giá trị a_j cao tạo ra các hàm đặc trưng câu hỏi rất dốc, trong khi giá trị a_j thấp dẫn đến các hàm đặc trưng câu hỏi tăng dần một cách từ từ theo năng lực. Điều này giúp mô hình hai tham số trở nên linh hoạt và phù hợp hơn trong việc đánh giá năng lực thí sinh một cách chính xác [18].

Khi năng lực thay đổi một lượng nhỏ $\Delta\theta$, xác suất trả lời đúng cũng thay đổi theo một lượng ΔP . Nếu đường cong có độ dốc lớn, sự thay đổi này sẽ rõ rệt hơn so với những đường cong có độ dốc nhỏ. Điều này có nghĩa là một câu hỏi có độ dốc lớn sẽ phân biệt tốt hơn giữa thí sinh có năng lực cao và thấp. Độ dốc của đường cong ĐTCH có thể được xác định thông qua đạo hàm của xác suất trả lời đúng $P(\theta)$ theo năng lực θ :

$$\frac{\partial P}{\partial \theta} = \frac{ae^{-a(\theta-b)}}{(1+e^{-a(\theta-b)})^2}.$$

Tại điểm $\theta = b$ (tức là mức năng lực bằng độ khó của câu hỏi), độ dốc đạt giá trị cực đại và được xác định bởi công thức:

$$\frac{a}{4}$$

đó là giá trị cực đại mà độ dốc đạt tại điểm uốn của đường cong [3].

1.4.3. Mô hình IRT 3 tham số

Trong các mô hình đường cong đặc trưng của câu hỏi trước đây, xác suất trả lời đúng giảm về 0 khi năng lực của thí sinh rất thấp. Điều này có nghĩa là, nếu một thí sinh hoàn toàn không có năng lực $\Theta \rightarrow 0$ hay $\theta = \ln \Theta \rightarrow -\infty$, thì xác suất trả lời đúng $P(\theta)$ cũng tiến về 0 [34]. Tuy nhiên, thực tế trắc nghiệm cho thấy ngay cả những thí sinh có năng lực thấp vẫn có thể trả lời đúng một số câu hỏi nhờ đoán mò hoặc chọn ngẫu nhiên. Để phản ánh hiện tượng này, người ta bổ sung thêm một tham số mới, gọi là tham số đoán mò c vào mô hình. Tham số này xác định một giá trị tiệm cận khác 0 cho xác suất trả lời đúng khi năng lực thí sinh rất thấp.

Birnbaum (1968) đã điều chỉnh mô hình logistic hai tham số bằng cách thêm

một tham số c_i để biểu thị ảnh hưởng của việc đoán lên xác suất trả lời đúng. Tuy nhiên, việc thêm tham số này khiến mô hình mất đi một số tính chất toán học đẹp của hàm logistic. Dù vậy, mô hình này vẫn được gọi là mô hình logistic ba tham số, mặc dù về mặt kỹ thuật, nó không còn là một mô hình logistic thuần túy [3]. Phương trình của mô hình ba tham số được biểu diễn như sau:

$$P(U_{ij} = 1 \mid \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}, \quad (1.5)$$

trong đó:

- θ_i là năng lực của thí sinh thứ i ,
- b_j là độ khó của câu hỏi thứ j ,
- a_j là độ phân biệt của câu hỏi thứ j ,
- c_j là tham số đoán mò, biểu thị xác suất tối thiểu mà một thí sinh có thể trả lời đúng một câu hỏi chỉ nhờ may mắn, ngay cả khi không có kiến thức hoặc kỹ năng cần thiết.
- U_{ij} là ứng đáp của thí sinh thứ i với câu hỏi thứ j :
 - + $U_{ij} = 1$ nếu thí sinh trả lời đúng câu hỏi và
 - + $U_{ij} = 0$ khi thí sinh trả lời sai.

Rõ ràng khi $\theta \rightarrow -\infty$ thì $P(\theta) \rightarrow c$. Nghĩa là, giá trị c không thay đổi theo năng lực, hay là cả thí sinh giỏi lẫn kém đều có cùng mức đoán đúng. Về mặt lý thuyết, tham số c nằm trong khoảng $0 \leq c \leq 1$. Trong thực tế, c thường nằm trong khoảng $0 < c < 0.35$ để đảm bảo tính hợp lý [3].

Việc bổ sung c làm thay đổi cách hiểu về độ khó b . Nếu trước đây b là điểm mà xác suất trả lời đúng đạt 0.5, thì giờ đây, do tồn tại xác suất đoán mò, b được xác định tại mức xác suất trung bình giữa c và 1.0:

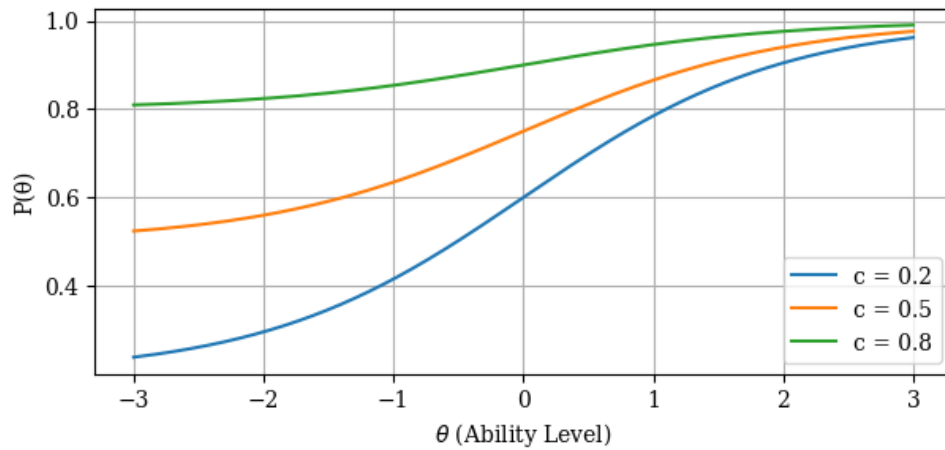
$$P(\theta) = \frac{1 + c}{2}$$

tại điểm $\theta = b$. Tham số a vẫn phản ánh độ dốc của đường cong đặc trưng câu hỏi tại $\theta = b$, nhưng trong mô hình 3 tham số, độ dốc này được điều chỉnh bởi yếu tố

$(1 - c)$:

$$\frac{a(1 - c)}{4}$$

Những thay đổi này giúp mô hình phản ánh chính xác hơn tác động của đoán mò đến xác suất trả lời đúng, đồng thời điều chỉnh cách hiểu về độ khó và độ phân biệt của câu hỏi [3].



Hình 1.4: Họ các đường cong đặc trưng câu hỏi ba tham số với $a = 1.0, b = 0.0$ và các giá trị c khác nhau.

CHƯƠNG 2. CÁC PHƯƠNG PHÁP ƯỚC LƯỢNG THAM SỐ TRONG MÔ HÌNH IRT

Ở Chương 1, chúng ta đã khám phá nền tảng của IRT thông qua các mô hình một tham số (Rasch), hai tham số và ba tham số. Các mô hình này đều giả định rằng xác suất trả lời đúng một câu hỏi được xác định bởi hai yếu tố: năng lực tiềm ẩn (θ) của thí sinh và tham số câu hỏi bao gồm độ khó (b), độ phân biệt (a), và khả năng đoán mò (c). IRT không chỉ đánh giá chính xác năng lực thí sinh mà còn phân tích chất lượng từng câu hỏi, giúp tối ưu hóa bài kiểm tra.

Sang Chương 2, trọng tâm chuyển sang ước lượng tham số – bước then chốt để áp dụng IRT vào thực tiễn. Trong IRT, xác suất thí sinh trả lời đúng một câu hỏi phụ thuộc vào năng lực tiềm ẩn (θ) của họ và các tham số đặc trưng của câu hỏi (chẳng hạn như độ khó, độ phân biệt). Tuy nhiên, cả hai yếu tố này đều không thể quan sát trực tiếp mà chỉ có dữ liệu phản hồi từ bài kiểm tra. Vì vậy, bài toán đặt ra là làm thế nào để suy ra θ của mỗi thí sinh cũng như các tham số câu hỏi từ dữ liệu phản hồi và đây là một bài toán khó.

Khác với hồi quy logistic (nơi biến độc lập quan sát được và mô hình hóa xác suất nhị phân), IRT phải xử lý bài toán phức tạp hơn do biến hồi quy (θ) là không quan sát được. Cụ thể:

- Mỗi quan hệ phi tuyến: Dù cả IRT và hồi quy logistic đều sử dụng hàm logistic để mô hình hóa xác suất, IRT yêu cầu đồng thời ước lượng θ (năng lực ẩn) và tham số câu hỏi từ dữ liệu phản hồi nhị phân.
- Tính ẩn của biến độc lập: Trong hồi quy logistic, các biến dự báo (ví dụ: tuổi, giới tính) đã được đo lường trực tiếp, nhưng trong IRT, θ chỉ được suy luận gián tiếp từ phản hồi của thí sinh.

Phương pháp ước lượng chủ yếu dựa trên hợp lý cực đại (MLE) hoặc Bayesian:

- Nếu tham số câu hỏi đã biết, việc ước lượng θ tương tự như hồi quy logistic, sử dụng hàm likelihood để tối ưu hóa giá trị năng lực.
- Khi cả θ và tham số đều chưa biết, bài toán trở nên phức tạp do cần giải đồng thời hai tập tham số.

Thách thức chính bao gồm:

- Dữ liệu thiếu/không đồng nhất: Mô hình IRT đòi hỏi cỡ mẫu lớn để ước lượng chính xác.
- Tính toán phức tạp: xử lý các phương trình phi tuyến và tối ưu hóa đa chiều.

Nếu biết trước năng lực thí sinh θ , việc xác định tham số câu hỏi sẽ trở nên đơn giản hơn, tương tự như ước lượng hệ số trong mô hình hồi quy logistic. Ngược lại, nếu các tham số câu hỏi đã được xác định, việc ước lượng năng lực thí sinh cũng trở nên dễ dàng hơn. Trong thực tế, mô hình IRT không bao giờ khớp hoàn toàn với dữ liệu, vì vậy cần tìm cách ước lượng tham số sao cho mô hình phù hợp nhất với dữ liệu. Trong hồi quy logistic, tiêu chí thường dùng để tối ưu hóa mô hình là hợp lý tối đa, giúp tìm ra tham số tối ưu để xác suất dự đoán sát với dữ liệu quan sát. Tương tự, trong IRT, phương pháp ước lượng hợp lý tối đa (MLE) cũng là một kỹ thuật phổ biến để xác định tham số câu hỏi và năng lực thí sinh. Phương pháp này có cơ sở lý thuyết vững chắc, hiệu quả với các mẫu lớn, và giúp cải thiện độ chính xác của mô hình [18].

Trong phần tiếp theo, chúng ta sẽ đi sâu vào quy trình ước lượng năng lực thí sinh khi đã biết tham số câu hỏi, sau đó xem xét cách xác định tham số câu hỏi từ dữ liệu thu thập được.

2.1. Phương pháp hợp lý cực đại

Ước lượng hợp lý cực đại (Maximum Likelihood Estimation - MLE) là một trong những phương pháp phổ biến nhất để ước lượng các tham số của mô hình thống kê. Ý tưởng cốt lõi của MLE là chọn các tham số sao cho chúng tối đa hóa xác suất quan sát được dữ liệu huấn luyện dưới mô hình đã cho. Nói cách khác, chúng ta tìm các tham số làm cho dữ liệu trở nên có khả năng xảy ra nhất theo mô hình. Tiếp theo, phần trình ở dưới dựa trên quyển “Probabilistic machine learning: an introduction” của Murphy (2022) [27].

Giả sử chúng ta có một tập dữ liệu huấn luyện $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, trong đó x_n là các biến đầu vào và y_n là các biến đầu ra tương ứng. Mục tiêu của MLE là tìm các tham số θ sao cho xác suất của dữ liệu \mathcal{D} dưới mô hình $p(\mathcal{D} | \theta)$ là lớn nhất. Hàm hợp lý được định nghĩa là

xác suất của dữ liệu \mathcal{D} khi biết tham số θ :

$$p(\mathcal{D} | \theta) = \prod_{n=1}^N p(y_n | x_n, \theta).$$

Ở đây, chúng ta giả định rằng các mẫu dữ liệu được lấy mẫu độc lập và cùng phân phối. Điều này có nghĩa là xác suất của toàn bộ tập dữ liệu là tích của xác suất của từng mẫu dữ liệu riêng lẻ.

Để đơn giản hóa việc tính toán, thay vì làm việc trực tiếp với hàm hợp lý, chúng ta thường sử dụng log likelihood:

$$\ell(\theta) \triangleq \log p(\mathcal{D} | \theta) = \sum_{n=1}^N \log p(y_n | x_n, \theta)$$

Log likelihood chuyển tích thành tổng, giúp việc tính toán trở nên dễ dàng hơn, đặc biệt khi làm việc với các mô hình phức tạp. Mục tiêu của MLE là tìm các tham số θ sao cho log likelihood đạt giá trị lớn nhất:

$$\hat{\theta}_{\text{mle}} = \arg \max_{\theta} \sum_{n=1}^N \log p(y_n | x_n, \theta).$$

Vì hầu hết các thuật toán tối ưu hóa được thiết kế để tối thiểu hóa hàm chi phí, chúng ta thường chuyển bài toán tối đa hóa log likelihood thành bài toán tối thiểu hóa negative log likelihood (NLL):

$$\text{NLL}(\theta) \triangleq -\log p(\mathcal{D} | \theta) = -\sum_{n=1}^N \log p(y_n | x_n, \theta)$$

Do đó, ước lượng MLE có thể được tính bằng cách tối thiểu hóa NLL:

$$\hat{\theta}_{\text{mle}} = \arg \min_{\theta} \text{NLL}(\theta).$$

Trong trường hợp mô hình không có điều kiện (ví dụ như các mô hình không giám sát), chúng ta chỉ có các đầu ra y_n mà không có đầu vào x_n . Khi đó, hàm hợp

lý trở thành:

$$p(\mathcal{D} | \theta) = \prod_{n=1}^N p(y_n | \theta)$$

Và MLE được tính bằng:

$$\hat{\theta}_{\text{mle}} = \arg \min_{\theta} - \sum_{n=1}^N \log p(y_n | \theta).$$

Trong trường hợp mô hình có điều kiện (ví dụ như các mô hình giám sát), chúng ta quan tâm đến xác suất đồng thời của cả đầu vào và đầu ra. Khi đó, hàm hợp lý trở thành:

$$p(\mathcal{D} | \theta) = \prod_{n=1}^N p(y_n, x_n | \theta)$$

Và MLE được tính bằng:

$$\hat{\theta}_{\text{mle}} = \arg \min_{\theta} - \sum_{n=1}^N \log p(y_n, x_n | \theta),$$

2.1.1. Ước lượng năng lực của thí sinh

Khi cần ước lượng năng lực của thí sinh, ta giả thiết rằng các tham số của câu hỏi đã được xác định.

Mỗi thí sinh khi làm bài kiểm tra sẽ tạo ra một vectơ ứng đáp, bao gồm các phản hồi nhị phân (1 nếu trả lời đúng, 0 nếu trả lời sai) cho N câu hỏi. Mục tiêu của quá trình ước lượng là sử dụng vectơ ứng đáp này cùng với các tham số câu hỏi đã biết để suy ra năng lực của thí sinh.

Quy trình ước lượng năng lực thí sinh sử dụng phương pháp MLE:

- Khởi tạo giá trị ban đầu: Gán một giá trị năng lực tiên nghiệm cho thí sinh.
- Tính toán xác suất trả lời đúng: Dựa trên giá trị năng lực giả định và các tham số câu hỏi đã biết, ta tính xác suất trả lời đúng của thí sinh cho từng câu hỏi.
- Điều chỉnh giá trị năng lực: So sánh các xác suất tính được với vectơ ứng đáp thực tế của thí sinh, sau đó điều chỉnh giá trị năng lực để tăng mức độ

phù hợp.

- Lặp lại quá trình: Tiếp tục điều chỉnh cho đến khi sự thay đổi trong giá trị năng lực trở nên rất nhỏ, tức là đã hội tụ đến giá trị tối ưu.

Dựa trên giả định độc lập cục bộ tức là xác suất trả lời đúng một câu hỏi nào đó không phụ thuộc vào các câu hỏi khác, xác suất chung của mẫu phản hồi này chính là tích của xác suất phản hồi cho từng câu hỏi, tức là:

$$P(U_1, U_2, \dots, U_N | \theta) = P(U_1 | \theta)P(U_2 | \theta) \dots P(U_N | \theta).$$

Biểu thức này có thể viết gọn hơn dưới dạng tích:

$$P(U_1, U_2, \dots, U_N | \theta) = \prod_{j=1}^N P(U_j | \theta).$$

Vì U_j có thể nhận giá trị 0 hoặc 1, ta có thể viết lại biểu thức trên dưới dạng hàm hợp lý (likelihood function) như sau:

$$L(U_1, U_2, \dots, U_N | \theta) = \prod_{j=1}^N P(U_j | \theta)^{U_j} [1 - P(U_j | \theta)]^{(1-U_j)}$$

hoặc có thể viết ngắn gọn hơn

$$L(\mathbf{U} | \theta) = \prod_{j=1}^N P_j^{U_j} Q_j^{1-U_j}. \quad (2.1)$$

Phương trình trên là biểu thức xác suất chung của một mẫu phản hồi. Khi mẫu phản hồi đã được quan sát, tức là $U_j = u_j$, thì cách diễn giải xác suất không còn phù hợp nữa; biểu thức xác suất chung này được gọi là hàm hợp lý (likelihood function) và được ký hiệu là $L(u_1, u_2, \dots, u_N | \theta)$, trong đó u_j là phản hồi quan sát được cho câu hỏi thứ j . Vậy, hàm hợp lý có dạng:

$$L(u_1, u_2, \dots, u_N | \theta) = \prod_{j=1}^N P_j^{u_j} Q_j^{(1-u_j)}. \quad (2.2)$$

Vì P_j và Q_j là các hàm của θ và các tham số câu hỏi, nên hàm hợp lý cũng là một

hàm của các tham số này [18].

Việc lấy logarit tự nhiên của hàm hợp lý cho kết quả:

$$\begin{aligned}\ell = \ln L(\mathbf{u} \mid \theta) &= \sum_{j=1}^N [u_j \ln P_j + (1 - u_j) \ln(1 - P_j)] \\ &= \sum_{j=1}^N [u_j \ln P_j + (1 - u_j) \ln Q_j].\end{aligned}\quad (2.3)$$

Trong đó, \mathbf{u} là vectơ các phản hồi của thí sinh đối với các câu hỏi trong bài kiểm tra. Giá trị θ được xác định sao cho hàm hợp lý đạt cực đại chính là ước lượng năng lực của thí sinh theo phương pháp MLE. Nói cách khác, đây là giá trị θ tối ưu nhất để mô hình có thể giải thích tốt nhất các phản hồi của thí sinh.

Bài toán tìm giá trị cực đại của một hàm số không hề đơn giản. Trong thực tế, giá trị cực đại của hàm có thể được tìm bằng các thuật toán tìm kiếm trên máy tính. Một phương pháp hiệu quả hơn dựa trên thực tế rằng tại điểm cực đại của hàm số, độ dốc của hàm (đạo hàm bậc nhất) bằng 0. Do đó, MLE có thể được xác định bằng cách giải phương trình thu được khi đặt đạo hàm bậc nhất của hàm hợp lý hoặc log-hợp lý bằng 0. Tuy nhiên, phương trình này không thể giải trực tiếp, vì vậy cần sử dụng các phương pháp xấp xỉ. Một trong những phương pháp phổ biến nhất là thuật toán Newton-Raphson, được mô tả chi tiết trong nghiên cứu của Hambleton và Swaminathan (1985) [18].

Tuy nhiên, trong một số trường hợp, hàm hợp lý có thể không có giá trị cực đại hữu hạn. Chẳng hạn, nếu một thí sinh trả lời đúng tất cả các câu hỏi hoặc sai tất cả các câu hỏi thì ước lượng hợp lý tối đa sẽ là $\theta = +\infty$ hoặc $\theta = -\infty$. Một số dạng phản hồi bất thường (không thể nhận biết trước) cũng có thể dẫn đến hàm hợp lý không có cực đại tuyệt đối hữu hạn. Lý do dẫn đến tình trạng này là do mô hình không thể giải thích hợp lý các phản hồi bất thường: những thí sinh này trả lời đúng một số câu hỏi khó có độ phân biệt cao, nhưng lại trả lời sai các câu hỏi dễ hơn. Trong trường hợp như vậy, thuật toán tìm cực đại thường sẽ hội tụ sai hoặc không hội tụ. Vấn đề này chủ yếu xảy ra trong mô hình ba tham số và không xuất hiện trong mô hình một hoặc hai tham số. Hiện tượng này có thể xảy ra ngay cả với các bài kiểm tra có tới 40 câu hỏi [18].

Ở đây sẽ trình bày cách tổng quát để ước lượng θ . Vì các tham số của tất cả N câu hỏi được giả định là đã biết và là giá trị thực, nên chỉ cần lấy đạo hàm của log-hàm hợp lý theo năng lực của thí sinh:

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^n \left[u_j \frac{1}{P_j} \frac{\partial P_j}{\partial \theta} + (1 - u_j) \frac{1}{Q_j} \frac{\partial Q_j}{\partial \theta} \right]. \quad (2.4)$$

Các đạo hàm của P_j và Q_j theo tham số năng lực sẽ phụ thuộc vào mô hình đường cong đặc trưng câu hỏi được sử dụng. Do đó, để trình bày một cách tổng quát, các đạo hàm này được giữ nguyên dưới dạng định nghĩa.

Kỹ thuật Newton-Raphson sẽ được sử dụng để tìm ước lượng tham số năng lực thông qua một quy trình lặp. Do đó, cần tính đến đạo hàm cấp hai của hàm hợp lý theo tham số năng lực. Với một thí sinh cụ thể, phương trình Newton-Raphson có thể được thiết lập để lặp và tìm giá trị cực đại hợp lý của năng lực như sau:

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \left[\frac{\partial^2 \ell}{\partial \theta^2} \right]_t^{-1} \left[\frac{\partial \ell}{\partial \theta} \right]_t. \quad (2.5)$$

Sau khi thực hiện quy trình Newton-Raphson, một giá trị ước lượng của tham số năng lực θ cho thí sinh được tìm ra. Trong các phần tiếp theo, phương trình Newton-Raphson để ước lượng năng lực sẽ được trình bày tương ứng với từng mô hình đường cong đặc trưng câu hỏi cụ thể.

Các ước lượng hợp lý cực đại khi tồn tại, có những tính chất tiệm cận quan trọng, đặc biệt khi kích thước mẫu lớn. Trong bối cảnh của một thí sinh cụ thể, khái niệm tiệm cận được hiểu là khi độ dài bài kiểm tra tăng lên. Khi số lượng câu hỏi trong bài kiểm tra càng nhiều, ước lượng MLE của năng lực θ , ký hiệu là $\hat{\theta}$, tiệm cận theo phân phối chuẩn với kỳ vọng đúng bằng giá trị thực của θ . Điều này có nghĩa là, đối với những bài kiểm tra dài, ước lượng MLE của θ là không chệch, tức là giá trị trung bình của nó trùng với giá trị thực của θ .

Sai số của ước lượng MLE, hay còn gọi là sai số chuẩn (Standard Error – SE), được tính theo công thức:

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\hat{\theta})}}, \quad (2.6)$$

trong đó $I(\theta)$ là hàm thông tin (information function), phản ánh mức độ chính xác của bài kiểm tra trong việc ước lượng θ , được tính bằng tổng các hàm thông tin của từng câu hỏi trong bài [3]:

$$I(\theta) = \sum_{j=1}^N I_j(\theta) = \sum_{j=1}^N \frac{\left[\frac{\partial P_j(\theta)}{\partial \theta} \right]^2}{P_j(\theta)Q_j(\theta)}. \quad (2.7)$$

Hay sai số chuẩn còn có thể dựa trên đạo hàm bậc hai của hàm ℓ [30]:

$$\sigma(\hat{\theta}) = \frac{1}{\sqrt{-\frac{\partial^2 \ell}{\partial \hat{\theta}^2}}}. \quad (2.8)$$

Điều này cho thấy rằng sai số tiêu chuẩn của năng lực ước lượng tỷ lệ nghịch với căn bậc hai của hàm thông tin. Nói cách khác, bài kiểm tra càng cung cấp nhiều thông tin về năng lực của thí sinh, thì độ chính xác của ước lượng càng cao, dẫn đến sai số càng nhỏ [3]. Đây là lý do tại sao việc thiết kế bài kiểm tra hiệu quả, với các câu hỏi có độ phân biệt tốt và độ khó phù hợp, đóng vai trò quan trọng trong việc nâng cao độ tin cậy của kết quả đánh giá.

Dựa trên tính chất phân phối chuẩn của $\hat{\theta}$, ta có thể xác định khoảng tin cậy cho θ . Khoảng tin cậy mức $(1 - \alpha)\%$ được tính theo công thức:

$$(\hat{\theta} - Z_{\alpha/2}SE(\hat{\theta}), \hat{\theta} + Z_{\alpha/2}SE(\hat{\theta})),$$

trong đó:

- $SE(\hat{\theta})$ là sai số chuẩn tại $\hat{\theta}$.
- $Z_{\alpha/2}$ bách phân vị (percentile) ở mức $1 - \alpha/2$ của phân phối chuẩn.

Ví dụ, với khoảng tin cậy 95%, ta chọn:

- $\alpha = 0.05$
- $Z_{\alpha/2} = 1.96$, tức là bách phân vị thứ 97.5% của phân phối chuẩn.

Điều này có nghĩa là nếu ta lặp lại quá trình ước lượng trên nhiều tập dữ liệu khác nhau, trong 95% số lần, khoảng tin cậy tính được sẽ chứa giá trị thực của θ [18].

2.1.1.1. Mô hình IRT 1 tham số

Như đã chỉ ra chương 2, mô hình IRT 1 tham số có dạng:

$$P(U_{ij} = 1 \mid \theta_i, b_j) = \frac{1}{1 + e^{-(\theta_i - b_j)}}.$$

Và một số đạo hàm hữu ích theo tham số θ_i :

$$\begin{aligned} \frac{\partial P_j}{\partial \theta_i} &= \frac{-\frac{\partial}{\partial \theta_i} (1 + e^{-(\theta_i - b_j)})}{(1 + e^{-(\theta_i - b_j)})^2} \\ &= \frac{e^{-(\theta_i - b_j)} \cdot \frac{\partial}{\partial \theta_i} (\theta_i - b_j)}{(1 + e^{-(\theta_i - b_j)})^2} \\ &= \frac{e^{-(\theta_i - b_j)}}{(1 + e^{-(\theta_i - b_j)})^2} \\ &= \frac{1}{1 + e^{-(\theta_i - b_j)}} \cdot \frac{1 + e^{-(\theta_i - b_j)} - 1}{1 + e^{-(\theta_i - b_j)}} \\ &= P_j \cdot Q_j, \end{aligned} \tag{2.9}$$

và

$$\frac{\partial Q_j}{\partial \theta_i} = -P_j \cdot Q_j. \tag{2.10}$$

Thay thế (2.9) và (2.10) này vào phương trình (2.4) sẽ thu được đạo hàm bậc nhất của hàm log-likelihood theo θ_i , và đó là:

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_i} &= \sum_{j=1}^N \left[u_j \frac{1}{P_j} (P_j \cdot Q_j) + (1 - u_j) \frac{1}{Q_j} (-P_j \cdot Q_j) \right] \\ &= \sum_{j=1}^N [u_j Q_j - (1 - u_j) P_j] \\ &= \sum_{j=1}^N [u_j - P_j]. \end{aligned} \tag{2.11}$$

Tiếp theo, đạo hàm bậc hai của hàm log-likelihood theo θ_i là:

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta_i^2} &= \sum_{j=1}^N \frac{\partial}{\partial \theta_i} [u_j - P_j] \\ &= \sum_{j=1}^N -P_j \cdot Q_j.\end{aligned}\quad (2.12)$$

Từ (2.11) và (2.12) theo phương pháp lặp Newton-Raphson, thì ta có phương trình ước lượng θ cho mô hình IRT 1 tham số như sau:

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t + \frac{\sum_{j=1}^N [u_j - P_j]}{\sum_{j=1}^N P_j \cdot Q_j}.\quad (2.13)$$

Và ta thu được sai số chuẩn là:

$$SE(\hat{\theta}_i) = \frac{1}{\sqrt{\sum_{j=1}^N P_j \cdot Q_j}}\quad (2.14)$$

2.1.1.2. Mô hình IRT 2 tham số

Mô hình IRT có dạng như sau:

$$P(U_{ij} = 1 \mid \theta_i, a_j, b_j) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}.$$

Tương tự như phần trên, ta có đạo hàm của P_j và Q_j theo θ_i như sau:

$$\begin{aligned}\frac{\partial P_j}{\partial \theta_i} &= \frac{-\frac{\partial}{\partial \theta_i} (1 + e^{-a_j(\theta_i - b_j)})}{(1 + e^{-a_j(\theta_i - b_j)})^2} \\ &= \frac{\frac{\partial}{\partial \theta_i} [a_j(\theta_i - b_j)] \cdot e^{-a_j(\theta_i - b_j)}}{(1 + e^{-a_j(\theta_i - b_j)})^2}\end{aligned}$$

$$\begin{aligned}
&= \frac{a_j e^{-a_j(\theta_i - b_j)}}{(1 + e^{-a_j(\theta_i - b_j)})^2} \\
&= a_j P_j \cdot Q_j,
\end{aligned} \tag{2.15}$$

và

$$\frac{\partial Q_j}{\partial \theta_i} = -a_j P_j \cdot Q_j. \tag{2.16}$$

Thế (2.15), (2.16) vào phương trình (2.4) sẽ thu được đạo hàm bậc nhất của hàm log-likelihood theo θ_i và đó là:

$$\begin{aligned}
\frac{\partial \ell}{\partial \theta_i} &= \sum_{j=1}^N \left[u_j \frac{1}{P_j} a_j P_j \cdot Q_j + (1 - u_j) \frac{1}{Q_j} (-a_j P_j \cdot Q_j) \right] \\
&= \sum_{j=1}^N [u_j a_j Q_j - (1 - u_j) a_j P_j] \\
&= \sum_{j=1}^N [a_j (u_j - P_j)].
\end{aligned} \tag{2.17}$$

Tiếp theo, đạo hàm bậc hai của hàm log-likelihood theo θ_i được:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \theta_i^2} &= \sum_{j=1}^N \frac{\partial}{\partial \theta_i} [a_j (u_j - P_j)] \\
&= \sum_{j=1}^N -a_j^2 P_j \cdot Q_j.
\end{aligned} \tag{2.18}$$

Từ (2.17) và (2.18) thế vào (2.5), thì ta có phương trình ước lượng θ cho mô hình IRT 2 tham số như sau:

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t + \frac{\sum_{j=1}^N [a_j (u_j - P_j)]}{\sum_{j=1}^N a_j^2 P_j Q_j}. \tag{2.19}$$

Và ta thu được sai số chuẩn là:

$$SE(\hat{\theta}_i) = \frac{1}{\sqrt{\sum_{j=1}^N a_j^2 P_j Q_j}} \quad (2.20)$$

2.1.1.3. Mô hình IRT 3 tham số

Như đã đề cập ở chương 2, mô hình IRT 3 tham số có dạng:

$$P(U_{ij} = 1 \mid \theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}.$$

Đầu tiên ta xét một số biểu thức sau $P_j^* = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$ và $Q_j^* = 1 - P_j^*$. Tiếp theo thực hiện một số phép biến đổi cơ bản sau đây:

$$\begin{aligned} P_j &= c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \\ P_j &= c_j + (1 - c_j) P_j^* \\ \Leftrightarrow P_j^* &= \frac{P_j - c_j}{1 - c_j}, \end{aligned} \quad (2.21)$$

và

$$\begin{aligned} Q_j^* &= 1 - P_j^* = 1 - \frac{P_j - c_j}{1 - c_j} \\ &= \frac{1 - c_j - P_j + c_j}{1 - c_j} = \frac{1 - P_j}{1 - c_j} \\ &= \frac{Q_j}{1 - c_j}. \end{aligned} \quad (2.22)$$

Tiếp tục đạo hàm của P_j và Q_j theo θ_i :

$$\begin{aligned} \frac{\partial P_j}{\partial \theta_i} &= (1 - c_j) \frac{\partial}{\partial \theta_i} \left[\frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \right] \\ &= (1 - c_j) \frac{-\frac{\partial}{\partial \theta_i} (1 + e^{-a_j(\theta_i - b_j)})}{(1 + e^{-a_j(\theta_i - b_j)})^2} \end{aligned}$$

$$\begin{aligned}
&= (1 - c_j) \frac{\frac{\partial}{\partial \theta_i} [a_j(\theta_i - b_j)] \cdot e^{-a_j(\theta_i - b_j)}}{(1 + e^{-a_j(\theta_i - b_j)})^2} \\
&= (1 - c_j) \frac{a_j e^{-a_j(\theta_i - b_j)}}{(1 + e^{-a_j(\theta_i - b_j)})^2} \\
&= a_j(1 - c_j)P_j^* \cdot Q_j^*,
\end{aligned} \tag{2.23}$$

thay (2.21) và (2.22) vào (2.23) ta được:

$$\begin{aligned}
\frac{\partial P_j}{\partial \theta_i} &= a_j(1 - c_j) \frac{P_j - c_j}{1 - c_j} \frac{Q_j}{1 - c_j} \\
&= a_j Q_j \frac{P_j - c_j}{1 - c_j}
\end{aligned} \tag{2.24}$$

và

$$\frac{\partial Q_j}{\partial \theta_i} = -a_j Q_j \frac{P_j - c_j}{1 - c_j}. \tag{2.25}$$

Thế (2.24), (2.25) vào phương trình (2.4) sẽ thu được được đạo hàm bậc nhất của hàm log-likelihood theo θ_i và đó là:

$$\begin{aligned}
\frac{\partial \ell}{\partial \theta_i} &= \sum_{j=1}^N \left[u_j \frac{1}{P_j} a_j Q_j \frac{P_j - c_j}{1 - c_j} + (1 - u_j) \frac{1}{Q_j} (-a_j) Q_j \frac{P_j - c_j}{1 - c_j} \right] \\
&= \sum_{j=1}^N a_j \frac{P_j - c_j}{1 - c_j} \left[u_j \frac{Q_j}{P_j} - (1 - u_j) \right] \\
&= \sum_{j=1}^N a_j \frac{P_j - c_j}{1 - c_j} \frac{u_j - P_j}{P_j}.
\end{aligned} \tag{2.26}$$

Tiếp theo, đạo hàm bậc hai của hàm log-likelihood theo θ_i được:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \theta_i^2} &= \frac{\partial}{\partial \theta_i} \left[\sum_{j=1}^n a_j \frac{P_j - c_j}{1 - c_j} \frac{u_j - P_j}{P_j} \right] = \frac{\partial}{\partial \theta_i} \left[\sum_{j=1}^n a_j P_j^* \frac{u_j - P_j}{P_j} \right] \\
&= \sum_{j=1}^N a_j \left\{ P_j^* \frac{-P_j(1 - c_j)P_j^* Q_j^* a_j - (u_j - P_j)(1 - c_j)P_j^* Q_j^* a_j}{P_j^2} + \frac{u_j - P_j}{P_j} P_j^* Q_j^* a_j \right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^N a_j^2 P_j^* Q_j^* \left\{ \frac{-(1-c_j)P_j P_j^* - (u_j - P_j)(1-c_j)P_j^*}{P_j^2} + \frac{u_j - P_j}{P_j} \right\} \\
&= \sum_{j=1}^N a_j^2 P_j^* Q_j^* \frac{u_j c_j - P_j^2}{P_j^2} = \sum_{j=1}^n a_j^2 \left(\frac{P_j - c_j}{1 - c_j} \right) \left(\frac{Q_j}{1 - c_j} \right) \left(\frac{u_j c_j - P_j^2}{P_j^2} \right) \\
&= \sum_{j=1}^N a_j^2 \frac{P_j - c_j}{(1 - c_j)^2} \frac{Q_j}{P_j} \frac{u_j c_j - P_j^2}{P_j}.
\end{aligned} \tag{2.27}$$

Việc thay thế u_j bằng P_j giúp giảm nhiễu, làm cho tính toán ổn định hơn, ta được:

$$\begin{aligned}
E \left(\frac{\partial^2 \ell}{\partial \theta_i^2} \right) &= \sum_{j=1}^N a_i^2 \frac{Q_j}{P_j} \frac{P_j - c_j}{(1 - c_j)^2} (c_i - P_{ij}) \\
&= - \sum_{j=1}^N a_i^2 \frac{Q_j}{P_j} \frac{(P_j - c_j)^2}{(1 - c_j)^2}.
\end{aligned} \tag{2.28}$$

Thay thế (2.26) và (2.28) vào phương trình (2.5), phương trình Fisher Scoring để ước lượng năng lực được thu được:

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t + \left[\frac{\sum_{j=1}^N a_j \frac{P_j - c_j}{1 - c_j} \frac{u_j - P_j}{P_j}}{\sum_{j=1}^N a_i^2 \frac{Q_j}{P_j} \frac{(P_j - c_j)^2}{(1 - c_j)^2}} \right]_t. \tag{2.29}$$

Phương sai mẫu lớn của $\hat{\theta}_j$ được cho bởi:

$$S_{\hat{\theta}_i}^2 = \frac{1}{-E \left(\frac{\partial^2 \ell}{\partial \theta_i^2} \right)} = \frac{1}{\sum_{j=1}^N a_i^2 \frac{Q_j}{P_j} \frac{(P_j - c_j)^2}{(1 - c_j)^2}},$$

và sai số chuẩn sẽ là:

$$SE(\hat{\theta}_i) = \sqrt{S_{\hat{\theta}_i}^2}. \tag{2.30}$$

2.1.2. Ước lượng các tham số của câu hỏi

Trong mô hình IRT (Item Response Theory), việc ước lượng tham số câu hỏi là một bước quan trọng để xác định các đặc trưng của câu hỏi, bao gồm độ khó b , độ phân biệt a , và độ đoán mò c . Ban đầu, chúng ta thường giả định rằng các tham số câu hỏi đã được biết trước để ước lượng năng lực θ của thí sinh. Tuy nhiên, trên thực tế, các tham số câu hỏi cũng cần được ước lượng từ dữ liệu phản ứng của thí sinh. Quá trình này dựa trên việc sử dụng hàm hợp lý để tìm ra các giá trị tham số phù hợp nhất với dữ liệu quan sát.

Khi ước lượng năng lực của một thí sinh, chúng ta sử dụng nhiều câu hỏi và xây dựng hàm hợp lý dựa trên phản ứng của thí sinh đó đối với N câu hỏi. Ngược lại, khi ước lượng tham số câu hỏi, chúng ta sử dụng phản ứng của n thí sinh đối với một câu hỏi cụ thể. Hàm hợp lý trong trường hợp này được biểu diễn như sau:

$$L(u_1, u_2, u_3, \dots, u_n | \theta, a, b, c) = \prod_{i=1}^n P^{u_i} Q^{1-u_i} \quad (2.31)$$

trong đó a , b , và c lần lượt là các tham số độ phân biệt, độ khó, và độ đoán mò.

Một điểm khác biệt quan trọng giữa hàm hợp lý cho thí sinh và hàm hợp lý cho câu hỏi là đối với câu hỏi, chúng ta chỉ cần giả định rằng các phản ứng của n thí sinh đối với một câu hỏi là độc lập, một giả định phổ biến trong thống kê. Trong khi đó, giả định về tính độc lập cục bộ nghiêm ngặt hơn, yêu cầu rằng các phản ứng của một thí sinh đối với N câu hỏi phải độc lập với nhau.

Khi các giá trị θ của thí sinh đã biết, việc ước lượng tham số câu hỏi trở nên tương đối đơn giản. Hàm hợp lý cho một câu hỏi là một hàm đa chiều, phụ thuộc vào ba tham số (a, b, c) . Để tìm ước lượng hợp lý cực đại của các tham số này, chúng ta cần tìm giá trị cực đại của một bề mặt trong không gian ba chiều. Quy trình này bao gồm các bước sau:

Bước 1: Tính đạo hàm bậc nhất của hàm hợp lý theo từng tham số.

Bước 2: Đặt các đạo hàm bằng θ và giải hệ phương trình phi tuyến ba ẩn.

Bước 3: Sử dụng phương pháp Newton-Raphson đa biến để giải hệ phương trình này.

Tùy thuộc vào mô hình được sử dụng (một, hai, hoặc ba tham số), số lượng

ần cần giải sẽ khác nhau. Khi năng lực của từng thí sinh đã biết, mỗi câu hỏi có thể được ước lượng riêng biệt mà không cần tham chiếu đến các câu hỏi khác. Do đó, quy trình ước lượng được lặp lại n lần, mỗi lần cho một câu hỏi.

Theo Baker 2001[3], trong một bài kiểm tra điển hình, n thí sinh sẽ phản ứng với N câu hỏi. Điểm năng lực của các thí sinh được phân bố trên một phạm vi rộng. Để đơn giản hóa quá trình ước lượng, các thí sinh được chia thành $g = 1, 2, 3, \dots, G$ nhóm dựa trên mức năng lực θ_g với f_g thí sinh trong mỗi nhóm. Trong mỗi nhóm, r_g thí sinh trả lời đúng câu hỏi đang xét. Tỷ lệ quan sát được của phản ứng đúng tại mức năng lực θ_g là:

$$p(\theta_g) = p_g = \frac{r_g}{f_g}. \quad (2.32)$$

Và tỷ lệ câu trả lời sai quan sát được là:

$$q(\theta_g) = q_g = \frac{f_g - r_g}{f_g}. \quad (2.33)$$

Giả định rằng số lượng câu trả lời đúng r_g tại mỗi mức năng lực θ_g tuân theo phân phối nhị thức (binomial distribution) với các tham số f_g và P_g , trong đó P_g là xác suất thực sự của việc trả lời đúng. Khi đó, kỳ vọng và phương sai của r_g là $E(r_g) = f_g P_g$ và $Var(r_g) = f_g P_g Q_g$ với $Q_g = 1 - P_g$. Do đó:

$$E(p_g) = E\left(\frac{r_g}{f_g}\right) = \frac{1}{f_g} E(r_g) = P_g. \quad (2.34)$$

Không có yêu cầu nào bắt buộc số lượng cá thể f_g phải giống nhau ở tất cả các mức năng lực. Việc lấy mẫu ngẫu nhiên cũng ngụ ý rằng các giá trị θ_g thu được sẽ là một mẫu ngẫu nhiên của các giá trị. Cuối cùng, phân phối mẫu của r_g quan sát được tại mỗi mức năng lực có thể được giả định theo nhiều dạng phân phối khác nhau. Tuy nhiên, lựa chọn phân phối nhị thức (binomial distribution) là cơ sở cho một số khái niệm quan trọng trong IRT.

Từ phương trình (2.31), hàm hợp lý L là xác suất quan sát được một tập hợp các giá trị r_g từ f_g với các tham số câu hỏi đã cho. Hàm hợp lý này được biểu diễn

như sau:

$$L = \prod_{g=1}^G C_{f_g}^{r_g} P(\theta_g)^{r_g} Q(\theta_g)^{f_g - r_g} \quad (2.35)$$

$$= \prod_{g=1}^G \frac{f_g!}{r_g! (f_g - r_g)!} P(\theta_g)^{r_g} Q(\theta_g)^{f_g - r_g}, \quad (2.36)$$

trong đó:

- C là hệ số tổ hợp,
- r_g là số thí sinh trả lời đúng trong nhóm g ,
- f_g là tổng số thí sinh trong nhóm g ,
- $P(\theta_g)$ là xác suất phản ứng đúng tại mức năng lực θ_g được tính từ mô hình ICC,
- $Q(\theta_g) = 1 - P(\theta_g)$ là xác suất phản ứng sai.

Mục tiêu của quá trình ước lượng là tìm các giá trị tham số câu hỏi (a, b, c) sao cho hàm hợp lý L đạt giá trị cực đại. Thay vì làm việc trực tiếp với L , người ta thường sử dụng logarit cơ sở tự nhiên của hàm hợp lý để đơn giản hóa các phép tính:

$$\ell = \ln L = \text{constant} + \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)]. \quad (2.37)$$

Bằng cách vẽ đồ thị các tỷ lệ này, chúng ta thu được một biểu đồ phân tán, từ đó cần tìm đường ICC phù hợp nhất. Để tìm ICC phù hợp, chúng ta sử dụng phương pháp MLE. Các bước cụ thể như sau:

Bước 1: Khởi tạo giá trị ban đầu cho các tham số câu hỏi, ví dụ $b = 0.0$ và $a = 1.0$ (cho mô hình irt 2 tham số).

Bước 2: Tính toán giá trị $P(\theta_g)$ tại mỗi mức năng lực dựa trên phương trình ICC.

Bước 3: So sánh giá trị quan sát được $p(\theta_g)$ với giá trị tính toán $P(\theta_g)$.

Bước 4: Điều chỉnh các tham số câu hỏi để cải thiện sự phù hợp giữa ICC và

dữ liệu quan sát.

Bước 5: Lặp lại quá trình điều chỉnh cho đến khi các thay đổi trở nên không đáng kể.

Khi quá trình kết thúc, các giá trị hiện tại của a và b được coi là ước lượng cuối cùng của tham số câu hỏi [3].

Một khía cạnh quan trọng trong IRT là đánh giá xem mô hình đường cong đặc trưng câu hỏi có phù hợp với dữ liệu phản ứng của thí sinh hay không. Để đo lường sự phù hợp này, người ta sử dụng chỉ số chi-square goodness-of-fit. Chỉ số này được định nghĩa như sau:

$$\chi^2 = \sum_{g=1}^G f_g \frac{[p(\theta_g) - P(\theta_g)]^2}{P(\theta_g)Q(\theta_g)}, \quad (2.38)$$

trong đó:

- G là số nhóm năng lực,
- θ_g là mức năng lực của nhóm g ,
- f_g là số thí sinh thuộc nhóm năng lực θ_g ,
- $p(\theta_g)$ là tỷ lệ quan sát được của phản ứng đúng trong nhóm g ,
- $P(\theta_g)$ là xác suất phản ứng đúng được tính từ mô hình ICC dựa trên các tham số ước lượng,
- $Q(\theta_g) = 1 - P(\theta_g)$.

Giá trị tiêu chuẩn là giá trị tối hạn của phân phối chi-bình phương với bậc tự do df và mức ý nghĩa α . Với df được tính như sau:

$$df = G - h,$$

trong đó:

- G là số nhóm năng lực,
- h là số tham số được ước tính từ mô hình.

Ta có bảng tra cứu sau:

Bảng 2.1: Bảng giá trị của phân phối χ -bình phương [22]

Bậc tự do	0.005	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99	0.995
1	7.87944	6.63490	5.02389	3.84146	2.70554	0.45494	0.01579	0.00393	0.00098	0.00016	0.00004
2	10.59663	9.21034	7.37776	5.99146	4.60517	1.38629	0.21072	0.10259	0.05064	0.02010	0.01003
3	12.83816	11.34487	9.34840	7.81473	6.25139	2.36597	0.58437	0.35185	0.21580	0.11483	0.07172
4	14.86026	13.27670	11.14329	9.48773	7.77944	3.35669	1.06362	0.71072	0.48442	0.29711	0.20699
5	16.74960	15.08627	12.83250	11.07050	9.23636	4.35146	1.61031	1.14548	0.83121	0.55430	0.41174
6	18.54758	16.81189	14.44938	12.59159	10.64464	5.34812	2.20413	1.63538	1.23734	0.87209	0.67573
7	20.27774	18.47531	16.01276	14.06714	12.01704	6.34581	2.83311	2.16735	1.68987	1.23904	0.98926
8	21.95495	20.09024	17.53455	15.50731	13.36157	7.34412	3.48954	2.73264	2.17973	1.64650	1.34441
9	23.58935	21.66599	19.02277	16.91898	14.68366	8.34283	4.16816	3.32511	2.70039	2.08790	1.73493
10	25.18818	23.20925	20.48318	18.30704	15.98718	9.34182	4.86518	3.94030	3.24697	2.55821	2.15586
11	26.75685	24.72497	21.92005	19.67514	17.27501	10.34100	5.57778	4.57481	3.81575	3.05348	2.60322
12	28.29952	26.21697	23.33666	21.02607	18.54935	11.34032	6.30380	5.22603	4.40379	3.57057	3.07382
13	29.81947	27.68825	24.73560	22.36203	19.81193	12.33976	7.04150	5.89186	5.00875	4.10692	3.55503
14	31.31935	29.14124	26.11895	23.68479	21.06414	13.33927	7.78953	6.57063	5.62873	4.66043	4.07467
15	32.80132	30.57791	27.48839	24.99579	22.30713	14.33886	8.54676	7.26094	6.26214	5.22935	4.60092
16	34.26719	31.99993	28.84535	26.29623	23.54183	15.33850	9.31224	7.96165	6.90766	5.81221	5.14221
17	35.71847	33.40866	30.19101	27.58711	24.76904	16.33818	10.08519	8.67176	7.56419	6.40776	5.69722
18	37.15645	34.80531	31.52638	28.86930	25.98942	17.33790	10.86494	9.39046	8.23075	7.01491	6.26480
19	38.58226	36.19087	32.85233	30.14353	27.20357	18.33765	11.65091	10.11701	8.90652	7.63273	6.84397
20	39.99685	37.56623	34.16961	31.41043	28.41198	19.33743	12.44261	10.85081	9.59078	8.26040	7.43384
21	41.40106	38.93217	35.47888	32.67057	29.61509	20.33723	13.23960	11.59131	10.28290	8.89720	8.03365
22	42.79565	40.28936	36.78071	33.92444	30.81328	21.33704	14.04149	12.33801	10.98232	9.54249	8.64272
23	44.18128	41.63840	38.07563	35.17246	32.00690	22.33688	14.84796	13.09051	11.68855	10.19572	9.26042
24	45.55851	42.97982	39.36408	36.41503	33.19624	23.33673	15.65868	13.84843	12.40115	10.85636	9.88623
25	46.92789	44.31410	40.64647	37.65248	34.38159	24.33659	16.47341	14.61141	13.11972	11.52398	10.51965
26	48.28988	45.64168	41.92317	38.88514	35.56317	25.33646	17.29188	15.37916	13.84390	12.19815	11.16024
27	49.64492	46.96294	43.19451	40.11327	36.74122	26.33634	18.11390	16.15140	14.57338	12.87850	11.80759
28	50.99338	48.27824	44.46079	41.33714	37.91592	27.33623	18.93924	16.92788	15.30786	13.56471	12.46134
29	52.33562	49.58788	45.72229	42.55697	39.08747	28.33613	19.76774	17.70837	16.04707	14.25645	13.12115
30	53.67196	50.89218	46.97924	43.77297	40.25602	29.33603	20.59923	18.49266	16.79077	14.95346	13.78672
40	66.76596	63.69074	59.34171	55.75848	51.80506	39.33534	29.05052	26.50930	24.43304	22.16426	20.70654

Nếu giá trị của chỉ số chi-square lớn hơn một giá trị tiêu chuẩn, điều này cho thấy mô hình ICC không phù hợp với dữ liệu. Nguyên nhân có thể là:

1. Mô hình ICC được chọn không phù hợp với đặc điểm của câu hỏi.
2. Dữ liệu quan sát quá phân tán, khiến không có mô hình nào có thể phù hợp tốt.

Trong hầu hết các bài kiểm tra, một số ít câu hỏi có thể có giá trị chi-square cao do dữ liệu phân tán. Tuy nhiên, nếu nhiều câu hỏi không đạt được độ phù hợp tốt, có thể nghi ngờ rằng mô hình đang sử dụng không phù hợp. Trong trường hợp này, việc phân tích lại bài kiểm tra với một mô hình khác (ví dụ: chuyển từ mô hình Rasch sang mô hình ba tham số) có thể mang lại kết quả tốt hơn [3].

Quy trình ước lượng tham số câu hỏi bằng MLE khá phức tạp về mặt toán

học và đòi hỏi tính toán nhiều bước cho từng câu hỏi trong bài kiểm tra. Trước đây, khi máy tính chưa phổ biến, IRT gần như không khả thi do yêu cầu tính toán quá lớn. Tuy nhiên, ngày nay với sự hỗ trợ của máy tính, quy trình này đã trở nên dễ dàng hơn.

2.1.2.1. Mô hình IRT 1 tham số

Như đã đề cập mô hình Rasch có dạng $P(U_{ij} = 1 | \theta_i, b_j) = \frac{1}{1 + e^{-(\theta_i - b_j)}}$. Ước lượng hợp lý cực đại của các tham số chưa biết b được xác định là giá trị thỏa mãn các phương trình:

$$\frac{\partial \ell}{\partial b} = 0. \quad (2.39)$$

Đầu tiên ta có đạo hàm P_g và Q_g theo b_j :

$$\begin{aligned} \frac{\partial P_g}{\partial b_j} &= \frac{-\frac{\partial}{\partial b_j} (1 + e^{-(\theta_g - b_j)})}{(1 + e^{-(\theta_g - b_j)})^2} = \frac{\frac{\partial}{\partial b_j} (\theta_g - b_j) \cdot e^{-(\theta_g - b_j)}}{(1 + e^{-(\theta_g - b_j)})^2} \\ &= \frac{-e^{-(\theta_g - b_j)}}{(1 + e^{-(\theta_g - b_j)})^2} = \frac{-1}{1 + e^{-(\theta_g - b_j)}} \cdot \frac{1 + e^{-(\theta_g - b_j)} - 1}{1 + e^{-(\theta_g - b_j)}} \\ &= -P_g \cdot Q_g, \end{aligned} \quad (2.40)$$

và

$$\frac{\partial Q_g}{\partial b_j} = P_g \cdot Q_g. \quad (2.41)$$

Khi lấy đạo hàm riêng cấp một, ta có:

$$\begin{aligned} \frac{\partial \ell}{\partial b_j} &= \sum_{g=1}^G \left[r_g \frac{1}{P_g} \frac{\partial P_g}{\partial b_j} + (f_g - r_g) \frac{1}{Q_g} \frac{\partial Q_g}{\partial b_j} \right] \\ &= \sum_{g=1}^G \left[r_g \frac{1}{P_g} (-P_g \cdot Q_g) + (f_g - r_g) \frac{1}{Q_g} P_g \cdot Q_g \right] \\ &= \sum_{g=1}^G [-r_g Q_g + (f_g - r_g) P_g] \\ &= \sum_{g=1}^G [-r_g + f_g P_g]. \end{aligned} \quad (2.42)$$

Tiếp theo, đạo hàm bậc hai của hàm log-likelihood theo b_j ta được:

$$\begin{aligned}\frac{\partial^2 \ell}{\partial b_j^2} &= \sum_{g=1}^G \frac{\partial}{\partial b_j} [-r_g + f_g P_g] \\ &= - \sum_{g=1}^G f_g P_g Q_g.\end{aligned}\quad (2.43)$$

Từ (2.42) và (2.43) theo phương pháp lặp Newton-Raphson, thì ta có phương trình ước lượng b cho mô hình IRT 1 tham số như sau:

$$[\hat{b}_j]_{t+1} = [\hat{b}_j]_t + \frac{\sum_{g=1}^G [-r_g + f_g P_g]}{\sum_{g=1}^G f_g P_g Q_g}.\quad (2.44)$$

2.1.2.2. Mô hình IRT 2 tham số

Trong mô hình $P(U_{ij} = 1 \mid \theta_i, a_j, b_j) = \frac{1}{1 + e^{-a_j(\theta_i - b_j)}}$ việc ước lượng tham số của câu hỏi gồm hai tham số a_j và b_j . Do đó để a_j, b_j làm hợp lý cực đại ℓ thỏa mãn phương trình sau:

$$\begin{aligned}\begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix}_{t+1} &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix}_t - \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_j^2} & \frac{\partial^2 \ell}{\partial b_j \partial a_j} \\ \frac{\partial^2 \ell}{\partial a \partial b} & \frac{\partial^2 \ell}{\partial b^2} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} \frac{\partial \ell}{\partial a_j} \\ \frac{\partial \ell}{\partial b_j} \end{bmatrix}_t \\ &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix}_t - \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}_t.\end{aligned}\quad (2.45)$$

Bây giờ, đạo hàm của P_g và Q_g theo a_j và b_j được biểu diễn như sau:

$$\frac{\partial P_g}{\partial a_j} = \frac{\partial}{\partial a_j} \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} = - \frac{\frac{\partial}{\partial a_j} [1 + e^{-a_j(\theta_g - b_j)}]}{[1 + e^{-a_j(\theta_g - b_j)}]^2}$$

$$\begin{aligned}
&= -\frac{\frac{\partial}{\partial a_j} [-a_j(\theta_g - b_j)] e^{-a_j(\theta_g - b_j)}}{[1 + e^{-a_j(\theta_g - b_j)}]^2} \\
&= (\theta_g - b_j) \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \cdot \frac{e^{-a_j(\theta_g - b_j)} + 1 - 1}{1 + e^{-a_j(\theta_g - b_j)}} \\
&= (\theta_g - b_j) P_g Q_g,
\end{aligned} \tag{2.46}$$

$$\frac{\partial Q_g}{\partial a_j} = -(\theta_g - b_j) P_g Q_g \tag{2.47}$$

và

$$\begin{aligned}
\frac{\partial P_g}{\partial b_j} &= \frac{\partial}{\partial b_j} \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} = -\frac{\partial}{\partial b_j} [1 + e^{-a_j(\theta_g - b_j)}] \frac{1}{[1 + e^{-a_j(\theta_g - b_j)}]^2} \\
&= -\frac{\partial}{\partial b_j} [-a_j(\theta_g - b_j)] \frac{e^{-a_j(\theta_g - b_j)}}{[1 + e^{-a_j(\theta_g - b_j)}]^2} \\
&= -a_j \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \frac{e^{-a_j(\theta_g - b_j)} + 1 - 1}{1 + e^{-a_j(\theta_g - b_j)}} \\
&= -a_j P_g Q_g,
\end{aligned} \tag{2.48}$$

$$\frac{\partial Q_g}{\partial b_j} = a_j P_g Q_g. \tag{2.49}$$

Tiếp theo, các đạo hàm riêng trong ma trận Hessian và ma trận Gradient được tính như sau:

$$\begin{aligned}
L_1 &= \frac{\partial \ell}{\partial a_j} = \sum_{g=1}^G \left[r_g \frac{1}{P_g} \frac{\partial P_g}{\partial a_j} + (f_g - r_g) \frac{1}{Q_g} \frac{\partial Q_g}{\partial a_j} \right] \\
&= \sum_{g=1}^G \left[r_g \frac{1}{P_g} (\theta_g - b_j) P_g Q_g - (f_g - r_g) \frac{1}{Q_g} (\theta_g - b_j) P_g Q_g \right] \\
&= \sum_{g=1}^G [r_g (\theta_g - b_j) Q_g - (f_g - r_g) (\theta_g - b_j) P_g] \\
&= \sum_{g=1}^G [(\theta_g - b_j) (r_g Q_g - (f_g - r_g) P_g)]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{g=1}^G [(\theta_g - b_j)(r_g - r_g P_g - f_g P_g + r_g P_g)] \\
&= \sum_{g=1}^G (\theta_g - b_j)(r_g - f_g P_g). \tag{2.50}
\end{aligned}$$

$$\begin{aligned}
L_2 &= \frac{\partial \ell}{\partial b_j} = \sum_{g=1}^G \left[r_g \frac{1}{P_g} \frac{\partial P_g}{\partial b_j} + (f_g - r_g) \frac{1}{Q_g} \frac{\partial Q_g}{\partial b_j} \right] \\
&= \sum_{g=1}^G \left[r_g \frac{1}{P_g} (-a_j P_g Q_g) + (f_g - r_g) \frac{1}{Q_g} a_j P_g Q_g \right] \\
&= \sum_{g=1}^G [-r_g a_j Q_g + a_j (f_g - r_g) P_g] \\
&= \sum_{g=1}^G [a_j (-r_g + r_g P_g + f_g P_g - r_g P_g)] \\
&= \sum_{g=1}^G a_j (-r_g + f_g P_g). \tag{2.51}
\end{aligned}$$

Thay thế biểu thức (2.32) vào các đạo hàm (2.50) và (2.51):

$$L_1 = \sum_{g=1}^G f_g (\theta_g - b_j) (p_g - P_g) \tag{2.52}$$

$$L_2 = \sum_{g=1}^G a_j f_g (-p_g + P_g). \tag{2.53}$$

Tiếp tục, tính đạo hàm cấp 2 của hàm ℓ :

$$\begin{aligned}
L_{11} &= \frac{\partial^2 \ell}{\partial a_j^2} = \sum_{g=1}^G \frac{\partial}{\partial a_j^2} [(\theta_g - b_j)(r_g - f_g P_g)] \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j) (\theta_g - b_j) P_g Q_g \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j)^2 P_g Q_g, \tag{2.54}
\end{aligned}$$

$$\begin{aligned}
L_{22} &= \frac{\partial^2 \ell}{b_j^2} = \sum_{g=1}^G \frac{\partial}{b_j^2} [a_j(-r_g + f_g P_g)] \\
&= \sum_{g=1}^G a_j f_g (-a_j P_g Q_g) \\
&= - \sum_{g=1}^G a_j^2 f_g P_g Q_g,
\end{aligned} \tag{2.55}$$

$$\begin{aligned}
L_{12} = L_{21} &= \frac{\partial^2 \ell}{\partial a_j \partial b_j} = \sum_{g=1}^G \frac{\partial}{\partial b_j} [(\theta_g - b_j)(r_g - f_g P_g)] \\
&= \sum_{g=1}^G \left[(r_g - f_g P_g) \frac{\partial}{\partial b_j} (\theta_g - b_j) + (\theta_g - b_j) \frac{\partial}{\partial b_j} (r_g - f_g P_g) \right] \\
&= \sum_{g=1}^G \left[-(r_g - f_g P_g) + (\theta_g - b_j) \left(-f_g \frac{\partial P_g}{\partial b_j} \right) \right] \\
&= \sum_{g=1}^G \left[-(r_g - f_g P_g) + (\theta_g - b_j) (-f_g (-a_j P_g Q_g)) \right] \\
&= \sum_{g=1}^G \left[-r_g + f_g P_g + a_j (\theta_g - b_j) f_g P_g Q_g \right].
\end{aligned} \tag{2.56}$$

2.1.2.3. Mô hình IRT 3 tham số

Mô hình 3 tham số có dạng:

$$\begin{aligned}
P(U_{ij} = 1 \mid \theta_i, a_j, b_j, c_j) &= c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \\
&= c_j + (1 - c_j) P_{ij}^*,
\end{aligned}$$

và

$$\begin{aligned}
Q(X_{ij} = 0 \mid \theta_i, a_j, b_j, c_j) &= 1 - c_j - (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \\
&= (1 - c_j) \left[1 - \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} \right]
\end{aligned}$$

$$= (1 - c_j) \left[1 - P_{ij}^* \right] = (1 - c_j) Q_{ij}^*,$$

trong đó P_{ij}^* là mô hình IRT 2 tham số.

Để ước lượng các tham số a_j, b_j, c_j ta cần giải phương trình sau:

$$\begin{aligned} \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_{t+1} &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_t - \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_j^2} & \frac{\partial^2 \ell}{\partial b_j \partial a_j} & \frac{\partial^2 \ell}{\partial c_j \partial a_j} \\ \frac{\partial^2 \ell}{\partial a_j \partial b_j} & \frac{\partial^2 \ell}{\partial b_j^2} & \frac{\partial^2 \ell}{\partial c_j \partial b_j} \\ \frac{\partial^2 \ell}{\partial a_j \partial c_j} & \frac{\partial^2 \ell}{\partial b_j \partial c_j} & \frac{\partial^2 \ell}{\partial c_j^2} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} \frac{\partial \ell}{\partial a_j} \\ \frac{\partial \ell}{\partial b_j} \\ \frac{\partial \ell}{\partial c_j} \end{bmatrix}_t \\ &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_t - \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}_t. \end{aligned} \quad (2.57)$$

Để thuận tiện trong quá trình tính toán và đơn giản một số kết quả, đặt:

$$\begin{aligned} P_j &= c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_i - b_j)}} = c_j + (1 - c_j) P_j^* \\ \Leftrightarrow (1 - c_j) P_j^* &= P_j - c_j \\ P_j^* &= \frac{P_j - c_j}{1 - c_j}, \end{aligned} \quad (2.58)$$

và

$$\begin{aligned} Q_j^* &= 1 - P_j^* = 1 - \frac{P_j - c_j}{1 - c_j} \\ &= \frac{1 - c_j - P_j + c_j}{1 - c_j} \\ &= \frac{1 - P_j}{1 - c_j} = \frac{Q_j}{1 - c_j} \end{aligned} \quad (2.59)$$

Đầu tiên, đạo hàm P_g và Q_g theo a_j, b_j, c_j được tính như sau:

$$\frac{\partial P_g}{\partial a_j} = \frac{\partial}{\partial a_j} \left[c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \right] = (1 - c_j) \frac{\partial}{\partial a_j} \frac{1}{1 + e^{-a_j(\theta_g - b_j)}}$$

$$\begin{aligned}
&= -(1-c_j) \frac{\frac{\partial}{\partial a_j} e^{-a_j(\theta_g - b_j)}}{(1 + e^{-a_j(\theta_g - b_j)})^2} = -(1-c_j) \frac{\partial}{\partial a_j} - a_j(\theta_g - b_j) \frac{e^{-a_j(\theta_g - b_j)}}{(1 + e^{-a_j(\theta_g - b_j)})^2} \\
&= (1-c_j)(\theta_g - b_j) \frac{e^{-a_j(\theta_g - b_j)}}{(1 + e^{-a_j(\theta_g - b_j)})^2} = (1-c_j)(\theta_g - b_j) P_g^* Q_g^* \\
&= (1-c_j)(\theta_g - b_j) \frac{P_g - c_j}{1 - c_j} \frac{Q_g}{1 - c_j} \\
&= \frac{(\theta_g - b_j)(P_g - c_j)Q_g}{1 - c_j}, \tag{2.60}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P_g}{\partial b_j} &= \frac{\partial}{\partial b_j} \left[c_j + (1-c_j) \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \right] = (1-c_j) \frac{\partial}{\partial b_j} \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \\
&= -(1-c_j) \frac{\frac{\partial}{\partial b_j} e^{-a_j(\theta_g - b_j)}}{(1 + e^{-a_j(\theta_g - b_j)})^2} = -(1-c_j) \frac{\partial}{\partial b_j} - a_j(\theta_g - b_j) \frac{e^{-a_j(\theta_g - b_j)}}{(1 + e^{-a_j(\theta_g - b_j)})^2} \\
&= -a_j(1-c_j) \frac{e^{-a_j(\theta_g - b_j)}}{(1 + e^{-a_j(\theta_g - b_j)})^2} = -a_j(1-c_j) P_g^* Q_g^* \\
&= -a_j(1-c_j) \frac{P_g - c_j}{1 - c_j} \frac{Q_g}{1 - c_j} \\
&= -\frac{a_j(P_g - c_j)Q_g}{1 - c_j}, \tag{2.61}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial P_g}{\partial c_j} &= \frac{\partial}{\partial c_j} \left[c_j + (1-c_j) \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \right] = 1 + \frac{\partial}{\partial c_j} (1-c_j) \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} \\
&= 1 - \frac{\frac{\partial}{\partial c_j} c_j}{1 + e^{-a_j(\theta_g - b_j)}} = 1 - \frac{1}{1 + e^{-a_j(\theta_g - b_j)}} = Q_g^* \\
&= \frac{Q_g}{1 - c_j} \tag{2.62}
\end{aligned}$$

Tiếp theo các đạo hàm riêng trong ma trận Hessian và ma trận Gradient được tính như sau:

$$L_1 = \frac{\partial \ell}{\partial a_j} = \sum_{g=1}^G \left[r_g \frac{1}{P_g} \frac{\partial P_g}{\partial a_j} + (f_g - r_g) \frac{1}{Q_g} \frac{\partial Q_g}{\partial a_j} \right]$$

$$\begin{aligned}
&= \sum_{g=1}^G \left[\frac{r_g}{P_g} (\theta_g - b_j) (1 - c_j) P_g^* Q_g^* - \frac{(f_g - r_g)}{Q_g} (\theta_g - b_j) (1 - c_j) P_g^* Q_g^* \right] \\
&= \sum_{g=1}^G (\theta_g - b_j) (1 - c_j) P_g^* Q_g^* \left[\frac{r_g}{P_g} - \frac{(f_g - r_g)}{Q_g} \right] \\
&= \sum_{g=1}^G (\theta_g - b_j) (1 - c_j) \frac{P_g^* Q_g^*}{P_g Q_g} [r_g Q_g - (f_g - r_g) P_g] \\
&= \sum_{g=1}^G (\theta_g - b_j) (1 - c_j) \frac{P_g^* Q_g^*}{P_g Q_g} [r_g - f_g P_g]
\end{aligned}$$

đặt $W_g = \frac{P_g^* Q_g^*}{P_g Q_g}$ và $r_g = f_g p_g$, như đã đề cập p_g là tỷ lệ quan sát được của ứng đáp đúng. Sau đó

$$L_1 = \sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) (1 - c_j) W_g. \quad (2.63)$$

Bằng cách đặt

$$(1 - c_j) W_g = (1 - c_j) \frac{P_g^* Q_g^*}{P_g Q_g} = (1 - c_j) \frac{\frac{P_g - c_j}{1 - c_j} \frac{Q_g}{1 - c_j}}{P_g Q_g} = \frac{P_g - c_j}{P_g (1 - c_j)}$$

thế vào biểu thức trên ta được:

$$\begin{aligned}
L_1 &= \sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) \left[\frac{P_g - c_j}{P_g (1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) \left[\frac{P_g^*}{P_g} \right]. \quad (2.64)
\end{aligned}$$

$$\begin{aligned}
L_2 &= \frac{\partial \ell}{\partial b_j} = \sum_{g=1}^G \left[\frac{r_g}{P_g} \frac{\partial P_g}{\partial b_j} + \frac{(f_g - r_g)}{Q_g} \frac{\partial Q_g}{\partial b_j} \right] \\
&= \sum_{g=1}^G \left[\frac{r_g}{P_g} (-a_j (1 - c_j) P_g^* Q_g^*) + \frac{(f_g - r_g)}{Q_g} (-a_j) (1 - c_j) P_g^* Q_g^* \right]
\end{aligned}$$

$$\begin{aligned}
&= -a_j \sum_{g=1}^G \frac{(1-c_j)P_g^*Q_g^*}{P_gQ_g} [r_gQ_g - (f_g - r_g)P_g] \\
&= -a_j \sum_{g=1}^G \frac{(1-c_j)P_g^*Q_g^*}{P_gQ_g} [f_gp_g - f_gP_g] \\
&= -a_j \sum_{g=1}^G f_g(p_g - P_g)(1-c)W_g
\end{aligned} \tag{2.65}$$

$$\begin{aligned}
&= -a_j \sum_{g=1}^G f_g(p_g - P_g) \frac{P_g - c_j}{P_g(1-c_j)} \\
&= -a_j \sum_{g=1}^G f_g(p_g - P_g) \left[\frac{P_g^*}{P_g} \right].
\end{aligned} \tag{2.66}$$

$$\begin{aligned}
L_3 &= \frac{\partial \ell}{\partial c_j} = \sum_{g=1}^G \left[\frac{r_g}{P_g} \frac{\partial P_g}{\partial b_j} + \frac{(f_g - r_g)}{Q_g} \frac{\partial Q_g}{\partial b_j} \right] = \sum_{g=1}^G \left[\frac{r_g}{P_g} Q_g^* - \frac{(f_g - r_g)}{Q_g} Q_g^* \right] \\
&= \sum_{g=1}^G \frac{Q_g^*}{P_gQ_g} [r_gQ_g - (f_g - r_g)P_g] = \sum_{g=1}^G \frac{Q_g^*}{P_gQ_g} [r_g(Q_g + P_g) - f_gP_g] \\
&= \sum_{g=1}^G \frac{Q_g^*}{P_gQ_g} [r_g - f_gP_g] = \sum_{g=1}^G \frac{Q_g^*}{P_gQ_g} [f_gp_g - f_gP_g] = \sum_{g=1}^G \frac{Q_g}{P_gQ_g} \frac{1-c_j}{1-c_j} [f_gp_g - f_gP_g] \\
&= \sum_{g=1}^G \frac{1}{(1-c_j)} \frac{1}{P_g} f_g(p_g - P_g) = \sum_{g=1}^G f_g(p_g - P_g) \frac{1}{P_g - c_j} \frac{1}{P_g} \frac{P_g - c_j}{1-c_j} \\
&= \sum_{g=1}^G f_g(p_g - P_g) \frac{1}{P_g - c_j} \frac{P_g^*}{P_g}.
\end{aligned} \tag{2.67}$$

Hãy nhớ $(1-c_j)W_g = (1-c_j) \frac{P_g^*Q_g^*}{P_gQ_g} = \frac{P_g - c_j}{P_g(1-c_j)}$. Để ý rằng các gradient L_1, L_2, L_3 đều xuất hiện chúng, để thuận tiện cho việc tính đạo hàm bậc hai, xét

đạo hàm với các tham số a_j, b_j, c_j với $\frac{P_g - c_j}{P_g(1 - c_j)}$:

$$\begin{aligned}
\frac{\partial}{\partial a_j} \left[\frac{P_g - c_j}{P_g(1 - c_j)} \right] &= \frac{\frac{\partial}{\partial a_j}(P_g - c_j) \cdot [P_g(1 - c_j)] - (P_g - c_j) \cdot \frac{\partial}{\partial a_j}[P_g(1 - c_j)]}{[P_g(1 - c_j)]^2} \\
&= \frac{\frac{(\theta_g - b_j)(P_g - c_j)}{1 - c_j} Q_g [P_g(1 - c_j)]}{[P_g(1 - c_j)]^2} \\
&= \frac{(P_g - c_j)(1 - c_j) \frac{(\theta_g - b_j)(P_g - c_j) Q_g}{1 - c_j}}{[P_g(1 - c_j)]^2} \\
&= \frac{(\theta_g - b_j)(P_g - c_j) Q_g P_g - (\theta_g - b_j)(P_g - c_j) Q_g \cdot (P_g - c_j)}{[P_g(1 - c_j)]^2} \\
&= \frac{(\theta_g - b_j)(P_g - c_j) Q_g (P_g - P_g + c_j)}{[P_g(1 - c_j)]^2} \\
&= c_j(\theta_g - b_j) \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)}, \tag{2.68}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial b_j} \left[\frac{P_g - c_j}{P_g(1 - c_j)} \right] &= \frac{\frac{\partial}{\partial b_j}(P_g - c_j) \cdot [P_g(1 - c_j)] - (P_g - c_j) \cdot \frac{\partial}{\partial b_j}[P_g(1 - c_j)]}{[P_g(1 - c_j)]^2} \\
&= \frac{-\frac{a_j(P_g - c_j) Q_g}{1 - c_j} [P_g(1 - c_j)] + (P_g - c_j)(1 - c_j) \frac{a_j(P_g - c_j) Q_g}{1 - c_j}}{[P_g(1 - c_j)]^2} \\
&= \frac{a_j(P_g - c_j) Q_g (-P_g + P_g - c_j)}{[P_g(1 - c_j)]^2} \\
&= -a_j c_j \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)}, \tag{2.69}
\end{aligned}$$

$$\frac{\partial}{\partial c_j} \left[\frac{P_g - c_j}{P_g(1 - c_j)} \right] = \frac{\frac{\partial}{\partial c_j}(P_g - c_j) \cdot [P_g(1 - c_j)] - (P_g - c_j) \cdot \frac{\partial}{\partial c_j}[P_g(1 - c_j)]}{[P_g(1 - c_j)]^2}$$

$$\begin{aligned}
&= \frac{\frac{Q_g - 1 + c_j}{1 - c_j} \cdot [P_g(1 - c_j)] - (P_g - c_j) \cdot \left[\frac{Q_g}{1 - c_j}(1 - c_j) - P_g \right]}{[P_g(1 - c_j)]^2} \\
&= \frac{-(P_g - c_j)P_g - (P_g - c_j)(Q_g - P_g)}{[P_g(1 - c_j)]^2} \\
&= \frac{-(P_g - c_j)(P_g + Q_g - P_g)}{[P_g(1 - c_j)]^2} \\
&= -\frac{P_g - c_j}{(1 - c_j)^2} \frac{Q_g}{P_g^2}.
\end{aligned} \tag{2.70}$$

Tiếp theo, đạo hàm bậc hai của hàm log-likelihood, được:

$$\begin{aligned}
L_{11} &= \frac{\partial^2 \ell}{\partial a_j^2} = \frac{\partial}{\partial a_j} \left\{ \sum_{g=1}^G f_g(p_g - P_g)(\theta_g - b_j) \left[\frac{P_g - c_j}{P_g(1 - c_j)} \right] \right\} \\
&= \sum_{g=1}^G f_g(\theta_g - b_j) \left[\frac{\partial}{\partial a_j}(p_g - P_g) \cdot \frac{P_g - c_j}{P_g(1 - c_j)} + (p_g - P_g) \cdot \frac{\partial}{\partial a_j} \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g(\theta_g - b_j) \left[-\frac{(\theta_g - b_j)(P_g - c_j)Q_g}{1 - c_j} \frac{P_g - c_j}{P_g(1 - c_j)} \right. \\
&\quad \left. + (p_g - P_g)c_j(\theta_g - b_j) \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g(\theta_g - b_j)^2 \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} [-(P_g - c_j)P_g + c_j(p_g - P_g)] \\
&= \sum_{g=1}^G f_g(\theta_g - b_j)^2 \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} (-P_g^2 + c_j p_g),
\end{aligned} \tag{2.71}$$

$$\begin{aligned}
L_{22} &= \frac{\partial^2 \ell}{\partial b_j^2} = \frac{\partial}{\partial b_j} \left[-a_j \sum_{g=1}^G f_g(p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= -a_j \sum_{g=1}^G f_g \left[\frac{\partial}{\partial b_j}(p_g - P_g) \cdot \frac{P_g - c_j}{P_g(1 - c_j)} + (p_g - P_g) \cdot \frac{\partial}{\partial b_j} \left(\frac{P_g - c_j}{P_g(1 - c_j)} \right) \right] \\
&= -a_j \sum_{g=1}^G f_g \left[\frac{a_j(P_g - c_j)Q_g}{1 - c_j} \cdot \frac{P_g - c_j}{P_g(1 - c_j)} + (p_g - P_g) \cdot -a_j c_j \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} \right]
\end{aligned}$$

$$\begin{aligned}
&= a_j^2 \sum_{g=1}^G f_g \frac{P_j - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} \left[-(P_g - c_j)P_g + c_j(p_g - P_g) \right] \\
&= a_j^2 \sum_{g=1}^G f_g \frac{P_j - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} (-P_g^2 + c_j p_g), \tag{2.72}
\end{aligned}$$

$$\begin{aligned}
L_{33} &= \frac{\partial^2 \ell}{\partial c_j^2} = \frac{\partial}{\partial c_j} \left\{ \sum_{g=1}^G \frac{1}{(1 - c_j)} \frac{1}{P_g} f_g (p_g - P_g) \right\} \\
&= \sum_{g=1}^G f_g \left[\frac{\partial}{\partial c_j} (p_g - P_g) \cdot \frac{1}{(1 - c_j)P_g} + (p_g - P_g) \cdot \frac{\partial}{\partial c_j} \frac{1}{(1 - c_j)P_g} \right] \\
&= \sum_{g=1}^G f_g \left[-\frac{Q_g}{1 - c_j} \frac{1}{(1 - c_j)P_g} - (p_g - P_g) \frac{1}{(1 - c_j)^2} \frac{Q_g}{P_g^2} \right] \\
&= - \sum_{g=1}^G f_g \frac{Q_g}{(1 - c_j)^2 P_g^2} (P_g + p_g - P_g) \\
&= - \sum_{g=1}^G \frac{f_g p_g}{(1 - c_j)^2} \frac{Q_g}{P_g^2}, \tag{2.73}
\end{aligned}$$

$$\begin{aligned}
L_{12} = L_{21} &= \frac{\partial L_1}{\partial b_j} = \frac{\partial}{\partial b_j} \left[\sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g \left[(p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} \frac{\partial}{\partial b_j} (\theta_g - b_j) + (\theta_g - b_j) \frac{\partial}{\partial b_j} (p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g \left\{ - (p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} \right. \\
&\quad \left. + (\theta_g - b_j) \left[\frac{P_g - c_j}{P_g(1 - c_j)} \frac{\partial}{\partial b_j} (p_g - P_g) + (p_g - P_g) \frac{\partial}{\partial b_j} \frac{P_g - c_j}{P_g(1 - c_j)} \right] \right\} \\
&= \sum_{g=1}^G f_g \left\{ - (p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} + (\theta_g - b_j) \left[\frac{P_g - c_j}{P_g(1 - c_j)} \frac{a_j (P_g - c_j) Q_g}{1 - c_j} \right. \right. \\
&\quad \left. \left. - (p_g - P_g) a_j c_j \frac{P_j - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} \right] \right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{g=1}^G f_g \left\{ -(p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} \right. \\
&\quad \left. + a_j(\theta_g - b_j) \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} [P_g(P_g - c_j) - (p_g - P_g)c_j] \right\} \\
&= \sum_{g=1}^G f_g \left[-(p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} + a_j(\theta_g - b_j) \frac{P_g - c_j}{P_g(1 - c_j)} \frac{Q_g}{P_g(1 - c_j)} (P_g^2 - p_g c_j) \right] \\
&= \sum_{g=1}^G f_g \frac{P_g - c_j}{P_g(1 - c_j)} \left[-(p_g - P_g) + a_j(\theta_g - b_j) \frac{Q_g}{P_g(1 - c_j)} (P_g^2 - p_g c_j) \right],
\end{aligned} \tag{2.74}$$

$$\begin{aligned}
L_{13} = L_{31} &= \frac{\partial L_1}{\partial c_j} = \frac{\partial}{\partial c_j} \left[\sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g (\theta_g - b_j) \left[\frac{P_g - c_j}{P_g(1 - c_j)} \frac{\partial}{\partial c_j} (p_g - P_g) + (p_g - P_g) \frac{\partial}{\partial c_j} \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= \sum_{g=1}^G f_g (\theta_g - b_j) \left[\frac{P_g - c_j}{P_g(1 - c_j)} \frac{-Q_g}{1 - c_j} - (p_g - P_g) \frac{P_g - c_j}{(1 - c_j)^2} \frac{Q_g}{P_g^2} \right] \\
&= \sum_{g=1}^G f_g (\theta_g - b_j) \frac{Q_g (P_g - c_j)}{[P_g(1 - c_j)]^2} [-P_g - (p_g - P_g)] \\
&= - \sum_{g=1}^G f_g p_g (\theta_g - b_j) \frac{Q_g (P_g - c_j)}{[P_g(1 - c_j)]^2},
\end{aligned} \tag{2.75}$$

$$\begin{aligned}
L_{23} = L_{32} &= \frac{\partial L_2}{\partial c_j} = \frac{\partial}{\partial c_j} \left[-a_j \sum_{g=1}^G f_g (p_g - P_g) \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= -a_j \sum_{g=1}^G f_g \left[\frac{P_g - c_j}{P_g(1 - c_j)} \frac{\partial}{\partial c_j} (p_g - P_g) + (p_g - P_g) \frac{\partial}{\partial c_j} \frac{P_g - c_j}{P_g(1 - c_j)} \right] \\
&= -a_j \sum_{g=1}^G f_g \left[\frac{P_g - c_j}{P_g(1 - c_j)} \frac{-Q_g}{1 - c_j} - (p_g - P_g) \frac{P_g - c_j}{(1 - c_j)^2} \frac{Q_g}{P_g^2} \right]
\end{aligned}$$

$$\begin{aligned}
&= -a_j \sum_{g=1}^G f_g \frac{P_g - c_j}{(1 - c_j)^2} \frac{Q_g}{P_g^2} [-P_g - (p_g - P_g)] \\
&= a_j \sum_{g=1}^G f_g p_g \frac{P_g - c_j}{(1 - c_j)^2} \frac{Q_g}{P_g^2}.
\end{aligned} \tag{2.76}$$

Trong quá trình ước lượng hợp lý cực đại các tham số câu hỏi dưới mô hình đường cong đặc trưng câu hỏi chuẩn, các đạo hàm bậc hai thường chứa các giá trị quan sát được. Để đảm bảo tính chính xác, chúng ta cần thay thế các giá trị quan sát p_g bằng kỳ vọng tương ứng P_g trong các đạo hàm bậc hai. Điều này khác biệt so với trường hợp mô hình hai tham số, nơi việc thay thế như vậy không cần thiết. Sự khác biệt này gợi ý rằng, khi các giá trị θ_g đã biết, có thể không tồn tại ước lượng đủ cho các tham số hồi quy trong mô hình ba tham số. Kỳ vọng của các đạo hàm bậc hai được tính như sau:

$$\begin{aligned}
\Lambda_{11} &= E(L_{11}) = \sum_{g=1}^G f_g (\theta_g - b_j)^2 \frac{P_g - c_j}{P_g (1 - c_j)} \frac{Q_g}{P_g (1 - c_j)} (-P_g^2 + c_j P_g) \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j)^2 \left[\frac{P_g - c_j}{P_g (1 - c_j)} \right]^2 P_g Q_g \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j)^2 \left[(1 - c_j) \frac{P_g^* Q_g^*}{P_g Q_g} \right]^2 P_g Q_g \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j)^2 \left[(1 - c_j) (P_g^* Q_g^*) \right]^2 \frac{1}{P_g Q_g} \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j)^2 P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2,
\end{aligned} \tag{2.77}$$

$$\begin{aligned}
\Lambda_{22} &= E(L_{22}) = a_j^2 \sum_{g=1}^G \frac{P_j - c_j}{P_g (1 - c_j)} \frac{Q_g}{P_g (1 - c_j)} (-P_g^2 + c_j p_g) \\
&= -a_j^2 \sum_{g=1}^G f_g P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2,
\end{aligned} \tag{2.78}$$

$$\begin{aligned}
\Lambda_{33} = E(L_{33}) &= - \sum_{g=1}^G \frac{f_g P_g}{(1-c_j)^2} \frac{Q_g}{P_g^2} \\
&= - \sum_{g=1}^G f_g \frac{Q_g}{1-c_j} \frac{P_g^*}{P_g},
\end{aligned} \tag{2.79}$$

$$\begin{aligned}
\Lambda_{12} = \Lambda_{21} = E(L_{12}) &= \sum_{g=1}^G f_g \frac{P_g - c_j}{P_g(1-c_j)} \left[a_j(\theta_g - b_j) \frac{Q_g}{P_g(1-c_j)} (P_g^2 - P_g c_j) \right] \\
&= a_j \sum_{g=1}^G f_g (\theta_g - b_j) \left[\frac{P_g - c_j}{P_g(1-c_j)} \right]^2 P_g Q_g \\
&= a_j \sum_{g=1}^G f_g (\theta_g - b_j) P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2,
\end{aligned} \tag{2.80}$$

$$\begin{aligned}
\Lambda_{13} = \Lambda_{31} = E(L_{13}) &= - \sum_{g=1}^G f_g P_g (\theta_g - b_j) \frac{Q_g (P_g - c_j)}{[P_g(1-c_j)]^2} \\
&= - \sum_{g=1}^G f_g P_g (\theta_g - b_j) \frac{Q_g}{P_g(1-c_j)} (1-c) \frac{P_g^* Q_g^*}{P_g Q_g} \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j) Q_g^* \frac{P_g^*}{P_g} \\
&= - \sum_{g=1}^G f_g (\theta_g - b_j) \frac{Q_g}{1-c_j} \frac{P_g^*}{P_g},
\end{aligned} \tag{2.81}$$

$$\begin{aligned}
\Lambda_{23} = \Lambda_{32} = L_{23} &= a_j \sum_{g=1}^G f_g P_g \frac{P_g - c_j}{(1-c_j)^2} \frac{Q_g}{P_g^2} \\
&= a_j \sum_{g=1}^G f_g \frac{P_g^* Q_g^*}{P_g Q_g} Q_g \\
&= a_j \sum_{g=1}^G f_g \frac{Q_g}{1-c_j} \frac{P_g^*}{P_g}.
\end{aligned} \tag{2.82}$$

Do đó, phương trình (2.57) được viết lại như sau:

$$\begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_{t+1} = \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_t - \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}_t \quad (2.83)$$

với các đạo hàm được tính ở trên:

$$\begin{aligned} L_1 &= \sum_{g=1}^G f_g(p_g - P_g)(\theta_g - b_j) \left[\frac{P_g^*}{P_g} \right] \\ L_2 &= -a_j \sum_{g=1}^G f_g(p_g - P_g) \left[\frac{P_g^*}{P_g} \right] \\ L_3 &= \sum_{g=1}^G f_g(p_g - P_g) \frac{1}{P_g - c_j} \frac{P_g^*}{P_g} \\ \Lambda_{11} &= - \sum_{g=1}^G f_g(\theta_g - b_j)^2 P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2 \\ \Lambda_{22} &= -a_j^2 \sum_{g=1}^G f_g P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2 \\ \Lambda_{33} &= - \sum_{g=1}^G f_g \frac{Q_g}{1 - c_j} \frac{P_g^*}{P_g} \\ \Lambda_{12} = \Lambda_{21} &= a_j \sum_{g=1}^G f_g(\theta_g - b_j) P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2 \\ \Lambda_{13} = \Lambda_{31} &= - \sum_{g=1}^G f_g(\theta_g - b_j) \frac{Q_g}{1 - c_j} \frac{P_g^*}{P_g} \\ \Lambda_{23} = \Lambda_{32} &= a_j \sum_{g=1}^G f_g \frac{Q_g}{1 - c_j} \frac{P_g^*}{P_g} \end{aligned}$$

Việc ước lượng các tham số trong mô hình ba tham số IRT phức tạp hơn nhiều so với mô hình hai tham số. Quá trình ước lượng đồng thời ba tham số gặp nhiều khó khăn do bề mặt log-likelihood phức tạp và khó hội tụ. Phương pháp Newton-Raphson thường không hội tụ dễ dàng, đòi hỏi các giá trị ban đầu gần với

giá trị thực và các bước tăng nhỏ, được kiểm soát chặt chẽ. Bề mặt log-likelihood trong mô hình ba tham số tồn tại trong không gian bốn chiều, khó hình dung và mô tả.

Để đảm bảo quá trình ước lượng hiệu quả, cần cân bằng giữa việc tăng các bước ước lượng để tránh sấn tìm cực đại và giảm các bước tăng để tránh hội tụ chậm. Việc này đòi hỏi sự tinh chỉnh cẩn thận trong quá trình lập trình và ước lượng.

2.1.3. Ước lượng đồng thời tham số câu hỏi và năng lực

Rõ ràng là tại một thời điểm nào đó, cả năng lực của thí sinh và các tham số câu hỏi đều không được biết trước. Đây là tình huống phổ biến nhất và cũng là vấn đề khó khăn nhất. Trong trường hợp này, các phản ứng của tất cả các thí sinh đối với tất cả các câu hỏi phải được xem xét đồng thời.

Hàm hợp lý khi n thí sinh trả lời N câu hỏi, sử dụng giả định độc lập cục bộ, có dạng:

$$L(u_1, u_2, \dots, u_j, \dots, u_n | \theta, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \prod_{i=1}^n \prod_{j=1}^N P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}, \quad (2.84)$$

trong đó:

- u_i là mẫu phản ứng của thí sinh i đối với N câu hỏi,
- θ là vector của n tham số năng lực,
- \mathbf{a}, \mathbf{b} , và \mathbf{c} là các vector tham số câu hỏi cho bài kiểm tra gồm N câu hỏi.

Số lượng tham số câu hỏi là $3N$ trong mô hình ba tham số ($2N$ cho mô hình hai tham số và N cho mô hình một tham số). Giả định độc lập cục bộ phải được áp dụng vì các giá trị θ không được biết trước. Số lượng tham số năng lực là n , và do đó, trong mô hình ba tham số, tổng cộng có $3N + n$ tham số cần được ước lượng.

Quy trình ước lượng hợp lý cực đại đồng thời (Joint Maximum Likelihood Estimation - JMLE) là một phương pháp phổ biến để ước lượng đồng thời các tham số câu hỏi và năng lực. Quy trình này được thực hiện qua hai giai đoạn lặp:

- Giai đoạn 1: Ước lượng tham số câu hỏi dựa trên giá trị năng lực ban đầu, thường được khởi tạo từ logarit của tỷ lệ câu trả lời đúng và sai.
- Giai đoạn 2: Ước lượng năng lực dựa trên các tham số câu hỏi đã được ước

lượng.

Quy trình này lặp lại cho đến khi các giá trị ước lượng hội tụ. Tuy nhiên, JMLE có một số hạn chế đáng kể:

- Thứ nhất, nó không thể ước lượng năng lực cho các thí sinh có điểm số hoàn hảo (toàn đúng) hoặc điểm số bằng không (toàn sai),
- Thứ hai, các câu hỏi được trả lời đúng hoặc sai bởi tất cả thí sinh cũng không thể ước lượng được tham số,
- Thứ ba, trong các mô hình hai và ba tham số JMLE không đảm bảo tính nhất quán của ước lượng,
- Thứ tư, trong mô hình ba tham số quy trình số học có thể thất bại nếu không có các ràng buộc phù hợp [18].

Để khắc phục các hạn chế của JMLE, các phương pháp ước lượng thay thế đã được phát triển.

- Phương pháp Bayesian: Phương pháp này sử dụng các phân phối tiên nghiệm (prior distributions) cho các tham số câu hỏi và năng lực. Nhờ đó, nó có thể xử lý các trường hợp cực đoan mà JMLE không thể giải quyết, chẳng hạn như các mẫu phản ứng toàn đúng hoặc toàn sai.
- Phương pháp hợp lý cực đại biên (Marginal Maximum Likelihood Estimation) - MMLE: MMLE khắc phục vấn đề về tính không nhất quán của JMLE bằng cách tích hợp năng lực ra khỏi hàm hợp lý. Cụ thể, năng lực được coi là một biến ngẫu nhiên với phân phối xác định, và hàm hợp lý biên (marginal likelihood function) chỉ phụ thuộc vào các tham số câu hỏi. Phương pháp này đòi hỏi nhiều tính toán hơn do cần thực hiện các phép tích phân phức tạp, nhưng nó cho ra các ước lượng nhất quán khi số lượng thí sinh đủ lớn. MMLE đã được triển khai trong phần mềm BILOG [18].

Sau khi các tham số câu hỏi đã được ước lượng bằng MMLE, năng lực của các thí sinh có thể được ước lượng bằng phương pháp hợp lý cực đại (MLE) hoặc kỳ vọng hậu nghiệm (EAP). Tuy nhiên, trong một số trường hợp, MMLE có thể thất bại, đặc biệt khi ước lượng tham số c (tham số đo lường khả năng đoán mò) trong mô hình ba tham số. Các ước lượng kém của c có thể ảnh hưởng tiêu cực đến chất lượng của các ước lượng khác. Để giải quyết vấn đề này, phương pháp

Bayesian được khuyến nghị, trong đó một phân phối tiên nghiệm được đặt lên các giá trị của c [18].

2.2. Phương pháp ước lượng hợp lý cực đại biên

Từ năm 1968, phương pháp phổ biến nhất để ước lượng tham số câu hỏi trong IRT là JMLE. Tuy nhiên, một nhược điểm lớn của JMLE là nó yêu cầu ước lượng đồng thời cả tham số câu hỏi (structural parameters) và năng lực thí sinh (incidental parameters). Vì số lượng tham số năng lực phụ thuộc vào kích thước mẫu, Neyman và Scott (1948) đã chỉ ra rằng việc ước lượng đồng thời có thể dẫn đến ước lượng không hội tụ khi kích thước mẫu tăng [28].

Một giải pháp được đề xuất là ước lượng hợp lý cực đại có điều kiện (CMLE - Conditional Maximum Likelihood Estimation) do Andersen (1972) phát triển, giúp đảm bảo tính hội tụ của tham số câu hỏi. Tuy nhiên, phương pháp này chỉ áp dụng được cho mô hình Rasch (một tham số - 1PL). Điều này đặt ra nhu cầu về một phương pháp tốt hơn cho các mô hình IRT hai tham số (2PL) và ba tham số (3PL) [1].

Bock và Lieberman (1970) đã giới thiệu ước lượng hợp lý cực đại biên (MMLE - Marginal Maximum Likelihood Estimation) để giải quyết vấn đề này. Tuy nhiên, cách tiếp cận ban đầu của họ có yêu cầu tính toán quá cao, chỉ khả thi với các bài kiểm tra ngắn [12].

Phương pháp ước lượng hợp lý cực đại biên (MMLE - Marginal Maximum Likelihood Estimation) giả định rằng năng lực của thí sinh được lấy ngẫu nhiên từ một tổng thể có phân bố xác suất được đặc trưng bởi một hàm mật độ $g(\theta | \tau)$, trong đó τ là vector tham số của phân bố năng lực [19].

Bock và Lieberman (1970) giải quyết vấn đề bằng cách tích phân theo phân bố năng lực, qua đó loại bỏ các tham số năng lực ra khỏi hàm hợp lý. Điều này giúp việc ước lượng tham số câu hỏi không còn phụ thuộc vào năng lực của từng thí sinh, mà chỉ phụ thuộc vào phân bố năng lực tổng thể [12].

Ưu điểm quan trọng của phương pháp này là đảm bảo ước lượng tham số câu hỏi hội tụ ngay cả khi kích thước mẫu tăng, vì không cần phải ước lượng thêm tham số năng lực của từng thí sinh. Mặc dù MMLE không phải là một phương pháp Bayes, nhưng nó vẫn sử dụng định lý Bayes trong quá trình tính toán [19].

Khác với phương pháp JMLE, trong đó năng lực của mỗi thí sinh được ước lượng bằng một giá trị điểm duy nhất, MMLE thay thế cách tiếp cận này bằng một phân phối xác suất. Cụ thể, năng lực của thí sinh không được xác định bởi một giá trị cố định, mà được xem là một biến ngẫu nhiên có phân bố theo xác suất hậu nghiệm. Xác suất hậu nghiệm này thể hiện mức độ tin cậy của từng giá trị năng lực có thể có của thí sinh, dựa trên:

- Phản hồi câu hỏi của thí sinh (\mathbf{u}_i),
- Tham số câu hỏi (ξ),
- Phân phối năng lực tổng thể của quần thể ($g(\theta | \tau)$).

Thay vì cố gắng tìm một giá trị năng lực duy nhất cho mỗi thí sinh, phương pháp MMLE tích phân theo phân bố năng lực, giúp loại bỏ sự phụ thuộc vào ước lượng cá nhân của từng thí sinh. Điều này khắc phục nhược điểm của JMLE, vốn có thể dẫn đến ước lượng không ổn định khi kích thước mẫu lớn. Quá trình này được thực hiện bằng cách sử dụng định lý Bayes, với công thức xác suất hậu nghiệm của năng lực:

$$P(\theta_i | \mathbf{u}_i, \tau, \xi) = \frac{P(\mathbf{u}_i | \theta_i, \xi) g(\theta_i | \tau)}{\int P(\mathbf{u}_j | \theta_j, \xi) g(\theta_j | \tau) d\theta_j}. \quad (2.85)$$

Với giả định độc lập có điều kiện, xác suất phản hồi của thí sinh có thể được biểu diễn dưới dạng hàm hợp lý (likelihood function):

$$P(\mathbf{u}_i | \theta_i, \xi) = \prod_{j=1}^N P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}}. \quad (2.86)$$

Đây chính là cơ sở toán học giúp MMLE ước lượng tham số câu hỏi một cách ổn định, không phụ thuộc vào việc ước lượng năng lực của từng thí sinh như JMLE [19].

Phân phối năng lực của thí sinh trong quần thể được mô hình hóa thông qua một hàm mật độ xác suất, ký hiệu là $g(\theta_i | \tau)$. Vì chưa biết trước dạng phân phối này, người nghiên cứu phải giả định dựa trên đặc điểm của bài kiểm tra và nhóm thí sinh. Phổ biến nhất là giả định phân phối chuẩn tắc đơn vị, nhưng các dạng phân phối khác cũng có thể được sử dụng. Phân phối này đóng vai trò là phân phối

tiên nghiệm trong thống kê Bayes, tức là mô tả xác suất ban đầu của năng lực trước khi có dữ liệu phản hồi. Giả định quan trọng trong MMLE là tất cả thí sinh đều có chung một phân phối tiên nghiệm của năng lực. Trong phương trình MMLE:

- Tử số chứa tích số của hàm hợp lý $P(\mathbf{u}_i | \theta_i, \xi)$ và phân phối tiên nghiệm $g(\theta_i | \tau)$, biểu diễn phân phối đồng thời (joint distribution) của phản hồi câu hỏi và năng lực thí sinh.
- Mẫu số thực hiện tích phân theo năng lực θ_j , giúp loại bỏ tham số gây nhiễu, tạo ra xác suất biên của phản hồi câu hỏi [19].

Bằng cách sử dụng định lý Bayes, MMLE tính toán phân phối hậu nghiệm của năng lực, phản ánh mức độ tin cậy của từng giá trị năng lực có thể có sau khi quan sát dữ liệu. Phân phối này giúp ước lượng số lượng kỳ vọng của thí sinh tại mỗi mức năng lực trên thang đo, tạo nền tảng cho các tính toán tiếp theo trong MMLE.

2.2.1. Giải pháp của Bock và Lieberman

Ở đây, chúng ta sẽ sử dụng mô hình hàm đặc trưng câu hỏi ba tham số với các tham số a, b, c . Để đơn giản hóa ký hiệu, ta đặt: $P(\mathbf{u}_i) = \int P(\mathbf{u}_i | \theta_i, \xi) g(\theta_i | \tau) d\theta_i$. Theo Bock và Lieberman (1970) [12], hàm hợp lý biên được viết dưới dạng:

$$L = \prod_{i=1}^n P(\mathbf{u}_i). \quad (2.87)$$

Lấy logarit cơ số tự nhiên của L :

$$\ell = \ln L = \sum_{i=1}^n \ln P(\mathbf{u}_i). \quad (2.88)$$

Để tìm phương trình hợp lý biên cho tham số a_j , ta lấy đạo hàm theo a_j :

$$\frac{\partial \ell}{\partial a_j} = 0.$$

Điều này dẫn đến:

$$\begin{aligned}\frac{\partial \ell}{\partial a_j} &= \sum_{i=1}^n \frac{\partial}{\partial a_j} [\ln P(\mathbf{u}_i)] \\ &= \sum_{i=1}^n [P(\mathbf{u}_i)]^{-1} \frac{\partial}{\partial a_j} \left[\int P(\mathbf{u}_i | \theta_i, \xi) g(\theta_i | \tau) d\theta_i \right].\end{aligned}$$

Để thuận tiện, chỉ số dưới i của θ được lược bỏ trong các biểu thức tiếp theo, vì có thể xem θ_i là một biến ngẫu nhiên đại diện cho một thí sinh được chọn ngẫu nhiên từ quần thể. Bằng cách hoán đổi phép vi phân và tích phân, ta có:

$$\frac{\partial}{\partial a_j} \ell = \sum_{i=1}^n \frac{1}{P(\mathbf{u}_i)} \int \frac{\partial}{\partial a_j} [P(\mathbf{u}_j | \theta, \xi)] g(\theta | \tau) d\theta.$$

Bằng cách biến đổi:

$$\frac{\partial}{\partial a_j} P(\mathbf{u}_i | \theta, \xi) = \frac{\partial}{\partial a_j} [\ln P(\mathbf{u}_i | \theta, \xi)] P(\mathbf{u}_i | \theta, \xi),$$

ta thu được

$$\begin{aligned}\frac{\partial}{\partial a_j} \ell &= \sum_{i=1}^n \frac{1}{P(\mathbf{u}_i)} \int \frac{\partial}{\partial a_j} [\ln P(\mathbf{u}_i | \theta, \xi)] P(\mathbf{u}_i | \theta, \xi) g(\theta | \tau) d\theta \\ &= \sum_{i=1}^n \int \frac{\partial}{\partial a_j} [\ln P(\mathbf{u}_i | \theta, \xi)] \left[\frac{P(\mathbf{u}_i | \theta, \xi) g(\theta | \tau)}{P(\mathbf{u}_i)} \right] d\theta.\end{aligned}$$

Về thứ hai trong dấu ngoặc vuông chính là phân phối hậu nghiệm của năng lực $P(\theta | \mathbf{u}_i, \xi, \tau)$ được định nghĩa trong phương trình (2.85). Thay thế (2.85) vào phương trình trên, ta thu được:

$$\frac{\partial}{\partial a_j} \ell = \sum_{i=1}^n \int \frac{\partial}{\partial a_j} [\ln P(\mathbf{u}_i | \theta, \xi)] P(\theta | \mathbf{u}_i, \xi, \tau) d\theta. \quad (2.89)$$

Thay thế $P(\mathbf{u}_i | \theta, \xi)$ trong dấu ngoặc vuông đầu tiên của phương trình (2.89) bằng biểu thức hàm hợp lý được cho trong phương trình (2.86), ta có:

$$\begin{aligned}
\frac{\partial}{\partial a_j} \ell &= \sum_{i=1}^n \int \frac{\partial}{\partial a_j} \left[\ln \prod_{j=1}^N P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}} \right] P(\theta | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta \\
&= \sum_{i=1}^n \int \left[\prod_{j=1}^N P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}} \right]^{-1} \frac{\partial}{\partial a_j} \left[\prod_{j=1}^N P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}} \right] \\
&\quad \times P(\theta | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta.
\end{aligned} \tag{2.90}$$

Tiếp tục, ta đạo hàm

$$\begin{aligned}
\frac{\partial}{\partial a_j} \left[\prod_{j=1}^N P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}} \right] &= \left[\prod_{h \neq j}^N P_h(\theta)^{u_{hj}} Q_h(\theta)^{1-u_{hj}} \right] \\
&\quad \times \frac{\partial}{\partial a_j} [P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}}],
\end{aligned} \tag{2.91}$$

trong đó,

$$\begin{aligned}
\frac{\partial}{\partial a_j} [P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}}] &= Q_j(\theta)^{1-u_{ij}} \frac{\partial}{\partial a_j} [P_j(\theta)^{u_{ij}}] + P_j(\theta)^{u_{ij}} \frac{\partial}{\partial a_j} [Q_j(\theta)^{1-u_{ij}}] \\
&= u_{ij} P_j(\theta)^{u_{ij}-1} Q_j(\theta)^{1-u_{ij}} \frac{\partial P_j(\theta)}{\partial a_j} \\
&\quad + (1-u_{ij}) P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}-1} \frac{\partial Q_j(\theta)}{\partial a_j}.
\end{aligned}$$

Mà $\frac{\partial Q_j(\theta)}{\partial a_j} = -\frac{\partial P_j(\theta)}{\partial a_j}$ và nhóm các hạng tử lại, ta có:

$$\begin{aligned}
&= \frac{\partial P_j(\theta)}{\partial a_j} [P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}}] \left[\frac{u_{ij}}{P_j(\theta)} - \frac{1-u_{ij}}{Q_j(\theta)} \right] \\
&= \frac{\partial P_j(\theta)}{\partial a_j} [P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}}] \left[\frac{u_{ij}(1-Q_j(\theta)) - (1-u_{ij})P_j(\theta)}{P_j(\theta)Q_j(\theta)} \right] \\
&= \frac{\partial P_j(\theta)}{\partial a_j} [P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}}] \left[\frac{u_{ij} - P_j(\theta)}{P_j(\theta)Q_j(\theta)} \right].
\end{aligned}$$

Từ các kết quả trên, ta có:

$$\frac{\partial}{\partial a_j} \left[\prod_{j=1}^N P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}} \right] = \left[\prod_{j=1}^N P_j(\theta)^{u_{ij}} Q_j(\theta)^{1-u_{ij}} \right] \frac{\partial P_j(\theta)}{\partial a_j} \left[\frac{u_{ij} - P_j(\theta)}{P_j(\theta) Q_j(\theta)} \right]. \quad (2.92)$$

Như đã tính toán ở các phần trước, theo phương trình (2.60) thì:

$$\frac{\partial P_j(\theta)}{\partial a_j} = (1 - c_j)(\theta - b_j)P_j^*(\theta)Q_j^*(\theta).$$

Từ các kết quả trên thay vào phương trình (2.90), ta có:

$$\begin{aligned} \frac{\partial}{\partial a_j} \ell &= (1 - c_j) \sum_{i=1}^n \int \left[\prod_{j=1}^N P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}} \right]^{-1} \left[\prod_{j=1}^N P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}} \right] \\ &\quad (\theta_i - b_j) P_j^*(\theta_i) Q_j^*(\theta_i) \left[\frac{u_{ij} - P_j(\theta_i)}{P_j(\theta_i) Q_j(\theta_i)} \right] P(\theta_i | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta_i \\ &= (1 - c_j) \sum_{i=1}^n \int (\theta_i - b_j) [u_{ij} - P_j(\theta_i)] \frac{P_j^*(\theta_i) Q_j^*(\theta_i)}{P_j(\theta_i) Q_j(\theta_i)} P(\theta_i | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta_i \\ &= (1 - c_j) \sum_{i=1}^n \int (\theta_i - b_j) [u_{ij} - P_j(\theta_i)] W_{ij} P(\theta_i | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta_j, \end{aligned} \quad (2.93)$$

với

$$W_{ji} = \frac{P_j^*(\theta_i) Q_j^*(\theta_i)}{P_j(\theta_i) Q_j(\theta_i)}$$

Từ (2.90) và (2.92), ta có các phương trình cho b_i và c_i như sau:

– Theo (2.61), ta có b_j : $\frac{\partial P_j(\theta)}{\partial b_j} = -a_j(1 - c_j)P_j^*(\theta_i)Q_j^*(\theta_i)$, nên:

$$\frac{\partial}{\partial a_j} \ell = -a_j(1 - c_j) \sum_{i=1}^n \int [u_{ij} - P_j(\theta_i)] W_{ij} P(\theta_i | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta_i. \quad (2.94)$$

– Theo (2.62), ta có c_j : $\frac{\partial P_j(\theta_i)}{\partial c_j} = \frac{Q_i(\theta_i)}{1 - c_j}$, nên:

$$\frac{\partial}{\partial a_j} \ell = \frac{1}{1 - c_j} \sum_{i=1}^n \int \frac{u_{ij} - P_j(\theta_i)}{P_j(\theta_i)} P(\theta_i | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) d\theta_i. \quad (2.95)$$

Vì việc tính toán tích phân trên máy tính kỹ thuật số là rất phức tạp, nên cần một phương pháp để xấp xỉ các tích phân này. Ở đây, kỹ thuật Hermite-Gauss quadrature có thể được áp dụng để xấp xỉ các tích phân dạng này. Nếu $g(\theta|\tau)$ là một phân phối liên tục với các moment hữu hạn, nó có thể được xấp xỉ đến mức độ chính xác mong muốn bằng một phân phối rời rạc trên một số điểm hữu hạn (tương tự như biểu đồ tần suất). Dưới đây là các bước thực hiện phương pháp xấp xỉ quadrature:

- Bước 1: Thay thế tích phân liên tục bằng tổng rời rạc: Thay vì tính toán diện tích dưới đường cong liên tục, phương pháp này thay thế bài toán bằng việc tính tổng diện tích của một số hữu hạn các hình chữ nhật xấp xỉ diện tích dưới đường cong.
- Bước 2: Xác định các nút (nodes) và trọng số
 - + Trên thang đo năng lực, các điểm X_k (với $k = 1, 2, \dots, q$) được gọi là các nút (nodes).
 - + Mỗi nút X_k có một trọng số liên kết $A(X_k)$, phản ánh chiều cao của hàm mật độ $g(\theta|\tau)$ trong lân cận của X_k và chiều rộng của các hình chữ nhật.
- Bước 3: Tính toán giá trị nút và trọng số: Các giá trị của X_k và $A(X_k)$ được xác định bằng cách giải một hệ phương trình liên quan đến phân phối liên tục cần xấp xỉ và số lượng nút được chọn.
- Bước 4: Điều chỉnh cho phân phối chuẩn đơn vị: Nếu cần xấp xỉ phân phối chuẩn đơn vị, các giá trị X_k trong bảng được nhân với $\sqrt{2}$, và các giá trị $A(X_k)$ được chia cho $\sqrt{\pi}$ (theo Bock và Lieberman, 1970 [12]).

Phương pháp này không yêu cầu $g(\theta|\tau)$ phải có dạng phân phối chuẩn. Nó có thể được áp dụng cho các phân phối được định nghĩa thực nghiệm. Các giá trị X_k tương tự như các mức năng lực đã biết θ_j [19].

Đối với một đối tượng ngẫu nhiên (không có chỉ số i) được lấy mẫu từ một quần thể với phân phối $g(\theta|\tau)$, việc thay thế tích phân trong phương trình (2.93) bằng phương pháp xấp xỉ quadrature dẫn đến kết quả sau:

$$a_j : (1 - c_j) \sum_{i=1}^n \sum_{k=1}^q [u_{ij} - P_j(X_k)] W_{jk}(X_k - b_j) [P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau})] = 0, \quad (2.96)$$

trong đó:

- $W_{jk} = \frac{P_j^*(X_k)Q_j^*(X_k)}{P_j(X_k)Q_j(X_k)}$ là trọng số,
- θ được thay thế bằng các nút X_k .

Từ phương trình trên, dạng quadrature của phương trình MMLE cho tham số độ khó b_j có thể được viết như sau:

$$b_j : -a_j(1 - c_j) \sum_{i=1}^n \sum_{k=1}^q [u_{ij} - P_j(X_k)] W_{jk} [P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau})] = 0. \quad (2.97)$$

Tương tự, phương trình cho tham số c_j có thể được viết:

$$c_j : (1 - c_j)^{-1} \sum_{i=1}^n \sum_{k=1}^q \left[\frac{u_{ij} - P_j(X_k)}{P_j(X_k)} \right] [P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau})] = 0. \quad (2.98)$$

Các biểu thức 2.96, 2.97, 2.98 đại diện cho dạng quadrature của các phương trình hợp lý biên cho mô hình logistic ba tham số. Theo cách tiếp cận của Bock và Lieberman (1970) [12], kỹ thuật Newton-Raphson được sử dụng để ước lượng đồng thời $3n$ tham số của một bài kiểm tra. Tương tự như trong JMLE, các giá trị quan sát u_{ij} trong đạo hàm bậc hai của hàm log-likelihood được thay thế bằng các giá trị kỳ vọng của chúng. Do đó, phương pháp Fisher's scoring thực sự được sử dụng. Tuy nhiên, phương pháp này không hấp dẫn về mặt tính toán vì nó đòi hỏi phải đảo ngược một ma trận có kích thước $3n \times 3n$, điều này có thể tốn kém và phức tạp khi số lượng tham số lớn [19].

2.2.2. Giải pháp của Bock và Aitkin

Bock và Aitkin (1981) [6] đã cải tiến phương trình hợp lý của Bock và Lieberman, tạo ra một phương pháp có thể tính toán hiệu quả hơn. Giải pháp

này, khi giả định rằng phân phối của quần thể là đã biết hoặc được ước lượng đồng thời với mô hình chính xác, sẽ cho ra ước lượng tham số câu hỏi hội tụ ổn định. Bock và Aitkin, giả định rằng:

- các câu hỏi là độc lập,
- các thí sinh là độc lập,
- câu hỏi và thí sinh là độc lập với nhau.

Nhờ giả định này, các tham số của từng câu hỏi có thể được ước lượng riêng lẻ từng câu một, trong khi năng lực của từng thí sinh có thể ước lượng trên cơ sở từng thí sinh riêng biệt. Ta bắt đầu từ phương trình hợp lý cho tham số a_j (2.96):

$$a_j : \quad (1 - c_j) \sum_{k=1}^q (X_k - b_j) \times \left[\sum_{i=1}^n u_{ij} P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) - P_j(X_k) \sum_{i=1}^n P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) \right] W_{jk} = 0. \quad (2.99)$$

Đưa phương trình (2.85) về dạng phép tính tổng quát (quadrature form) ta được:

$$P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) = \frac{\prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} A(X_k)}{\sum_{k=1}^q \prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} A(X_k)}. \quad (2.100)$$

Tiếp tục, có thể định nghĩa biểu thức sau:

$$\bar{f}_{jk} = \sum_{i=1}^n P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) = \sum_{i=1}^n \left[\frac{\prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} A(X_k)}{\sum_{k=1}^q \prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} A(X_k)} \right], \quad (2.101)$$

biểu thức trên đơn giản chỉ là tổng xác suất hậu nghiệm của một thí sinh có năng

lực X_k trên toàn bộ n thí sinh. Đặt

$$\bar{r}_{jk} = \sum_{i=1}^n u_{ij} P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) = \sum_{i=1}^n \left[\frac{\prod_{j=1}^N u_{ij} P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} A(X_k)}{\sum_{k=1}^q \prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} A(X_k)} \right]. \quad (2.102)$$

Ta đặt,

$$L(X_k) = \prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}} \quad (2.103)$$

để biểu diễn dạng tổng quát của xác suất có điều kiện của \mathbf{u}_i với điều kiện $\theta = X_k$ và các tham số của câu hỏi. Khi đó, các phương trình (2.100), (2.101), (2.102) được viết lại thành như sau:

$$P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) = \frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}, \quad (2.104)$$

và

$$\bar{f}_{jk} = \sum_{i=1}^n P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) = \sum_{i=1}^n \left[\frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \right], \quad (2.105)$$

$$\bar{r}_{jk} = \sum_{i=1}^n u_{ij} P(X_k | \mathbf{u}_i, \boldsymbol{\xi}, \boldsymbol{\tau}) = \sum_{i=1}^n \left[\frac{u_{ij} L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \right]. \quad (2.106)$$

Hai thuật ngữ \bar{r}_{jk} và \bar{f}_{jk} có ý nghĩa như sau:

- \bar{f}_{jk} : Đây là số lượng thí sinh trong quần thể (cỡ mẫu n) được kỳ vọng có mức năng lực X_k . Giá trị này được tính bằng cách phân bổ phản hồi của mỗi thí sinh lên các q nút tích phân (quadrature nodes) dựa trên xác suất hậu nghiệm của thí sinh đó tại từng nút. Nói cách khác, \bar{f}_{jk} cho biết có bao

nhiều thí sinh được ước lượng là có năng lực X_k sau khi phân tích dữ liệu phản hồi.

- \bar{r}_{jk} : Đây là số lượng thí sinh có mức năng lực X_k được kỳ vọng trả lời đúng câu hỏi.

Sự khác biệt giữa hai phương trình (2.105) và (2.106) nằm ở biến u_{ij} xuất hiện trong phương trình (2.106). Vì u_{ij} chỉ có giá trị 1 (nếu thí sinh trả lời đúng) hoặc 0 (nếu thí sinh trả lời sai), nên khi tính tổng, chỉ các thí sinh có câu trả lời đúng mới được cộng vào. Điều này giúp xác định tần suất thí sinh trả lời đúng tại mỗi mức năng lực X_k , tức là giá trị r_{jk} [19].

Với các đại lượng \bar{r}_{jk} và \bar{f}_{jk} đã được xác định, dạng tích phân của phương trình hợp lý biên có thể được viết như sau:

$$a_j : \quad (1 - c_j) \sum_{k=1}^q (X_k - b_j) \left[\bar{r}_{jk} - \bar{f}_{jk} P_j(X_k) \right] W_{jk} = 0 \quad (2.107)$$

$$b_j : \quad -a_j(1 - c_j) \sum_{k=1}^q \left[\bar{r}_{jk} - \bar{f}_{jk} P_j(X_k) \right] W_{jk} = 0 \quad (2.108)$$

$$c_j : \quad (1 - c_j)^{-1} \sum_{k=1}^q \left[\frac{\bar{r}_{jk} - \bar{f}_{jk} P_j(X_k)}{P_j(X_k)} \right] = 0 \quad (2.109)$$

Trong phương pháp MMLE, tổng số thí sinh ở mức năng lực X_k và số thí sinh trả lời đúng không được quan sát trực tiếp. Thay vào đó, chúng được thay thế bằng kỳ vọng về số lượng thí sinh tham gia trả lời câu hỏi j và số lượng kỳ vọng của thí sinh trả lời đúng tại mỗi mức năng lực X_k . Các giá trị kỳ vọng này được tính toán thông qua các xác suất hậu nghiệm, dựa trên mô hình ước lượng tham số câu hỏi và năng lực thí sinh. Điều này giúp phương pháp MMLE có thể hoạt động ngay cả khi không có thông tin trực tiếp về phân bố năng lực thí sinh, miễn là có giả định hợp lý về phân phối tiềm ẩn của θ . Trong IRT, các đại lượng \bar{f}_{jk} và \bar{r}_{jk} được gọi là “dữ liệu nhân tạo”, vì chúng không được thu thập trực tiếp mà được tính toán thông qua các phương trình (2.105) và (2.106) [19].

Sau đây sẽ tóm tắt lại bốn bước trong phương pháp Bock và Aitkin cho MMLE của một câu hỏi:

- Bước 1: Tính toán xác suất hợp lý của từng thí sinh: Sử dụng phương trình

(2.103) với

$$P_j(X_k) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(X_k - b_j)}}$$

và

$$Q_j(X_k) = 1 - P_j(X_k)$$

để tính toán likelihood của vector điểm số câu hỏi của mỗi thí sinh tại mỗi nút cầu phương q . Điều này giúp xác định xác suất quan sát được điểm số của thí sinh dựa trên năng lực tại các nút cầu phương.

- Bước 2: Tính toán xác suất hậu nghiệm của năng lực thí sinh: Sử dụng biểu thức (2.104) và trọng số cầu phương $A(X_k)$ tại mỗi nút q để tính toán xác suất hậu nghiệm rằng năng lực của thí sinh thứ i là X_k . Xác suất này phản ánh khả năng thí sinh có năng lực X_k khi biết điểm số của họ.
- Bước 3: Tính toán các giá trị kỳ vọng: Dùng phương trình (2.105) và (2.106) để tạo \bar{f}_{ik} và \bar{r}_{ik} , lần lượt là số lượng kỳ vọng của thí sinh tham gia trả lời câu hỏi j và số lượng kỳ vọng của thí sinh trả lời đúng tại mỗi nút năng lực q . Các giá trị này giúp ước lượng tần suất thí sinh trả lời câu hỏi và tỷ lệ trả lời đúng tại các mức năng lực khác nhau.
- Bước 4: Giải phương trình hợp lý và cập nhật tham số: Giải các phương trình hợp lý (2.107, 2.108 và 2.109) để ước lượng tham số câu hỏi bằng cách sử dụng các giá trị \bar{r}_{jk} và \bar{n}_{jk} . Vì các giá trị này phụ thuộc vào các thành phần $P(X_k | \mathbf{u}_i, \boldsymbol{\epsilon}, \boldsymbol{\tau})$, mà các thành phần này lại phụ thuộc vào các tham số câu hỏi chưa biết, nên các phương trình hợp lý là phương trình ẩn và cần được giải lặp (ví dụ: sử dụng chuỗi Taylor và phương pháp Newton-Raphson). Nếu các ước lượng tham số câu hỏi thay đổi so với chu kỳ trước, quay lại Bước 1 với các ước lượng tham số mới nhất.

Sử dụng 10 giá trị X_k trong quy trình cầu phương Gauss, ta có bảng:

Node	X_k	$A(X_k)$
1	-4.000	0.000119
2	-3.111	0.002805
3	-2.222	0.030002
4	-1.333	0.145800
5	-0.444	0.321300
6	0.444	0.321300
7	1.333	0.145800
8	2.222	0.030002
9	3.111	0.002805
10	4.000	0.000119

Bảng 2.2: Bảng giá trị X_k và $A(X_k)$ [19]

Đối với mô hình 1 tham số ta có phương trình:

$$b_j : - \sum_{k=1}^q \left[\bar{r}_{jk} - \bar{f}_{jk} P_j(X_k) \right] W_{jk} = 0.$$

Đối với mô hình 2 tham số ta có phương trình:

$$a_j : \sum_{k=1}^q (X_k - b_j) \left[\bar{r}_{jk} - \bar{f}_{jk} P_j(X_k) \right] W_{jk} = 0$$

$$b_j : -a_j \sum_{k=1}^q \left[\bar{r}_{jk} - \bar{f}_{jk} P_j(X_k) \right] W_{jk} = 0.$$

2.3. Phương pháp ước lượng hậu nghiệm cực đại

Phương pháp Maximum A Posteriori estimation-MAP (hay còn gọi là Bayes Modal Estimation-BME) là một cách tiếp cận Bayesian, trong đó các tham số được ước lượng bằng cách tìm giá trị tối đa của phân phối hậu nghiệm. Phân phối hậu nghiệm này kết hợp thông tin từ dữ liệu quan sát với thông tin tiên nghiệm về các tham số. Cụ thể:

- Prior distribution: Phân phối tiên nghiệm phản ánh niềm tin ban đầu về các

tham số trước khi quan sát dữ liệu. Ví dụ, năng lực của thí sinh có thể được giả định tuân theo phân phối chuẩn $\theta \sim N(0, 1)$.

- Likelihood: Hàm hợp lý dựa trên mô hình IRT, chẳng hạn như mô hình Rasch hoặc mô hình 2PL, 3PL.
- Posterior distribution: Phân phối hậu nghiệm là kết quả của việc kết hợp thông tin từ prior và likelihood, được tính toán bằng định lý Bayes.

Phương pháp MAP tìm giá trị của các tham số sao cho phân phối hậu nghiệm đạt giá trị tối đa. So với phương pháp JMLE, MAP có ưu điểm là tránh được các vấn đề về tính không nhất quán và có thể xử lý tốt hơn các trường hợp cực đoan (ví dụ: thí sinh trả lời đúng hoặc sai toàn bộ câu hỏi). Và MAP không cần tích phân ra biến năng lực như MMLE, nhưng vẫn đảm bảo tính nhất quán và độ tin cậy của ước lượng nhờ việc kết hợp thông tin tiên nghiệm.

Sau đây sẽ trình bày về định lý Bayes, đầu tiên coi các tham số θ là chưa biết, và dữ liệu \mathcal{D} là cố định và đã biết, khi đó

$$p(\theta | \mathcal{D}) = \frac{p(\theta)p(\mathcal{D}|\theta)}{p(\mathcal{D})} \quad (2.110)$$

trong đó:

- Phân phối tiên nghiệm $p(\theta)$ (prior): Niềm tin ban đầu về tham số θ trước khi có dữ liệu.
- Khả năng $p(\mathcal{D}|\theta)$ (likelihood): Mô tả khả năng quan sát dữ liệu \mathcal{D} với tham số θ .
- Phân phối hậu nghiệm $p(\theta|\mathcal{D})$ (posterior): Niềm tin về tham số θ sau khi có dữ liệu \mathcal{D} .
- Khả năng biên $p(\mathcal{D})$ (evidence): Hằng số chuẩn hóa để đảm bảo tính hợp lệ của phân phối hậu nghiệm.

Theo Prince (2012)[29], phương pháp MAP được trình bày như sau. Trong MAP, chúng ta đưa vào thông tin tiên nghiệm về các tham số θ . Thông tin này có thể dựa trên kinh nghiệm trước đây, ví dụ như trong chuỗi thời gian, giá trị của tham số tại thời điểm t có thể cung cấp thông tin quan trọng về giá trị tại thời điểm $t + 1$, và thông tin này được mã hóa trong phân phối tiên nghiệm.

Mục tiêu của MAP là tối đa hóa xác suất hậu nghiệm $P(\theta | \mathbf{x}_{1...N})$ của các tham số:

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} P(\theta | \mathbf{x}_{1...N}) \\ &= \arg \max_{\theta} \left[\frac{P(\mathbf{x}_{1,...,N} | \theta) P(\theta)}{P(\mathbf{x}_{1,...,N})} \right] \\ &= \arg \max_{\theta} \left[\frac{\prod_{i=1}^N P(\mathbf{x}_i | \theta) P(\theta)}{P(\mathbf{x}_{1,...,N})} \right],\end{aligned}$$

sử dụng định lý Bayes và giả định tính độc lập, ta có thể đơn giản hóa biểu thức trên bằng cách loại bỏ mẫu số (vì nó không ảnh hưởng đến vị trí cực đại):

$$\hat{\theta}_{MAP} = \arg \max_{\theta} \left[\prod_{i=1}^N P(\mathbf{x}_i | \theta) P(\theta) \right]. \quad (2.111)$$

So sánh với phương pháp hợp lý cực đại MLE, MAP chỉ khác biệt ở việc thêm thành phần tiên nghiệm $P(\theta)$. Trong trường hợp phân phối tiên nghiệm là không thông tin (uninformative), MAP trở thành MLE.

Sau khi ước lượng tham số, xác suất dự đoán cho một dữ liệu mới \mathbf{x}^* được tính bằng cách đánh giá hàm mật độ xác suất $P(\mathbf{x}^* | \hat{\theta})$ dựa trên các tham số mới. Phương pháp MAP kết hợp thông tin từ dữ liệu quan sát và thông tin tiên nghiệm, giúp cải thiện độ chính xác và ổn định của ước lượng, đặc biệt trong các trường hợp dữ liệu không đầy đủ hoặc nhiễu.

2.3.1. Ước lượng năng lực của thí sinh

Với thí sinh thứ i trả lời N câu hỏi, theo phương pháp MAP, ta có hàm hợp lý:

$$\begin{aligned}L &= \arg \max_{\theta} P(\theta, \mathbf{a}, \mathbf{b}, \mathbf{c} | \mathbf{u}) \\ &= \arg \max_{\theta} P(\mathbf{u} | \theta, \mathbf{a}, \mathbf{b}, \mathbf{c}) P(\theta),\end{aligned} \quad (2.112)$$

trong đó:

– $\mathbf{u} = [u_1, u_2, \dots, u_N]$ là vector của các phản ứng đã quan sát,

- $\mathbf{a} = [a_1, a_2, \dots, a_N]$ là vector độ phân biệt của N câu hỏi,
- $\mathbf{b} = [b_1, b_2, \dots, b_N]$ là vector độ khó của N câu hỏi,
- $\mathbf{c} = [c_1, c_2, \dots, c_N]$ là vector độ đoán mò của N câu hỏi.

Đầu tiên, chúng ta giả định rằng tham số θ được phân phối độc lập và giống nhau, nghĩa là thông tin về các tham số này có thể hoán đổi được, và hơn nữa chúng được phân phối chuẩn [32]. Do đó, chúng ta có:

$$P(\theta) \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2}\right). \quad (2.113)$$

Tiếp theo, với hàm khả năng, ta có:

$$\begin{aligned} P(\mathbf{u} \mid \theta, \mathbf{a}, \mathbf{b}, \mathbf{c}) &= \prod_{j=1}^N P_j(u_j \mid \theta)^{u_j} [1 - P_j(u_j \mid \theta)]^{1-u_j} \\ &= \prod_{j=1}^N P_j^{u_j} Q_j^{1-u_j}. \end{aligned} \quad (2.114)$$

Tiếp theo, ta lấy logarit cơ số tự nhiên cho hàm L , ta được:

$$\begin{aligned} \ell &= \sum_{j=1}^N [u_j \ln P_j + (1 - u_j) \ln(1 - P_j)] + \ln \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2}\right) \\ &= \sum_{j=1}^N [u_j \ln P_j + (1 - u_j) \ln(1 - P_j)] - \ln \sigma\sqrt{2\pi} - \frac{(\theta - \mu)^2}{2\sigma^2} \\ &\propto \sum_{j=1}^N [u_j \ln P_j + (1 - u_j) \ln(1 - P_j)] - \frac{(\theta - \mu)^2}{2\sigma^2}. \end{aligned} \quad (2.115)$$

Tiếp theo, ta lấy đạo hàm của hàm hợp lý, ta được:

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_i} &= \sum_{j=1}^N \left[u_j \frac{1}{P_j} \frac{\partial P_j}{\partial \theta_i} + (1 - u_j) \frac{1}{Q_j} \frac{\partial Q_j}{\partial \theta_i} \right] + \frac{\partial}{\partial \theta_i} \left(-\frac{(\theta_i - \mu)^2}{2\sigma^2} \right) \\ &= \sum_{j=1}^N \left[u_j \frac{1}{P_j} \frac{\partial P_j}{\partial \theta_i} + (1 - u_j) \frac{1}{Q_j} \frac{\partial Q_j}{\partial \theta_i} \right] - \frac{1}{2\sigma^2} 2(\theta_i - \mu) \frac{\partial}{\partial \theta_i} (\theta_i - \mu) \end{aligned}$$

$$= \sum_{j=1}^N \left[u_j \frac{1}{P_j} \frac{\partial P_j}{\partial \theta_i} + (1 - u_j) \frac{1}{Q_j} \frac{\partial Q_j}{\partial \theta_i} \right] - \frac{\theta_i - \mu}{\sigma^2} \quad (2.116)$$

Sử dụng thuật toán Newton-Raphson, ta được:

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t - \left[\frac{\partial^2 \ell}{\partial \theta_i^2} \right]_t^{-1} \left[\frac{\partial \ell}{\partial \theta_i} \right]_t. \quad (2.117)$$

2.3.1.1. Mô hình IRT 1 tham số

Sử dụng kết quả (2.9) và (2.10) thay vào phương trình (2.116), ta được:

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_i} &= \sum_{j=1}^N \left[u_j \frac{1}{P_j} P_j Q_j + (1 - u_j) \frac{1}{Q_j} (-P_j Q_j) \right] - \frac{\theta_i - \mu}{\sigma^2} \\ &= \sum_{j=1}^N [u_j - P_j] - \frac{\theta_i - \mu}{\sigma^2}. \end{aligned} \quad (2.118)$$

Tiếp theo đạo hàm bậc hai của hàm hợp lý ta được:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \theta_i^2} &= \frac{\partial}{\partial \theta_i} \left[\sum_{j=1}^N (u_j - P_j) - \frac{\theta_i - \mu}{\sigma^2} \right] \\ &= \sum_{j=1}^N -P_j Q_j - \frac{\partial}{\partial \theta_i} \frac{\theta_i - \mu}{\sigma^2} \\ &= \sum_{j=1}^N -P_j Q_j - \frac{1}{\sigma^2} \frac{\partial}{\partial \theta_i} (\theta_i - \mu) \\ &= \sum_{j=1}^N -P_j Q_j - \frac{1}{\sigma^2}. \end{aligned} \quad (2.119)$$

Từ (2.118) và (2.119) thay vào phương trình (2.117), ta thu được phương trình

Newton-Raphson ước lượng θ_i cho mô hình IRT 1 tham số là:

$$\begin{aligned} [\hat{\theta}_i]_{t+1} &= [\hat{\theta}_i]_t - \frac{\sum_{j=1}^N [u_j - P_j] - \frac{\theta_i - \mu}{\sigma^2}}{\sum_{j=1}^N -P_j Q_j - \frac{1}{\sigma^2}} \\ &= [\hat{\theta}_i]_t + \frac{\sum_{j=1}^N [u_j - P_j] - \frac{\theta_i - \mu}{\sigma^2}}{\sum_{j=1}^N P_j Q_j + \frac{1}{\sigma^2}}. \end{aligned} \quad (2.120)$$

Và có sai số chuẩn là:

$$SE(\hat{\theta}_i) = \frac{1}{\sqrt{\sum_{j=1}^N P_j Q_j + \frac{1}{\sigma^2}}} \quad (2.121)$$

2.3.1.2. Mô hình IRT 2 tham số

Theo kết quả của phương trình (2.17), ta có đạo hàm bậc nhất của hàm hợp lý:

$$\frac{\partial}{\partial \theta_i} \ell = \sum_{j=1}^N [a_j(u_j - P_j)] - \frac{\theta_i - \mu}{\sigma^2}. \quad (2.122)$$

Sử dụng kết quả của phương trình (2.18) và phương trình (2.119), ta có đạo hàm bậc hai của phương trình hợp lý như sau:

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i^2} \ell &= \frac{\partial}{\partial \theta_i} \left\{ \sum_{j=1}^N [a_j(u_j - P_j)] - \frac{\theta_i - \mu}{\sigma^2} \right\} \\ &= \sum_{j=1}^N -a_j^2 P_j Q_j - \frac{1}{\sigma^2}. \end{aligned} \quad (2.123)$$

Thay (2.122) và (2.123) vào phương trình (2.117), ta có phương trình lặp Newton-Raphson như sau:

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t + \frac{\sum_{j=1}^N [a_j(u_j - P_j)] - \frac{\theta_i - \mu}{\sigma^2}}{\sum_{j=1}^N a_j^2 P_j Q_j + \frac{1}{\sigma^2}}. \quad (2.124)$$

Và có sai số chuẩn là:

$$SE(\hat{\theta}_i) = \frac{1}{\sqrt{\sum_{j=1}^N a_j^2 P_j Q_j + \frac{1}{\sigma^2}}} \quad (2.125)$$

2.3.1.3. Mô hình IRT 3 tham số

Sử dụng kết quả của phương trình (2.26), phương trình (2.117) cho mô hình 3 tham số có dạng:

$$\frac{\partial}{\partial \theta_i} \ell = \sum_{j=1}^N a_j \frac{P_j - c_j}{1 - c_j} \frac{u_j - P_j}{P_j} - \frac{\theta_i - \mu}{\sigma^2}. \quad (2.126)$$

Sử dụng kết quả của phương trình (2.28) và phương trình (2.119), ta có đạo hàm bậc hai của phương trình hợp lý như sau:

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i^2} \ell &= \frac{\partial}{\partial \theta_i} \left\{ \sum_{j=1}^N a_j \frac{P_j - c_j}{1 - c_j} \frac{u_j - P_j}{P_j} - \frac{\theta_i - \mu}{\sigma^2} \right\} \\ &= \sum_{j=1}^N -a_j^2 \frac{Q_j}{P_j} \frac{(P_j - c_j)^2}{(1 - c_j)^2} - \frac{1}{\sigma^2}. \end{aligned} \quad (2.127)$$

Thay (2.126) và (2.127) vào phương trình (2.117), ta có phương trình lặp Newton-Raphson như sau:

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t + \frac{\sum_{j=1}^N a_j \frac{P_j - c_j}{1 - c_j} \frac{u_j - P_j}{P_j} - \frac{\theta_i - \mu}{\sigma^2}}{\sum_{j=1}^N a_j^2 \frac{Q_j}{P_j} \frac{(P_j - c_j)^2}{(1 - c_j)^2} + \frac{1}{\sigma^2}}. \quad (2.128)$$

2.3.2. Ước lượng các tham số của câu hỏi

Với câu hỏi thứ j được trả lời bởi n thí sinh, theo phương pháp MAP, ta có hàm hợp lý:

$$\begin{aligned} L &= \arg \max_{a,b,c} P(\boldsymbol{\theta}, a, b, c | \mathbf{u}) \\ &= \arg \max_{a,b,c} P(\mathbf{u} | \boldsymbol{\theta}, a, b, c) P(a) P(b) P(c), \end{aligned} \quad (2.129)$$

trong đó:

- $\mathbf{u} = [u_1, u_2, \dots, u_n]$ là vector của các phản ứng đã quan sát,
- $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_n]$ là vector năng lực của n thí sinh.

Tiếp theo, chúng ta sẽ xác định phân phối tiên nghiệm cho a, b, c . Tương tự như θ , các tham số độ khó b được giả định là độc lập và đồng nhất, tuân theo phân phối chuẩn:

$$P(b) \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(b - \mu)^2}{2\sigma^2}\right). \quad (2.130)$$

Việc xác định phân phối tiên nghiệm cho các tham số phân biệt a là một vấn đề phức tạp. Tham số a đại diện cho độ dốc của đường cong đặc trưng câu hỏi tại điểm uốn do đó gần như chắc chắn là dương và bị giới hạn cách xa 0. Vì chỉ các giá trị dương của a là có ý nghĩa, điều này gợi ý rằng phân phối tiên nghiệm nên thuộc họ Gamma[31].

Theo Murphy (2022) [27], phân phối Gamma là một phân phối linh hoạt cho các biến ngẫu nhiên nhận giá trị thực dương, $x > 0$. Nó được định nghĩa bởi hai tham số, gọi là hình dạng $a > 0$ và tỷ lệ $b > 0$:

$$Ga(x \mid \text{shape} = a, \text{rate} = b) \triangleq \frac{b^a}{\Gamma(a)} x^{a-1} e^{-xb}.$$

Đôi khi phân phối được tham số hóa theo hình dạng a và tỷ lệ $s = \frac{1}{b}$

$$Ga(x \mid \text{shape} = a, \text{rate} = s) \triangleq \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-\frac{x}{s}}.$$

Phân phối Chi-bình phương được định nghĩa bởi:

$$\chi_v^2(x) \triangleq \text{Ga}(x \mid \text{shape} = \frac{v}{2}, \text{rate} = \frac{1}{2})$$

trong đó v được gọi là bậc tự do. Đây là phân phối của tổng bình phương các biến ngẫu nhiên Gaussian. Cụ thể hơn, nếu $Z_i \sim \mathcal{N}(0, 1)$, và $S = \sum_{i=1}^v Z_i^2$, thì $S \sim \chi_v^2$.

Theo trên, thì phân phối tiên nghiệm của a theo phân phối Chi-bình phương có dạng:

$$P(a) = \frac{\left(\frac{1}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} a^{v/2-1} e^{-\frac{a}{2}} \quad (2.131)$$

với $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Mà, khi độ lệch chuẩn σ của phân phối năng lực tăng, độ dốc a của ICC thường giảm. Điều này có nghĩa là a tỷ lệ nghịch với σ , tức là $a \propto \frac{1}{\sigma}$ [31]. Vì, nếu phân phối năng lực của thí sinh quá rộng (độ lệch chuẩn lớn), câu hỏi sẽ khó phân biệt giữa các thí sinh có năng lực khác nhau, dẫn đến độ dốc a thấp. Ngược lại, nếu phân phối năng lực tập trung (độ lệch chuẩn nhỏ), câu hỏi sẽ phân biệt tốt hơn, dẫn đến độ dốc a cao. Do đó để phù hợp, ở đây phân phối tiên nghiệm của a sẽ chọn là phân phối Chi. Phân phối Chi được định nghĩa là căn bậc

hai của phân phối Chi-bình phương, nên ta có hàm mật độ xác suất sau:

$$\begin{aligned}
P(a) &= 2a \frac{\left(\frac{1}{2}\right)^{v/2}}{\Gamma\left(\frac{v}{2}\right)} a^{2(v/2-1)} e^{-\frac{a^2}{2}} \\
&= 2^{1-v/2} \left[\Gamma\left(\frac{v}{2}\right)\right]^{-1} a^{v-1} e^{-\frac{a^2}{2}} \\
&= \left[2^{v/2-1} \Gamma\left(\frac{v}{2}\right)\right]^{-1} a^{v-1} \exp\left(-\frac{a^2}{2}\right) \\
&\propto a^{v-1} \exp\left(-\frac{a^2}{2}\right).
\end{aligned} \tag{2.132}$$

Như đã trình bày thì tham số độ đoán mò, c rõ ràng bị giới hạn trên bởi 1 và dưới bởi 0. Phân phối tiên nghiệm cho c có thể được chọn là phân phối Beta [32], chúng ta có:

$$P(c) \sim \text{Beta}(c \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} c^{\alpha-1} (1-c)^{\beta-1}$$

Với $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha\beta)}$ và $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. Vì $B(\alpha, \beta)$ là hệ số, nên:

$$P(c) \propto c^{\alpha-1} (1-c)^{\beta-1} \tag{2.133}$$

Tương tự như phương pháp MLE, trong quá trình ước lượng tham số câu hỏi, chia tập hợp thí sinh thành G nhóm trên thang đo năng lực, trong đó các thí sinh thuộc cùng một nhóm g sẽ có cùng giá trị năng lực θ_g , theo phương trình (2.36), ta có hàm khả năng:

$$P(\mathbf{u} \mid \boldsymbol{\theta}, a, b, c) = \prod_{g=1}^G \frac{f_g!}{r_g! (f_g - r_g)!} P(\theta_g)^{r_g} Q(\theta_g)^{f_g - r_g},$$

với r_g là số thí sinh trả lời đúng trong nhóm g , f_g là tổng số thí sinh trong nhóm g .

Chúng ta có hàm hợp lý theo phương pháp MAP, có dạng như sau:

$$L = \prod_{g=1}^G \frac{f_g!}{r_g!(f_g - r_g)!} P(\theta_g)^{r_g} Q(\theta_g)^{f_g - r_g} \times P(a_j) \times P(b_j) \times P(c_j). \quad (2.134)$$

Tiếp theo ta lấy logarit cơ số tự nhiên cho L , ta được:

$$\ell = \ln P(\mathbf{u} \mid \boldsymbol{\theta}, a_j, b_j, c_j) + \ln P(a_j) + \ln P(b_j) + \ln P(c_j), \quad (2.135)$$

với

$$\begin{aligned} -\ln P(\mathbf{u} \mid \boldsymbol{\theta}, a_j, b_j, c_j) &= \text{constant} + \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)], \\ -\ln P(a_j) &\propto (v-1) \ln a_j - \frac{a_j^2}{2}, \\ -\ln P(b_j) &\propto -\frac{(b_j - \mu)^2}{2\sigma^2}, \\ -\ln P(c_j) &\propto (\alpha-1) \ln c_j + (\beta-1) \ln(1-c_j) \end{aligned}$$

2.3.2.1. Mô hình IRT 1 tham số

Đầu tiên, ta đạo hàm bậc 1 hàm hợp lý theo b :

$$\begin{aligned} \frac{\partial}{\partial b_j} \ell &= \frac{\partial}{\partial b_j} \left\{ \text{constant} + \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] + \ln P(b_j) \right\} \\ &= \frac{\partial}{\partial b_j} \left\{ \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] \right\} + \frac{\partial}{\partial b_j} \left[-\frac{(b_j - \mu)^2}{2\sigma^2} \right]. \end{aligned} \quad (2.136)$$

Đầu tiên ta tính đạo hàm ở vế thứ nhất của phương trình (2.136), theo kết quả của phương trình (2.42), ta có:

$$\frac{\partial}{\partial b_j} \left\{ \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] \right\} = \sum_{g=1}^G [-r_g + f_g P_g].$$

Về thứ hai của phương trình (2.136), có kết quả:

$$\frac{\partial}{\partial b_j} \left[-\frac{(b_j - \mu)^2}{2\sigma^2} \right] = -\frac{b_j - \mu}{\sigma^2}$$

Do đó, phương trình (2.136), được viết lại như sau:

$$\frac{\partial}{\partial b_j} \ell = \sum_{g=1}^G [-r_g + f_g P_g] - \frac{b_j - \mu}{\sigma^2}. \quad (2.137)$$

Tiếp theo, đạo hàm bậc 2 của hàm hợp lý, ta được:

$$\frac{\partial^2}{\partial b_j^2} \ell = \frac{\partial}{\partial b_j} \left\{ \sum_{g=1}^G [-r_g + f_g P_g] - \frac{b_j - \mu}{\sigma^2} \right\}.$$

Theo kết quả của phương trình (2.43), ta có:

$$\frac{\partial}{\partial b} \sum_{g=1}^G [-r_g + f_g P_g] = -\sum_{g=1}^G f_g P_g Q_g.$$

Và

$$\frac{\partial}{\partial b_j} \left\{ -\frac{b_j - \mu}{\sigma^2} \right\} = -\frac{1}{\sigma^2}.$$

Do đó, ta có phương trình đạo hàm bậc 2 cho ℓ :

$$\frac{\partial^2}{\partial b_j^2} \ell = -\sum_{g=1}^G f_g P_g Q_g - \frac{1}{\sigma^2} \quad (2.138)$$

Từ (2.137) và (2.138), theo phương pháp lặp Newton-Raphson, ta có phương trình ước lượng b_j như sau:

$$[b_j]_{t+1} = [b_j]_t + \frac{\sum_{g=1}^G [-r_g + f_g P_g] - \frac{b_j - \mu}{\sigma^2}}{\sum_{g=1}^G f_g P_g Q_g + \frac{1}{\sigma^2}}. \quad (2.139)$$

2.3.2.2. Mô hình IRT 2 tham số

Ta có hàm hợp lý cho mô hình 2 tham số như sau:

$$\ell = \text{constant} + \sum_{g=1}^G \left[r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g) \right] + \ln P(a_j) + \ln P(b_j),$$

ở đây, đặc trưng của câu hỏi gồm hai tham số a_j, b_j , nên để a_j, b_j làm hợp lý cực đại ℓ thỏa mãn phương trình sau:

$$\begin{aligned} \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix}_{t+1} &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix}_t - \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_j^2} & \frac{\partial^2 \ell}{\partial b_j \partial a_j} \\ \frac{\partial^2 \ell}{\partial a_j \partial b_j} & \frac{\partial^2 \ell}{\partial b_j^2} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} \frac{\partial \ell}{\partial a_j} \\ \frac{\partial \ell}{\partial b_j} \end{bmatrix}_t \\ &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix}_t - \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}_t \end{aligned}$$

Ma trận đạo hàm được tính như sau:

$$\begin{aligned} L_1 &= \frac{\partial \ell}{\partial a_j} = \frac{\partial}{\partial a_j} \left\{ \text{constant} + \sum_{g=1}^G \left[r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g) \right] + \ln P(a_j) \right\} \\ &= \frac{\partial}{\partial a_j} \left\{ \text{constant} + \sum_{g=1}^G \left[r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g) \right] \right\} \\ &\quad + \frac{\partial}{\partial a_j} \left\{ (v-1) \ln a_j - \frac{a_j^2}{2} \right\}. \end{aligned} \quad (2.140)$$

Theo kết quả của phương trình (2.52), ta có về thứ nhất của phương trình trên:

$$\frac{\partial}{\partial a_j} \left\{ \text{constant} + \sum_{g=1}^G \left[r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g) \right] \right\} = \sum_{g=1}^G f_g (\theta_g - b) (p_g - P_g),$$

tiếp theo, về thứ 2 của phương trình (2.140) được tính như sau:

$$\frac{\partial}{\partial a_j} \left\{ (v-1) \ln a_j - \frac{a_j^2}{2} \right\} = \frac{v-1}{a_j} - a_j.$$

Nên ta thu được phương trình (2.140) có dạng như sau:

$$L_1 = \sum_{g=1}^G f_g(\theta_g - b_j)(p_g - P_g) + \frac{v-1}{a_j} - a_j \quad (2.141)$$

Đạo hàm bậc 1 của hàm hợp lý theo b , được tính như sau:

$$L_2 = \frac{\partial \ell}{\partial b_j} = \frac{\partial}{\partial a_j} \left\{ \text{constant} + \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] + \ln P(b) \right\},$$

từ kết quả của phương trình (2.53) và (2.137), ta có kết quả sau:

$$L_2 = \frac{\partial \ell}{\partial b_j} = \sum_{g=1}^G a_j f_g(-p_g + P_g) - \frac{b_j - \mu}{\sigma^2}, \quad (2.142)$$

Ma trận Hessian được tính như sau. Đầu tiên:

$$L_{11} = \frac{\partial^2 \ell}{\partial a_j^2} = \frac{\partial}{\partial a_j} \left\{ \sum_{g=1}^G f_g(\theta_g - b_j)(p_g - P_g) + \frac{v-1}{a_j} - a_j \right\},$$

về thứ nhất của phương trình trên theo kết quả của phương trình (2.54), ta có kết quả:

$$\frac{\partial}{\partial a_j} \left\{ \sum_{g=1}^G f_g(\theta_g - b_j)(p_g - P_g) \right\} = - \sum_{g=1}^G f_g(\theta_g - b_j)^2 P_g Q_g,$$

về thứ hai của phương trình L_{11} được tính như sau:

$$\frac{\partial}{\partial a_j} \left\{ \frac{v-1}{a_j} - a_j \right\} = -\frac{v-1}{a_j^2} - 1,$$

nên ta thu được L_{11} có dạng như sau:

$$L_{11} = - \sum_{g=1}^G f_g (\theta_g - b_j)^2 P_g Q_g - \frac{\nu - 1}{a_j^2} - 1. \quad (2.143)$$

Đạo hàm bậc hai của hàm hợp lý theo b được tính như sau:

$$L_{22} = \frac{\partial^2}{\partial b_j^2} \ell = \frac{\partial}{\partial b_j} \left\{ \sum_{g=1}^G a_j f_g (-p_g + P_g) - \frac{b_j - \mu}{\sigma^2} \right\},$$

theo kết quả của phương trình (2.55) và (2.138), ta có kết quả như sau:

$$L_{22} = - \sum_{g=1}^G a_j^2 f_g P_g Q_g - \frac{1}{\sigma^2}. \quad (2.144)$$

Đạo hàm hỗn hợp của hàm hợp lý theo a và b được tính như sau:

$$L_{12} = L_{21} = \frac{\partial^2 \ell}{\partial a_j \partial b_j} = \frac{\partial}{\partial b} \left\{ \sum_{g=1}^G f_g (\theta_g - b_j) (p_g - P_g) + \frac{\nu - 1}{a_j} - a_j \right\},$$

ta có

$$\frac{\partial}{\partial b_j} \left[\frac{\nu - 1}{a_j} - a_j \right] = 0,$$

và theo kết quả của phương trình (2.56), ta có kết quả sau:

$$L_{12} = L_{21} = \sum_{g=1}^G [-r_g + f_g P_g + a_j (\theta_g - b_j) f_g P_g Q_g] \quad (2.145)$$

2.3.2.3. Mô hình IRT 3 tham số

Hàm hợp lý cho mô hình 3 tham số có dạng:

$$\begin{aligned} \ell = \text{constant} + \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] \\ + \ln P(a_j) + \ln P(b_j) + \ln P(c_j), \end{aligned}$$

đặc trưng của câu hỏi gồm 3 tham số đó là a, b và c , nên để ước lượng các tham số a, b, c ta cần giải phương trình sau:

$$\begin{aligned} \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_{t+1} &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_t - \begin{bmatrix} \frac{\partial^2 \ell}{\partial a_j^2} & \frac{\partial^2 \ell}{\partial b_j \partial a_j} & \frac{\partial^2 \ell}{\partial c_j \partial a_j} \\ \frac{\partial^2 \ell}{\partial a_j \partial b_j} & \frac{\partial^2 \ell}{\partial b_j^2} & \frac{\partial^2 \ell}{\partial c_j \partial b_j} \\ \frac{\partial^2 \ell}{\partial a_j \partial c_j} & \frac{\partial^2 \ell}{\partial b_j \partial c_j} & \frac{\partial^2 \ell}{\partial c_j^2} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} \frac{\partial \ell}{\partial a_j} \\ \frac{\partial \ell}{\partial b_j} \\ \frac{\partial \ell}{\partial c_j} \end{bmatrix}_t \\ &= \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \\ \hat{c}_j \end{bmatrix}_t - \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}_t^{-1} \cdot \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}_t. \end{aligned}$$

Các đạo hàm trong ma trận Gradient, được tính như sau. Đầu tiên ta tính đạo hàm của hàm hợp lý đối với tham số a :

$$\frac{\partial}{\partial a_j} \ell = \frac{\partial}{\partial a_j} \left\{ \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] + \ln P(a_j) \right\},$$

theo kết quả của phương trình (2.141) và (2.64), ta có:

$$L_1 = \sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) \left[\frac{P_g^*}{P_g} \right] + \frac{v-1}{a_j} - a_j. \quad (2.146)$$

Đạo hàm của hàm hợp lý theo tham số b_j :

$$\frac{\partial}{\partial b_j} \ell = \frac{\partial}{\partial b_j} \left\{ \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] + \ln P(b_j) \right\},$$

theo kết quả của phương trình (2.137) và (2.66), ta có:

$$L_2 = -a_j \sum_{g=1}^G f_g (p_g - P_g) \left[\frac{P_g^*}{P_g} \right] - \frac{b_j - \mu}{\sigma^2}. \quad (2.147)$$

Đạo hàm bậc nhất của hàm hợp lý theo c_j được tính như sau:

$$\frac{\partial}{\partial c_j} \ell = \frac{\partial}{\partial c_j} \left\{ \sum_{g=1}^G [r_g \ln P(\theta_g) + (f_g - r_g) \ln Q(\theta_g)] + \ln P(c_j) \right\},$$

ta có:

$$\frac{\partial}{\partial c_j} [(\alpha - 1) \ln c_j + (\beta - 1) \ln (1 - c_j)] = \frac{\alpha - 1}{c_j} + \frac{\beta - 1}{1 - c_j},$$

và theo kết quả của phương trình (2.67), ta có kết quả:

$$L_3 = \sum_{g=1}^G f_g (p_g - P_g) \frac{1}{P_g - c_j} \frac{P_g^*}{P_g} + \frac{\alpha - 1}{c_j} + \frac{\beta - 1}{1 - c_j} \quad (2.148)$$

Các đạo hàm trong ma trận Hessian, được tính như sau. Đầu tiên đạo hàm bậc hai của hàm hợp lý theo a_j có dạng:

$$\frac{\partial^2}{\partial a_j^2} \ell = \frac{\partial}{\partial a_j} \left\{ \sum_{g=1}^G f_g (p_g - P_g) (\theta_g - b_j) \left[\frac{P_g^*}{P_g} \right] + \frac{v-1}{a_j} - a_j \right\},$$

theo kết quả của phương trình (2.77) và (2.143), ta có kết quả sau:

$$L_{11} = - \sum_{g=1}^G f_g (\theta_g - b_j)^2 P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2 - \frac{v-1}{a_j^2} - 1. \quad (2.149)$$

Đạo hàm bậc hai của hàm hợp lý theo b có dạng:

$$\frac{\partial^2}{\partial b_j^2} \ell = \frac{\partial}{\partial b_j} \left\{ -a_j \sum_{g=1}^G f_g (p_g - P_g) \left[\frac{P_g^*}{P_g} \right] - \frac{b_j - \mu}{\sigma^2} \right\},$$

theo kết quả của phương trình (2.78) và (2.144), ta có kết quả:

$$L_{22} = -a_j^2 \sum_{g=1}^G f_g P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2 - \frac{1}{\sigma^2}. \quad (2.150)$$

Đạo hàm bậc hai của hàm hợp lý theo c_j có dạng như sau:

$$\frac{\partial^2}{\partial c_j^2} \ell = \frac{\partial}{\partial c_j} \left\{ \sum_{g=1}^G f_g(p_g - P_g) \frac{1}{P_g - c_j} \frac{P_g^*}{P_g} + \frac{\alpha - 1}{c_j} + \frac{\beta - 1}{1 - c_j} \right\},$$

ta có:

$$\frac{\partial}{\partial c_j} \left\{ \frac{\alpha - 1}{c_j} + \frac{\beta - 1}{1 - c_j} \right\} = -\frac{\alpha - 1}{c_j^2} + \frac{\beta - 1}{(1 - c_j)^2},$$

và theo kết quả của phương trình (2.79), ta có kết quả sau:

$$L_{33} = - \sum_{g=1}^G f_g \frac{Q_g}{1 - c_j} \frac{P_g^*}{P_g} - \frac{\alpha - 1}{c_j^2} + \frac{\beta - 1}{(1 - c_j)^2}. \quad (2.151)$$

Các đạo hàm hỗn hợp trong ma trận Hessian được tính như sau, đầu tiên đạo hàm hỗn hợp của ℓ theo a_j và b_j :

$$L_{12} = L_{21} = \frac{\partial^2}{\partial a_j \partial b_j} \ell = \frac{\partial}{\partial b_j} \left\{ \sum_{g=1}^G f_g(p_g - P_g)(\theta_g - b_j) \left[\frac{P_g^*}{P_g} \right] + \frac{\nu - 1}{a_j} - a_j \right\},$$

theo kết quả của phương trình (2.80), ta có đạo hàm:

$$L_{12} = L_{21} = a_j \sum_{g=1}^G f_g(\theta_g - b_j) P_g Q_g \left[\frac{P_g^*}{P_g} \right]^2. \quad (2.152)$$

Tiếp theo, đạo hàm hỗn hợp của ℓ theo a_j và c_j có dạng:

$$L_{13} = L_{31} = \frac{\partial^2}{\partial a_j \partial c_j} \ell = \frac{\partial}{\partial c_j} \left\{ \sum_{g=1}^G f_g(p_g - P_g)(\theta_g - b_j) \left[\frac{P_g^*}{P_g} \right] + \frac{\nu - 1}{a_j} - a_j \right\},$$

theo kết quả của phương trình (2.81), ta có kết quả sau:

$$L_{13} = L_{31} = - \sum_{g=1}^G f_g(\theta_g - b_j) \frac{Q_g}{1 - c_j} \frac{P_g^*}{P_g}. \quad (2.153)$$

Cuối cùng đạo hàm hỗn hợp của ℓ theo b_j và c_j :

$$L_{23} = L_{32} = \frac{\partial^2}{\partial b_j \partial c_j} \ell = \frac{\partial}{\partial c_j} \left\{ -a_j \sum_{g=1}^G f_g(p_g - P_g) \left[\frac{P_g^*}{P_g} \right] - \frac{b_j - \mu}{\sigma^2} \right\},$$

, theo kết quả của phương trình (2.82), ta có kết quả:

$$L_{23} = L_{32} = a_j \sum_{g=1}^G f_g \frac{Q_g}{1 - c_j} \frac{P_g^*}{P_g} \quad (2.154)$$

2.3.3. Cách chọn các tham số trong phân phối tiên nghiệm

Như đã trình bày ở phần trên, thì θ và b theo phân phối chuẩn, nên ta chọn:

$$\begin{aligned} \mu_\theta &= 0 \text{ và } \sigma_\theta^2 = 1 \\ \mu_b &= 0 \text{ và } \sigma_b^2 = 1. \end{aligned}$$

Theo Swaminathan và Gifford (1985) [31], tham số v trong phân phối tiên nghiệm của a_j , có thể chọn bằng cách xác định điểm đầu cuối của khoảng tin cậy $1 - \alpha$ cho a . Đối với v đủ lớn, Khoảng ước lượng $(a_1; a_2)$ của a với độ tin cậy $1 - \alpha$ được xác định bởi:

$$a_1 = \left(v - \frac{1}{2} \right)^{1/2} - \frac{z_{\alpha/2}}{2^{1/2}}, \quad a_2 = \left(v - \frac{1}{2} \right)^{1/2} + \frac{z_{\alpha/2}}{2^{1/2}},$$

trong đó $z_{\alpha/2}$ là phân vị mức $\frac{\alpha}{2}$ của phân phối chuẩn tắc. Do đó, ta có phương trình để ước lượng v :

$$v = \frac{1}{2} \left[1 + z_{\alpha/2}^2 \frac{(a_2 + a_1)^2}{a_2 - a_1} \right]. \quad (2.155)$$

Ta có bảng phân vị của phân phối chuẩn tắc như sau:

Bảng 2.3: Bảng phân vị của phân phối chuẩn tắc[22]

$\alpha = \text{tail area}$	central area = $1 - 2\alpha$	Z_α
0.10	0.80	$z_{0.10} = \mathbf{1.28}$
0.05	0.90	$z_{0.05} = \mathbf{1.645}$
0.025	0.95	$z_{0.025} = \mathbf{1.96}$
0.01	0.98	$z_{0.01} = \mathbf{2.33}$
0.005	0.99	$z_{0.005} = \mathbf{2.58}$

Nếu không có thông tin tiên nghiệm cụ thể, bạn có thể chọn giá trị mặc định $\nu = 10$. Đây là giá trị thường được sử dụng trong thực tế vì nó cân bằng giữa độ linh hoạt và độ tập trung của phân phối. Đối với $\nu \geq 10$, các phép xấp xỉ cho giá trị trung bình và độ lệch chuẩn của phân phối chỉ có sai số tương đối nhỏ (dưới 1.5%). Nếu chọn $\nu < 10$, hãy kiểm tra lại các phép xấp xỉ để đảm bảo độ chính xác. Nếu có dữ liệu lịch sử hoặc thông tin tiên nghiệm về tham số a , hãy sử dụng dữ liệu này để điều chỉnh ν sao cho phân phối tiên nghiệm phù hợp với dữ liệu [31].

Theo Swaminathan và Gifford (1986) [32], tham số α và β trong phân phối Beta của tham số c được chọn như sau. Giá trị trung bình M của phân phối tiên nghiệm có thể được chọn dựa trên thông tin tiên nghiệm (ví dụ: xác suất một thí sinh có năng lực thấp trả lời đúng một mục là 0.15). Số lượng quan sát m mà thông tin tiên nghiệm đáng giá được sử dụng để xác định α và β :

$$\alpha = mM \quad \beta = m(1 - M) - 2. \quad (2.156)$$

Ví dụ, với $M = 0.15$ và $m = 15$:

$$\alpha = 2.25 \approx 2.0$$

$$\beta = 10.75 \approx 11.0.$$

Ví dụ này cho khoảng tin cậy của c :

$$90\% : (0.04, 0.36),$$

$$95\% : (0.03, 0.40),$$

99% : (0.01, 0.49).

Nếu khoảng tin cậy quá rộng, giá trị m có thể được tăng lên để làm cho phân phối tiên nghiệm chính xác hơn. Một phân phối tiên nghiệm không quá mơ hồ và cũng không quá chính xác thường được ưu tiên. Do đó, một ước lượng bảo thủ của m như trong ví dụ minh họa là hợp lý trong hầu hết các tình huống.

2.4. Ước lượng kỳ vọng hậu nghiệm

Quy trình ước lượng Bayes kỳ vọng hậu nghiệm (Expected A Posteriori - EAP) dựa trên dạng sau của định lý Bayes:

$$g(\theta_i | \mathbf{u}, \xi) = \frac{P(\mathbf{u} | \theta_i, \xi) g(\theta)}{P(\mathbf{u})}. \quad (2.157)$$

Dưới giả định độc lập cục bộ, xác suất để thí sinh i phản hồi với vector hồi đáp $\mathbf{u} = [u_{i1}, u_{i2}, \dots, u_{in}]$ là

$$P(\mathbf{u} | \theta_i, \xi) = \prod_{j=1}^N P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}}$$

đây là hàm hợp lý (likelihood) có điều kiện dựa trên giá trị của θ_i . Đối với một thí sinh được chọn ngẫu nhiên từ một quần thể có phân phối khả năng $g(\theta)$, mẫu số của phương trình trên được cho bởi

$$P(U_j) = \int P(U_j | \theta) g(\theta) d\theta$$

đây là xác suất không điều kiện. Từ phương trình (2.157), kỳ vọng không điều kiện của θ_i , với \mathbf{u} , là

$$E(\theta_i | \mathbf{u}, \xi) = \frac{\int \theta_i g(\theta) \prod_{j=1}^N P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}} d\theta}{\int g(\theta) \prod_{j=1}^N P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}} d\theta} \quad (2.158)$$

Như trình bày ở phương pháp MMLE, ta có phương trình:

$$E(\theta_i | \mathbf{u}, \xi) = \bar{\theta}_i = \frac{\sum_{k=1}^q X_k L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}, \quad (2.159)$$

trong đó, $L(X_k) = \prod_{j=1}^N P_j(X_k)^{u_{ij}} Q_j(X_k)^{1-u_{ij}}$. Phương trình này nó không lặp và cho ước lượng EAP của năng lực thí sinh trực tiếp.

Đối với ước lượng EAP, việc tính toán độ lệch chuẩn từ giá trị trung bình (là điểm EAP ước lượng) một cách trực tiếp là điều tự nhiên, bằng cách lấy các giá trị của hàm hợp lý tại các điểm cầu phương [8], ta có:

$$SE(\hat{\theta}) = \sqrt{\frac{\sum_{k=1}^q (X_k - \hat{\theta})^2 L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}} \quad (2.160)$$

CHƯƠNG 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Ở chương này, với phương pháp MLE và MAP chúng tôi sẽ trình bày code và thực nghiệm ước lượng 1 câu hỏi với nhiều thí sinh và 1 thí sinh với nhiều câu hỏi trên tập dữ liệu tự tạo. Còn phương pháp MMLE và EAP chúng tôi sẽ thực nghiệm trên dữ liệu thực tế với các gói thư viện trong R.

3.1. Phương pháp MLE và MAP

3.1.1. Dữ liệu thực nghiệm

Dữ liệu được mô phỏng để ước lượng các tham số của một câu hỏi được mô phỏng như sau. Năng lực của thí sinh θ được giả định phân bố trong khoảng từ $(-3; 3)$, tạo thành 33 mức năng lực khác nhau. Mỗi mức năng lực có 21 thí sinh. Giả định câu hỏi có tham số thực: độ phân biệt $a = 1.27$, độ khó $b = -0.39$ và độ đoán mò $c = 0.17$. Xác suất trả lời đúng $P(\theta)$ được tính theo công thức của mô hình IRT ((1.2), (1.4) và (1.5)). Sau đó, số lượng thí sinh trả lời đúng tại mỗi mức năng lực được mô phỏng bằng phân phối nhị thức $r \sim \text{Binomial}(n = 21, p = P(\theta))$. Kết quả thu được là tỉ lệ trả lời đúng thực nghiệm $p = r/f$ tại từng mức năng lực, được sử dụng làm dữ liệu đầu vào cho quá trình ước lượng tham số. Dữ liệu được mô phỏng để ước lượng năng lực của một thí sinh với nhiều câu hỏi d mô tả như sau. Đầu tiên, tham số độ khó b được sinh ra từ phân phối chuẩn với giá trị trung bình bằng 0 và độ lệch chuẩn bằng 1, sau đó giới hạn trong khoảng từ -3 đến 3 để đảm bảo tính hợp lý. Tiếp theo, tham số độ phân biệt a được tạo từ phân phối chi bình phương với bậc tự do bằng 3, sau đó chuyển đổi về khoảng từ 0.1 đến 2.8 để phù hợp với các giá trị thực tế. Đối với tham số đoán đúng c , dữ liệu được sinh từ phân phối beta với tham số (5, 17) rồi nhân với 0.35 để giới hạn giá trị trong khoảng từ 0 đến 0.35, phản ánh xác suất đoán đúng thấp.

Sau khi có các tham số, tùy thuộc vào mô hình IRT ((1.2), (1.4) và (1.5)) mà một số tham số có thể được cố định và cách chọn lựa tham số được mô tả chi tiết ở mục 2.3.3. Chẳng hạn, trong mô hình 1PL, tất cả các câu hỏi có cùng độ phân biệt a bằng 1 và không có đoán đúng ($c = 0$), trong khi mô hình 2PL cho phép a thay đổi nhưng vẫn giữ c bằng 0.

Cuối cùng, xác suất trả lời đúng của thí sinh được tính bằng công thức IRT ((1.2), (1.4) và (1.5)), kết hợp năng lực θ của thí sinh với các tham số câu hỏi.

Dựa trên xác suất này, phản hồi của thí sinh (0 hoặc 1) được mô phỏng bằng phân phối nhị thức. Kết quả trả về bao gồm các phản hồi cùng với các tham số đã sử dụng, giúp phân tích và đánh giá mô hình một cách chi tiết.

Dưới đây là dữ liệu sinh ra cho từng mô hình để ước lượng các đặc trưng của câu hỏi và năng lực của thí sinh. Ở đây với phương pháp MLE và MAP khi ước lượng các đặc trưng của câu hỏi thì năng lực của thí sinh được nhóm nên dữ liệu gồm đầu vào gồm năng lực của từng nhóm thí sinh, r_g , f_g và $p(\theta_g)$ và đầu ra là các đặc trưng của câu hỏi. Đối với năng lực của thí sinh thì đầu vào là các hồi đáp của thí sinh với các câu hỏi và các đặc trưng của câu hỏi và đầu ra năng lực của thí sinh.

Ở đây ta chọn các giá trị đúng cho các tham số cho đặc trưng của câu hỏi và năng lực thí sinh với:

- Năng lực của thí sinh: $\theta = 0.5$
- Độ khó của câu hỏi: $b = 0.4$
- Độ phân biệt của câu hỏi: $a = 1.4$
- Độ đoán mò của câu hỏi: $c = 0.17$

3.1.1.1. Dữ liệu ước lượng đặc trưng của câu hỏi cho mô hình IRT 1 tham số

Với mô hình IRT 1 tham số (1.2), chọn các tham số đúng như ở trên với $b = 0.4$. Chia năng lực của thí sinh thành 33 nhóm năng lực và mỗi nhóm gồm 21 thí sinh. Ta có bảng dữ liệu đầu vào như bảng dữ liệu ở dưới:

Bảng 3.1: Dữ liệu ước lượng đặc trưng câu hỏi cho mô hình IRT 1PL

Nhóm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
θ	-3.0000	-2.8125	-2.6250	-2.4375	-2.2500	-2.0625	-1.8750	-1.6875	-1.5000	-1.3125	-1.1250	-0.9375	-0.7500	-0.5625	-0.3750	-0.1875
r_{1PL}	0	0	0	1	1	2	1	2	2	5	4	6	8	9	7	8
f_{1PL}	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
p_{1PL}	0.0476	0.2381	0.1905	0.2857	0.1905	0.0952	0.0476	0.0952	0.0952	0.2381	0.1905	0.2857	0.3809	0.4286	0.3333	0.3809

Nhóm	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
θ	0.0000	0.1875	0.3750	0.5625	0.7500	0.9375	1.1250	1.3125	1.5000	1.6875	1.8750	2.0625	2.2500	2.4375	2.6250	2.8125	3.0000
r_{1PL}	9	11	16	14	14	13	11	17	18	17	19	19	18	17	17	19	17
f_{1PL}	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
p_{1PL}	0.4286	0.5238	0.7619	0.6667	0.6667	0.6190	0.5238	0.6667	0.8571	0.8095	0.9048	0.9048	0.8571	0.8095	0.7619	0.9048	0.8095

3.1.1.2. Dữ liệu ước lượng đặc trưng của câu hỏi cho mô hình IRT 2 tham số

Với mô hình IRT 2 tham số (1.4), chọn các tham số đúng như ở trên với $b = 0.4, a = 1.4$. Chia năng lực của thí sinh thành 33 nhóm năng lực và mỗi nhóm gồm 21 thí sinh. Ta có bảng dữ liệu đầu vào như bảng dữ liệu ở dưới:

Bảng 3.2: Dữ liệu ước lượng đặc trưng câu hỏi cho mô hình IRT 2PL

Nhóm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
θ	-3.0000	-2.8125	-2.6250	-2.4375	-2.2500	-2.0625	-1.8750	-1.6875	-1.5000	-1.3125	-1.1250	-0.9375	-0.7500	-0.5625	-0.3750	-0.1875
r_{2PL}	0	0	0	1	0	0	2	0	1	3	3	1	2	3	9	8
f_{2PL}	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
p_{2PL}	0.0000	0.0000	0.0000	0.0476	0.0000	0.0000	0.0952	0.0000	0.0476	0.1429	0.1429	0.0476	0.0952	0.1429	0.4286	0.3810

Nhóm	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
θ	0.0000	0.1875	0.3750	0.5625	0.7500	0.9375	1.1250	1.3125	1.5000	1.6875	1.8750	2.0625	2.2500	2.4375	2.6250	2.8125	3.0000
r_{2PL}	13	9	4	10	14	15	18	15	20	17	18	20	21	20	19	18	21
f_{2PL}	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
p_{2PL}	0.6190	0.4286	0.1905	0.4762	0.6667	0.7143	0.8571	0.7143	0.9524	0.8095	0.8571	0.9524	1.0000	0.9524	0.9048	0.8571	1.0000

3.1.1.3. Dữ liệu ước lượng đặc trưng của câu hỏi cho mô hình IRT 3 tham số

Với mô hình IRT 3 tham số (1.5), chọn các tham số đúng như ở trên với $b = 0.4, a = 1.4, c = 0.17$. Chia năng lực của thí sinh thành 33 nhóm năng lực và mỗi nhóm gồm 21 thí sinh. Ta có bảng dữ liệu đầu vào như bảng dữ liệu ở dưới:

Bảng 3.3: Dữ liệu ước lượng đặc trưng câu hỏi cho mô hình IRT 3PL

Nhóm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
θ	-3.0000	-2.8125	-2.6250	-2.4375	-2.2500	-2.0625	-1.8750	-1.6875	-1.5000	-1.3125	-1.1250	-0.9375	-0.7500	-0.5625	-0.3750	-0.1875
r_{3PL}	2	3	6	6	5	6	2	5	8	5	4	7	10	4	10	11
f_{3PL}	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
p_{3PL}	0.0952	0.1429	0.2857	0.2857	0.2381	0.2857	0.0952	0.2381	0.3810	0.2381	0.1905	0.3333	0.4762	0.1905	0.4762	0.5238

Nhóm	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
θ	0.0000	0.1875	0.3750	0.5625	0.7500	0.9375	1.1250	1.3125	1.5000	1.6875	1.8750	2.0625	2.2500	2.4375	2.6250	2.8125	3.0000
r_{3PL}	9	9	13	12	14	14	19	18	18	20	20	21	21	20	17	21	21
f_{3PL}	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
p_{3PL}	0.4286	0.4286	0.6190	0.5714	0.6667	0.6667	0.9048	0.8571	0.8571	0.9524	0.9524	1.0000	1.0000	0.9524	0.8095	1.0000	1.0000

3.1.1.4. Dữ liệu ước lượng năng lực thí sinh

Với từng loại mô hình IRT ta có những vector đặc trưng câu hỏi khác nhau và dữ liệu phản hồi của thí sinh đó với từng câu hỏi.

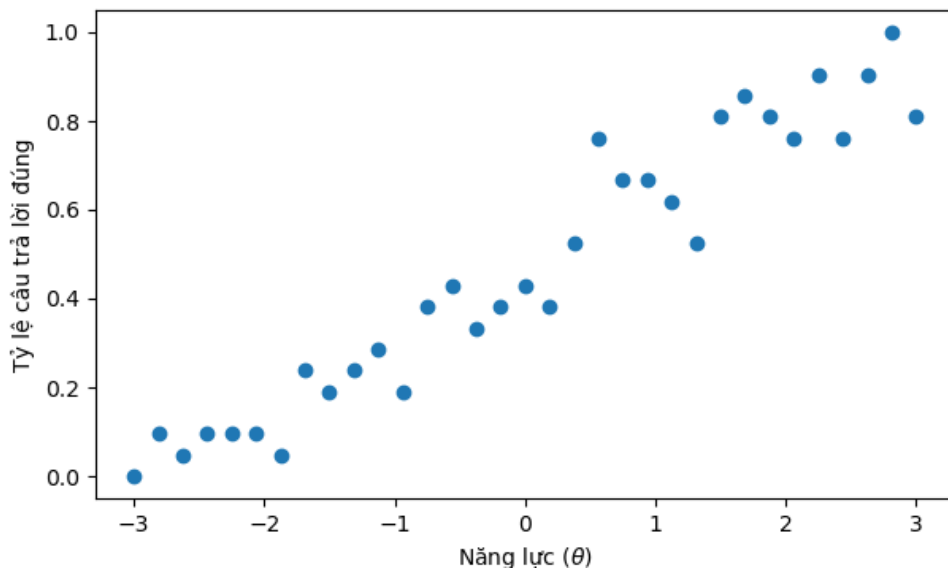
```
array([0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0,
      1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0,
      1, 1, 1, 1, 1, 0])
```

Hình 3.1: Dữ liệu phản hồi của 1 thí sinh với nhiều câu hỏi

3.1.2. Ước lượng các tham số của câu hỏi

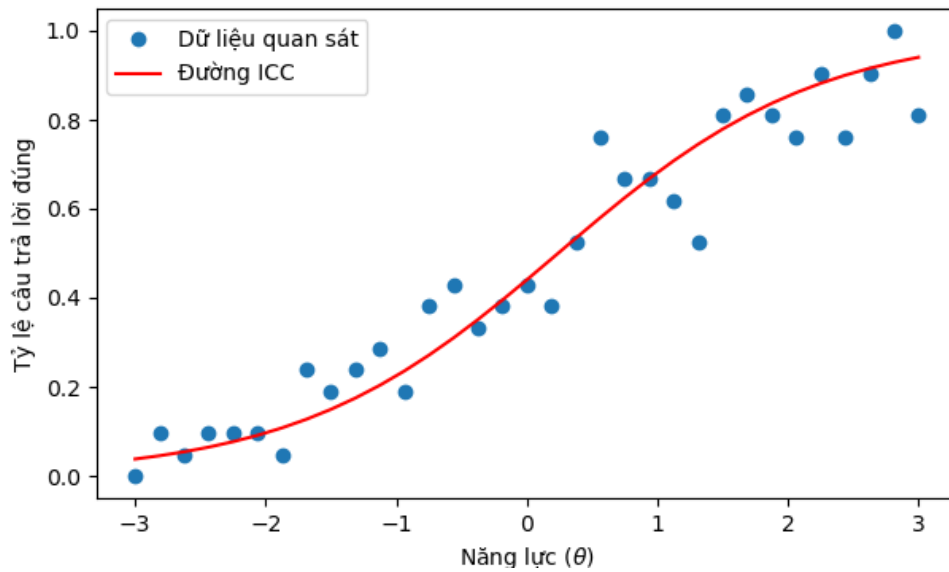
3.1.2.1. Phương pháp MLE

Với mô hình 1 tham số (1.2), sau khi sinh dữ liệu ta có biểu đồ tỷ lệ quan sát đúng tại từng mức năng lực như sau:



Hình 3.2: Tỷ lệ quan sát trả lời đúng tại mỗi mức năng lực theo mô hình 1pl

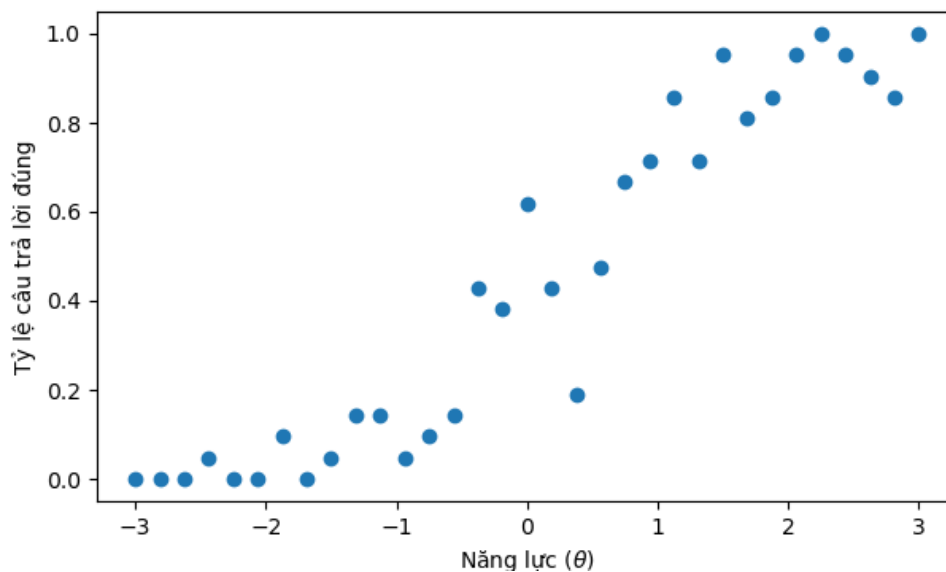
Sau khi, ước lượng ta thu được kết quả là độ khó $b = 0.23949523$, chỉ số chi bình phương $\chi^2 = 35.40720796$ và ngưỡng tiêu chuẩn bằng 46.19425952027847. Và ta có đường cong ICC trùng khớp tốt nhất với dữ liệu quan sát được ở hình 3.2 như sau:



Hình 3.3: Đường cong DTCH 1 tham số với tỷ lệ quan sát được dùng MLE

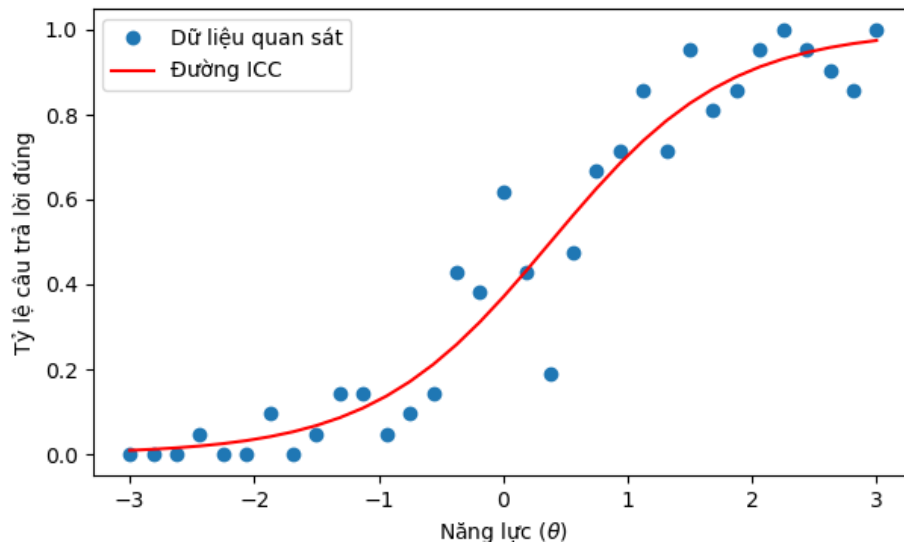
Ở đây giá trị độ khó với giá trị đúng ta chọn nằm trong khoảng trung bình $[-0.5; 0.5]$ với $b = 0.4$, sau khi ta ước lượng thì được $b = 0.23949523$ vẫn nằm trong khoảng trung bình, vì vậy không thay đổi bản chất mức độ khó của câu hỏi. Câu hỏi được thiết kế hợp lý về mặt độ khó. Sai số giữa thiết kế và thực tế sau ước lượng là nhỏ và chấp nhận được trong thực hành IRT. Giá trị $b = 0.239$ vẫn đảm bảo mục tiêu đánh giá thí sinh có năng lực trung bình.

Với mô hình 2 tham số (1.4), sau khi sinh dữ liệu ta có biểu đồ tỷ lệ quan sát đúng tại từng mức năng lực như sau:



Hình 3.4: Tỷ lệ quan sát trả lời đúng tại mỗi mức năng lực theo mô hình 2pl

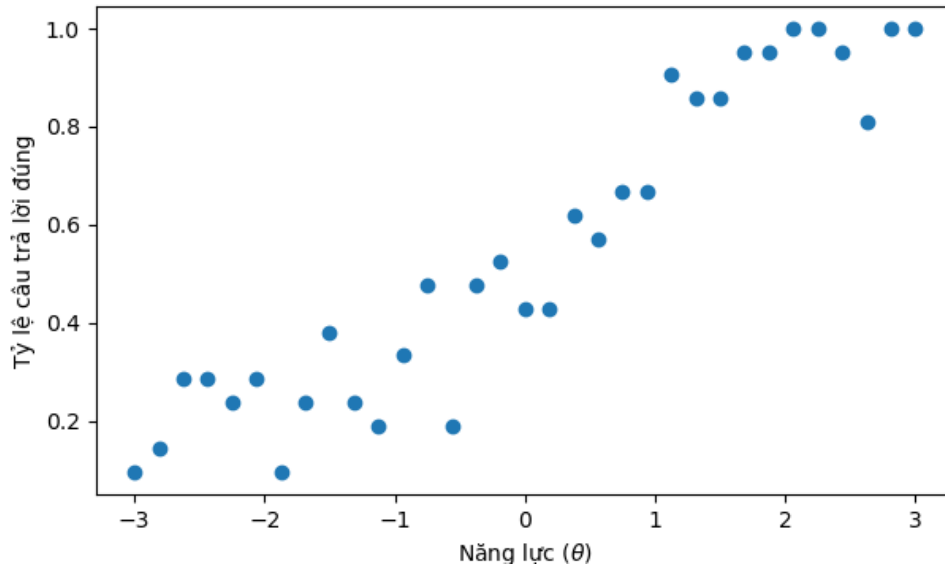
Sau khi, ước lượng ta thu được kết quả là độ phân biệt $a = 1.3968010225970338$, độ khó $b = 0.38082703143684193$, chỉ số chi bình phương $\chi^2 = 45.14324047$ và ngưỡng tiêu chuẩn bằng 44.98534328036513. Và ta có đường cong ICC trùng khớp tốt nhất với dữ liệu quan sát được ở hình 3.4 như sau:



Hình 3.5: Đường cong DTCH 2 tham số với tỷ lệ quan sát được dùng MLE

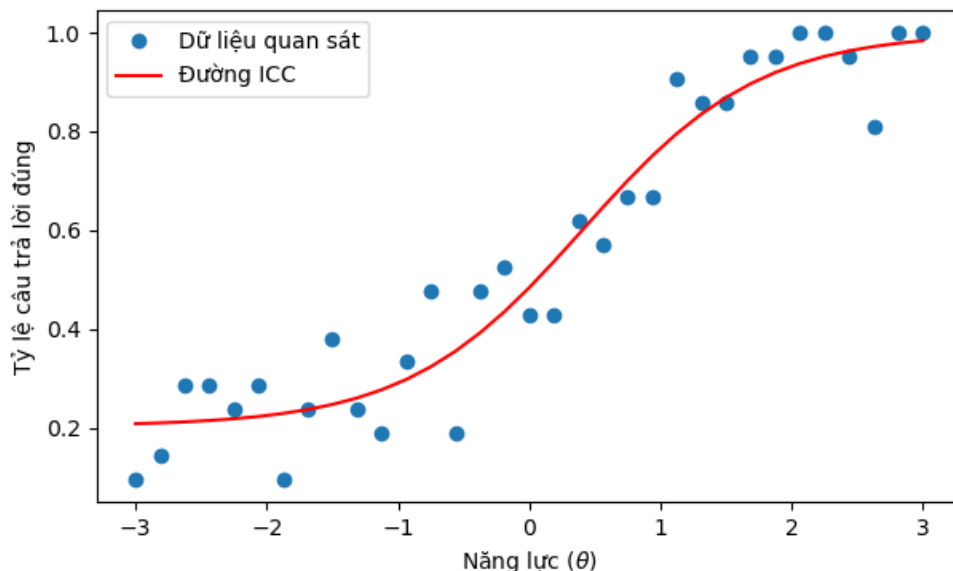
Ở đây giá trị độ khó với giá trị đúng ta chọn nằm trong khoảng trung bình $[-0.5; 0.5)$ với $b = 0.4$ và độ phân biệt nằm trong khoảng tốt $[1.35; 1.7)$ sau khi ta ước lượng thì được $b = 0.38082703143684193$ và $a = 1.3968010225970338$. Câu hỏi có độ khó trung bình và khả năng phân biệt tốt, đúng như mục tiêu thiết kế ban đầu. Sai số giữa giá trị giả định và giá trị ước lượng là rất nhỏ, chứng tỏ chất lượng câu hỏi được đảm bảo, phù hợp để đánh giá hiệu quả năng lực thí sinh trong khoảng trung bình đến khá.

Với mô hình 3 tham số (1.5), sau khi sinh dữ liệu ta có biểu đồ tỷ lệ quan sát đúng tại từng mức năng lực như sau:



Hình 3.6: Tỷ lệ quan sát trả lời đúng tại mỗi mức năng lực theo mô hình 3pl

Sau khi, ước lượng ta thu được kết quả là độ phân biệt $a = 1.4887052775483185$, độ khó $b = 0.41087808000023074$, độ đoán mò $c = 0.2031761422158637$, chỉ số chi bình phương $\chi^2 = 43.84246510872816$ và ngưỡng tiêu chuẩn bằng 43.77297182574219. Và ta có đường cong ICC trùng khớp tốt nhất với dữ liệu quan sát được ở hình 3.6 như sau:



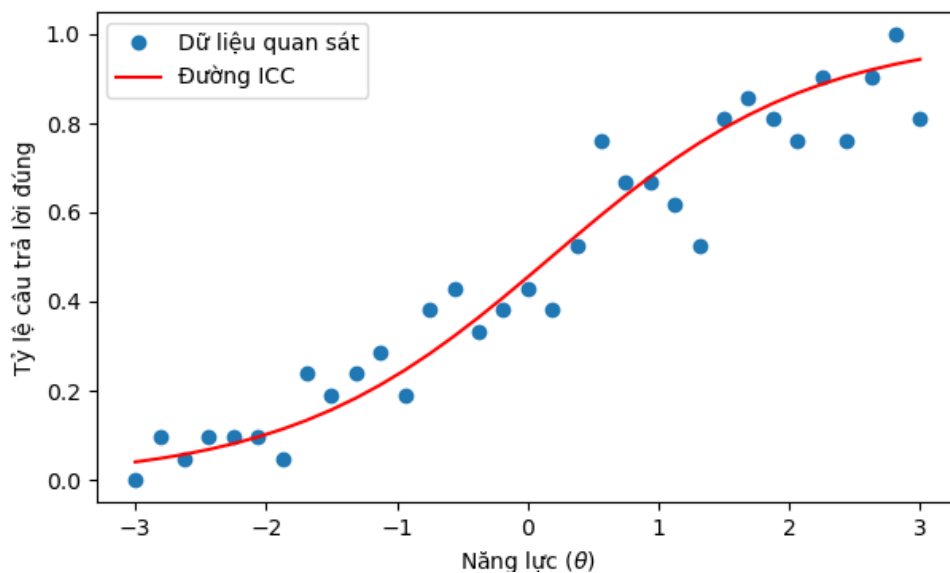
Hình 3.7: Đường cong ĐTCH 3 tham số với tỷ lệ quan sát được dùng MLE

Ở đây giá trị độ khó với giá trị đúng ta chọn nằm trong khoảng trung bình $[-0.5; 0.5]$ với $b = 0.4$ và độ phân biệt nằm trong khoảng tốt $[1.35; 1.7)$ độ đoán

mò trong khoảng chấp nhận được $c = 0.17$ sau khi ta ước lượng thì được $b = 0.41087808000023074$ và $a = 1.4887052775483185$ và độ đoán mò trong khoảng chấp nhận được $c = 0.2031761422158637$. Câu hỏi sau khi hiệu chỉnh bằng mô hình 3 tham số có các tham số ước lượng ổn định và hợp lý, duy trì được mục tiêu thiết kế ban đầu về độ khó, độ phân biệt và độ đoán mò. Đây là một câu hỏi chất lượng cao, phù hợp để sử dụng trong đánh giá năng lực theo mô hình IRT.

3.1.2.2. Phương pháp MAP

Với mô hình 1 tham số ta ước lượng được độ khó $b = 0.18108110929899324$, chỉ số chi bình phương $\chi^2 = 36.35636540374039$ và ngưỡng tiêu chuẩn bằng 46.1942595202784. Và ta có đường cong ICC trùng khớp tốt nhất với dữ liệu quan sát được ở hình 3.2 như sau:

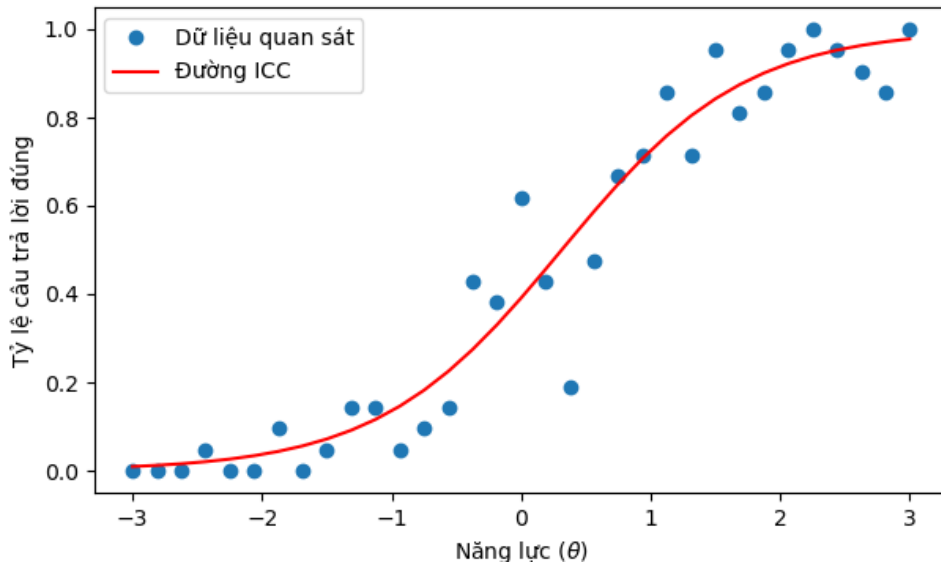


Hình 3.8: Đường cong DTCH 1 tham số với tỷ lệ quan sát được dùng MAP

Ở đây giá trị độ khó với giá trị đúng ta chọn nằm trong khoảng trung bình $[-0.5; 0.5)$ với $b = 0.4$, sau khi ta dùng MAP ước lượng thì được $b = 0.18108110929899324$ so với MLE ước lượng được $b = 0.23949523$ vẫn nằm trong khoảng trung bình, vì vậy không thay đổi bản chất mức độ khó của câu hỏi. Câu hỏi được thiết kế hợp lý về mặt độ khó. Sai số giữa thiết kế và thực tế sau ước lượng là nhỏ và chấp nhận được trong thực hành IRT. Trong trường hợp này, cả hai phương pháp đều dưới ước lượng tham số độ khó b , nhưng MAP cho kết quả nhỏ hơn MLE do bị ảnh hưởng bởi prior. MLE có thể chính xác hơn nếu dữ liệu đủ mạnh, trong khi MAP ổn định hơn trong trường hợp dữ liệu ít nhưng có

thể bị lệch nếu prior không phản ánh đúng thực tế.

Với mô hình 2 tham số ta ước lượng được độ phân biệt $a = 1.4132168050684217$, độ khó $b = 0.3132575026857251$, chỉ số chi bình phương $\chi^2 = 47.16270966077577$ và ngưỡng tiêu chuẩn bằng 44.98534328036513. Và ta có đường cong ICC trùng khớp tốt nhất với dữ liệu quan sát được ở hình 3.4 như sau:



Hình 3.9: Đường cong DTCH 2 tham số với tỷ lệ quan sát được dùng MAP

Ở đây giá trị độ khó với giá trị đúng ta chọn nằm trong khoảng trung bình $[-0.5; 0.5)$ với $b = 0.4$ và độ phân biệt nằm trong khoảng tốt $[1.35; 1.7)$ sau khi ta dùng MAP ước lượng thì được $b = 0.3132575026857251$ và $a = 1.4132168050684217$ so với dùng MLE ước lượng thì được $b = 0.38082703143684193$ và $a = 1.3968010225970338$. Với tham số b , MLE cho kết quả chính xác hơn so với MAP, do không bị ảnh hưởng bởi prior kéo về 0. Với tham số a , cả MLE và MAP đều ước lượng tốt, và sai số nhỏ, trong đó MAP hơi cao hơn MLE một chút nhưng vẫn trong phạm vi hợp lý. Trong các mô hình thực tế, MAP sẽ hữu ích hơn nếu bạn có prior đáng tin cậy, còn MLE thường chính xác hơn khi dữ liệu đủ lớn và muốn tránh ảnh hưởng của prior. Câu hỏi có độ khó trung bình và khả năng phân biệt tốt, đúng như mục tiêu thiết kế ban đầu. Sai số giữa giá trị giả định và giá trị ước lượng là rất nhỏ, chứng tỏ chất lượng câu hỏi được đảm bảo, phù hợp để đánh giá hiệu quả năng lực thí sinh trong khoảng trung bình đến khá.

3.1.3. Ước lượng năng lực của thí sinh

3.1.3.1. Thông tin cập nhật

- Giá trị đúng của năng lực: $\theta = 0.5$, nằm trong ngưỡng đầu của mức giỏi: $[0.5, 2.0)$.
- Mục tiêu: kiểm tra các phương pháp ước lượng có nhận diện đúng thí sinh thuộc nhóm giỏi hay không.

3.1.3.2. Phân tích lại theo mức năng lực

Bảng 3.4: So sánh kết quả ước lượng năng lực theo các mô hình với phương pháp MLE và MAP

Mô hình	Phương pháp	$\hat{\theta}$	SE	Đánh giá năng lực
1PL	MLE	1.3664	0.3473	Giỏi
2PL	MLE	0.6363	0.2903	Giỏi
3PL	MLE	1.0797	0.3726	Giỏi
1PL	MAP	0.2277	0.1279	Khá
2PL	MAP	0.1685	0.1255	Khá
3PL	MAP	0.1777	0.1329	Khá

3.1.3.3. Nhận xét theo mức năng lực

- Phương pháp MLE: Tất cả các mô hình đều xác định đúng mức giỏi cho thí sinh.
 - + 1PL và 3PL ước lượng cao hơn nhiều, có thể gây hiểu nhầm về năng lực thực sự.
 - + 2PL cho kết quả chính xác và hợp lý nhất.
- Phương pháp MAP: Tất cả mô hình đều ước lượng thấp hơn 0.5, xếp sai mức năng lực thành trung bình:
 - + Nguyên nhân do prior kéo giá trị về trung tâm (thường là 0).
 - + Hậu quả: nếu dùng MAP không hiệu chỉnh prior, dễ đánh giá thấp thí, đặc biệt gần ranh giới phân loại.

3.1.3.4. Kết luận

- MLE phù hợp hơn trong việc phân loại đúng mức năng lực
- MAP ổn định nhưng dễ bị underestimate, đặc biệt nguy hiểm ở các ngưỡng

phân loại.

- Nếu vẫn sử dụng MAP, nên cân nhắc hiệu chỉnh prior để phản ánh đúng phân bố năng lực thực tế.

3.2. Phương pháp MMLE và EAP

3.2.1. Mô tả dữ liệu

Trong nghiên cứu này, tôi sử dụng tập dữ liệu Dicht_Data2.csv, được tải từ nền tảng Kaggle [23]. Tập dữ liệu này phản ánh kết quả của một bài kiểm tra trắc nghiệm thực tế, trong đó mỗi dòng biểu diễn dữ liệu của một học sinh, được định danh thông qua mã số học sinh (Student ID). Các cột tiếp theo tương ứng với các câu hỏi trong bài kiểm tra, mỗi câu hỏi bao gồm bốn phương án lựa chọn. Và có gồm 60 câu hỏi và 239 học sinh tham gia trả lời. Phản hồi của học sinh đối với từng câu hỏi được mã hóa nhị phân: giá trị 1 nếu học sinh chọn đúng đáp án và 0 nếu chọn sai. Cần lưu ý rằng tập dữ liệu này không bao gồm nội dung cụ thể của các câu hỏi và phương án, mà những thông tin đó được lưu trữ trong một tập tin riêng biệt.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1		Student ID 1. If the rad 2. A cable t 3. Bolling w 4. A differe 5. Air bubb 6. Sound w 7. The heat 8. Equal ch 9. Wavelen 10. When v 11. Dopple 12. In myor 13. Two bul 14. Materi 15. In case 16. In horiz 17. The cur 18. The nor 19. The pre 20. Inorgan 2																				
2	0	3402	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
3	1	2179	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	2	1061	1	0	1	0	0	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0
5	3	2781	1	1	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0	1	1	1
6	4	3688	1	0	1	0	1	1	0	0	0	0	0	0	1	1	0	1	1	0	1	0
7	5	530	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0	0	1	0
8	6	2400	1	0	0	0	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0
9	7	3035	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0
10	8	1320	1	1	0	0	1	1	0	0	0	0	0	1	1	1	0	1	0	0	0	0
11	9	3372	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0
12	10	3642	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0
13	11	2729	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0
14	12	2877	1	1	0	0	0	1	1	1	1	0	1	1	0	0	0	0	1	0	0	1
15	13	4260	1	0	0	1	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0
16	14	1500	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	0	1	1	1	0
17	15	1501	0	0	0	1	0	0	0	1	1	0	1	1	1	1	0	0	0	0	1	0
18	16	1502	0	1	0	0	0	1	0	0	1	0	1	0	1	1	1	0	0	0	1	0
19	17	1503	1	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	0	0	0	1
20	18	1504	1	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0
21	19	1508	1	1	0	0	0	1	1	1	1	0	0	1	1	1	1	0	1	0	0	0
22	20	1512	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1
23	21	1525	1	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0

Hình 3.10: Dữ liệu phản hồi của thí sinh khi làm bài kiểm tra trong tập dữ liệu Dicht_Data2.csv

Trong nghiên cứu này, tôi sử dụng gói TAM trong ngôn ngữ R để ước lượng các tham số của mô hình IRT thông qua hai phương pháp MMLE và EAP. Trước khi tiến hành phân tích, tôi thực hiện bước tiền xử lý dữ liệu bằng cách loại bỏ hai cột không cần thiết là STT và Student ID khỏi tập dữ liệu. Sau đó, tôi chuẩn hóa tên các biến quan sát bằng cách thay thế tên cột hiện có bằng định dạng mới lần lượt là $I_1, I_2, I_3, \dots, I_{60}$, tương ứng với 60 câu hỏi trắc nghiệm trong bài kiểm tra. Và trong tập dữ liệu được điền đầy đủ giá trị và không có giá trị NULL và N/A.

3.2.2. Ước lượng các tham số trong mô hình

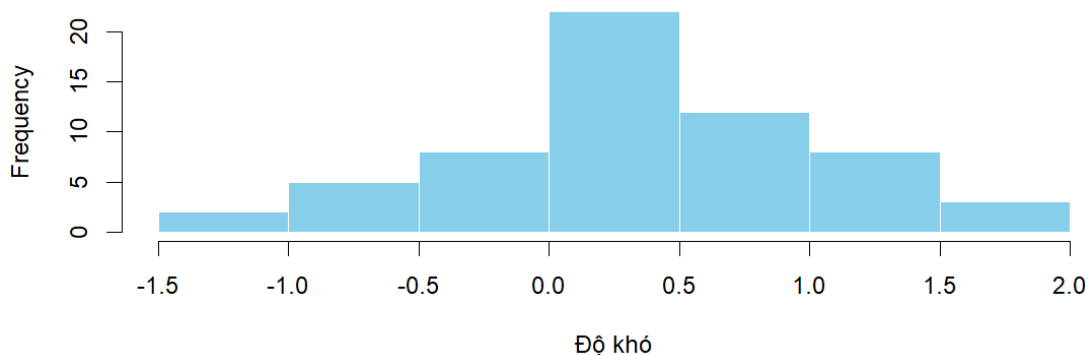
3.2.2.1. Mô hình IRT 1 tham số

Đầu tiên ta sử dụng mmle để ước lượng tham số độ khó của câu hỏi ta được. Và sau khi lặp 17 lần, ta thu được kết quả

Bảng 3.5: Tham số độ khó trong mô hình 1pl sau khi ước lượng bằng MMLE

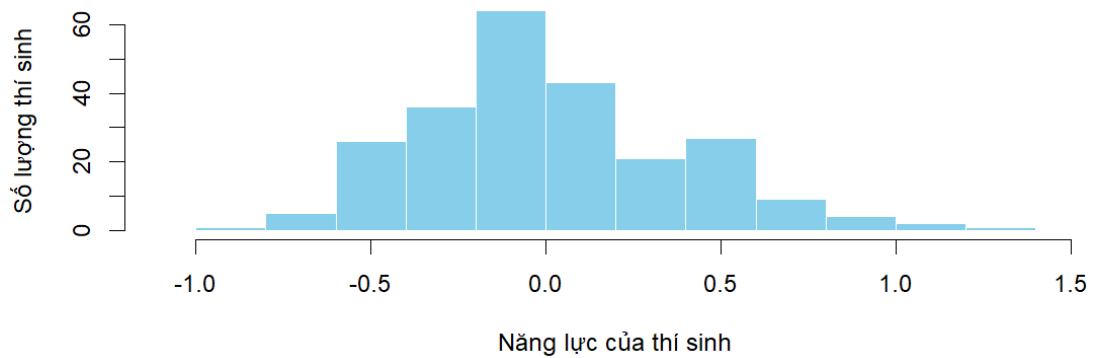
Item	Độ khó	Item	Độ khó	Item	Độ khó	Item	Độ khó
I1	-0.71626820	I16	1.26651804	I31	0.13495586	I46	1.01257278
I2	0.18796703	I17	0.86423463	I32	-0.37849368	I47	-0.61961970
I3	0.18796703	I18	0.33065477	I33	0.70383237	I48	0.90576150
I4	0.29475640	I19	0.62651388	I34	1.64671412	I49	1.47368973
I5	-0.18185394	I20	0.88491868	I35	0.25902282	I50	0.64568724
I6	-0.43309251	I21	1.50133022	I36	-0.25283851	I51	1.10155167
I7	0.72343615	I22	1.29120727	I37	0.72343615	I52	1.21803757
I8	0.70383237	I23	0.72343615	I38	0.92676901	I53	0.40303243
I9	-0.07606258	I24	0.38485846	I39	0.01180103	I54	1.77321052
I10	0.33065477	I25	0.53203735	I40	-0.54405568	I55	-0.09365565
I11	0.45791297	I26	0.09971191	I41	0.25902282	I56	-1.33203714
I12	-0.56281554	I27	0.22343316	I42	0.24121129	I57	-0.94003079
I13	-0.09365565	I28	0.11732582	I43	0.43955672	I58	0.33065477
I14	1.10155167	I29	-1.11575480	I44	0.40303243	I59	0.29475640
I15	0.31268366	I30	1.36722576	I45	0.08211176	I60	-0.41483463

Biểu đồ Histogram cho độ khó như sau:



Hình 3.11: Biểu đồ Histogram cho tham số độ khó của mô hình IRT 1 tham số sau khi ước lượng bằng MMLE

Tiếp theo ta ước lượng năng lực của thí sinh bằng phương pháp EAP. Và thu được biểu đồ Histogram:



Hình 3.12: Biểu đồ Histogram cho năng lực thí sinh của mô hình IRT 1 tham số sau khi ước lượng bằng EAP

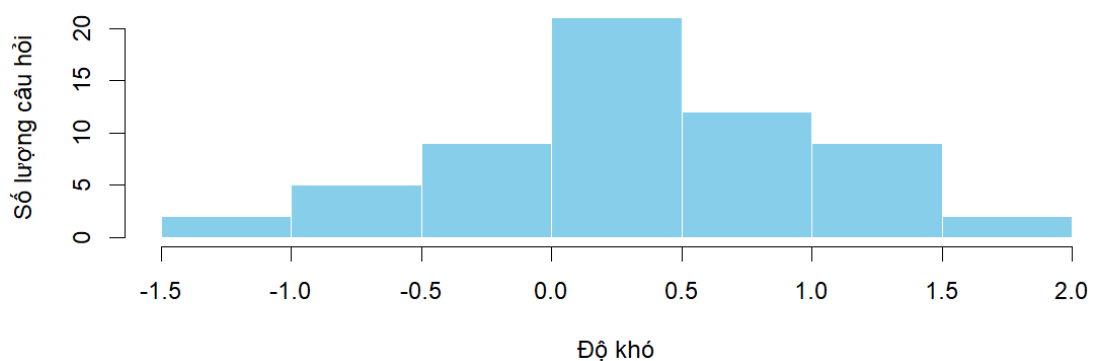
3.2.2.2. Mô hình IRT 2 tham số

Với mô hình irt 2pl sau khi sử dụng MMLE, sau 26 lần lặp ta thu được bảng kết quả ước lượng tham số độ khó và độ phân biệt như sau:

Bảng 3.6: Các tham số trong mô hình IRT 2 tham số sau khi ước lượng bằng MMLE

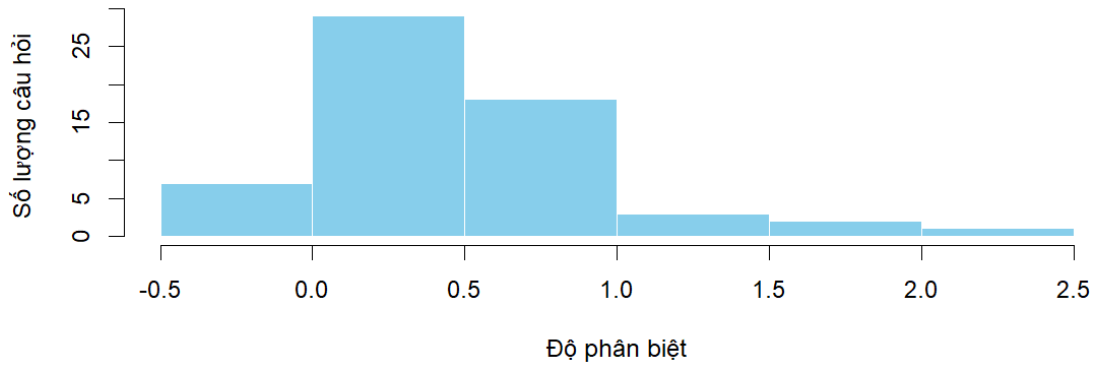
Item	Độ phân biệt	Độ khó	Item	Độ phân biệt	Độ khó	Item	Độ phân biệt	Độ khó
I1	0.3974	-0.7133	I21	-0.2000	1.4462	I41	0.6464	0.2600
I2	0.3076	0.1794	I22	-0.3067	1.2567	I42	0.4290	0.2346
I3	1.1293	0.1924	I23	0.6323	0.7428	I43	0.1520	0.4182
I4	1.5804	0.3434	I24	0.3351	0.3726	I44	0.7314	0.4169
I5	0.3728	-0.1836	I25	0.8545	0.5696	I45	0.7592	0.0739
I6	0.7300	-0.4737	I26	0.1508	0.0926	I46	0.4192	1.0007
I7	0.4139	0.7127	I27	0.2694	0.2131	I47	1.2805	-0.8101
I8	0.7573	0.7426	I28	0.8217	0.1114	I48	0.1583	0.8663
I9	0.5711	-0.0862	I29	0.6854	-1.1794	I49	-0.0741	1.4090
I10	0.6161	0.3331	I30	-0.0927	1.3070	I50	0.1869	0.6176
I11	0.1059	0.4346	I31	0.9164	0.1304	I51	-0.0131	1.0491
I12	0.4777	-0.5710	I32	0.2637	-0.3704	I52	0.0223	1.1611
I13	0.6790	-0.1099	I33	0.3299	0.6845	I53	0.3054	0.3889
I14	0.8053	1.1836	I34	-0.2231	1.5907	I54	0.1406	1.7042
I15	0.1424	0.2964	I35	-0.0432	0.2440	I55	0.0831	-0.0923
I16	0.2899	1.2299	I36	1.5162	-0.3990	I56	0.2993	-1.3047
I17	0.2503	0.8333	I37	0.3389	0.7046	I57	0.4615	-0.9461
I18	0.7598	0.3413	I38	0.1529	0.8862	I58	0.6567	0.3353
I19	0.6212	0.6405	I39	1.0266	-0.0132	I59	0.2709	0.2824
I20	0.8468	0.9578	I40	2.0344	-0.9483	I60	0.0906	-0.3993

Biểu đồ Histogram cho độ khó như sau:



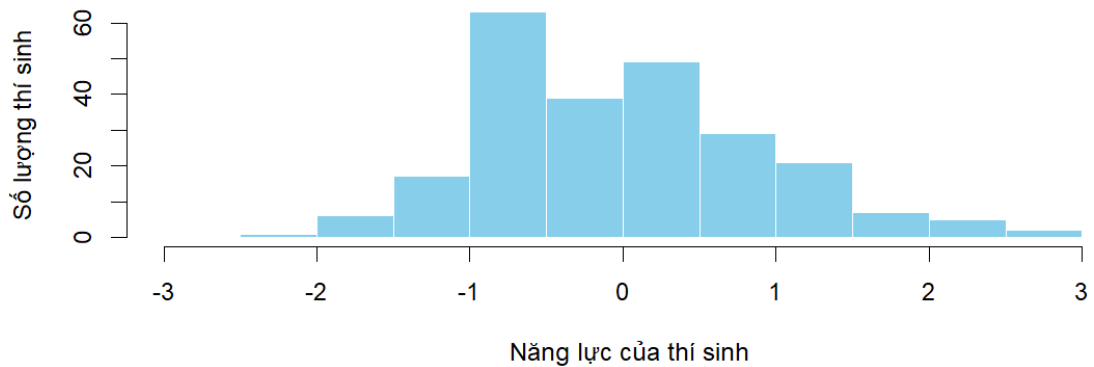
Hình 3.13: Biểu đồ Histogram cho tham số độ khó của mô hình IRT 2 tham số sau khi ước lượng bằng MMLE

và biểu đồ Histogram cho độ phân biệt như sau:



Hình 3.14: Biểu đồ Histogram cho tham số độ phân biệt của mô hình IRT 2 tham số sau khi ước lượng bằng MMLE

Tiếp theo ta ước lượng năng lực của thí sinh bằng phương pháp EAP. Và thu được biểu đồ Histogram:



Hình 3.15: Biểu đồ Histogram cho năng lực thí sinh của mô hình IRT 2 tham số sau khi ước lượng bằng EAP

3.2.2.3. Mô hình IRT 3 tham số

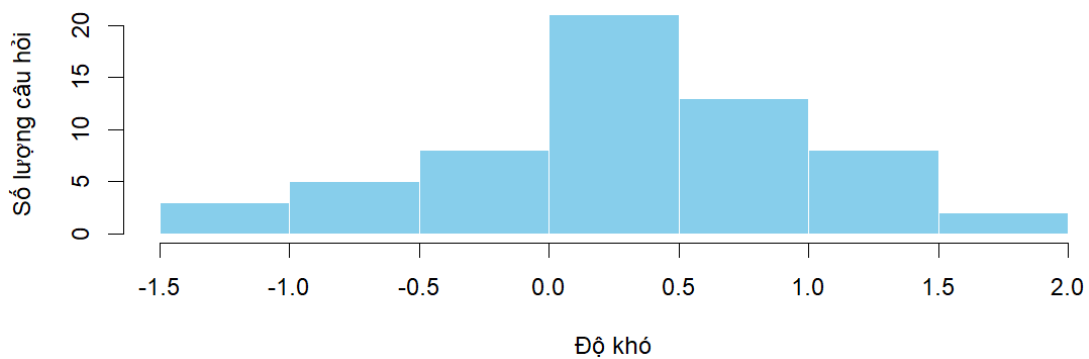
Ở đây tôi sẽ chỉ định cho câu 3, 7, 9, 12 sẽ được ước lượng tham số độ phân biệt. Với MMLE Sau 118 lần lặp ta thu được bảng kết quả ước lượng tham số độ khó và độ phân biệt như sau:

Bảng 3.7: Các tham số độ phân biệt và độ khó trong mô hình IRT 3 tham số sau khi ước lượng bằng MMLE

Item	Độ phân biệt	Độ khó	Item	Độ phân biệt	Độ khó	Item	Độ phân biệt	Độ khó
I1	0.6204	-0.7311	I21	-0.2949	1.4532	I41	1.0317	0.2319
I2	0.4709	0.1663	I22	-0.4933	1.2724	I42	0.6326	0.2165
I3	1.8185	0.1407	I23	0.9761	0.7157	I43	0.2345	0.4117
I4	2.4173	0.2730	I24	0.5149	0.3583	I44	1.1057	0.3848
I5	0.5718	-0.1993	I25	1.3325	0.5334	I45	1.1704	0.0409
I6	1.1281	-0.5056	I26	0.2344	0.0861	I46	0.6465	0.9828
I7	0.6440	0.6953	I27	0.4062	0.2018	I47	1.9913	-0.8716
I8	1.1741	0.7105	I28	1.2689	0.0755	I48	0.2580	0.8597
I9	0.8849	-0.1110	I29	1.0564	-1.2085	I49	-0.1152	1.4121
I10	0.9547	0.3066	I30	-0.1547	1.3117	I50	0.2929	0.6097
I11	0.1493	0.4303	I31	1.4100	0.0904	I51	-0.0293	1.0499
I12	0.7493	-0.5928	I32	0.4102	-0.3819	I52	0.0203	1.1604
I13	1.0720	-0.1410	I33	0.5149	0.6706	I53	0.4794	0.3759
I14	1.2563	1.1516	I34	-0.3343	1.5990	I54	0.2311	1.6988
I15	0.2067	0.2905	I35	-0.0614	0.2456	I55	-0.0959	-0.0923
I16	0.4678	1.2190	I36	2.3924	-0.4769	I56	-1.3185	-1.2085
I17	0.3834	0.8226	I37	0.5357	0.6906	I57	-0.9676	-0.9676
I18	1.1583	0.3082	I38	0.2455	0.8798	I58	0.3071	0.3071
I19	0.9335	0.6124	I39	1.5848	-0.0586	I59	0.2707	0.2707
I20	1.2674	0.9167	I40	3.0767	-1.0322	I60	0.1437	-0.4033

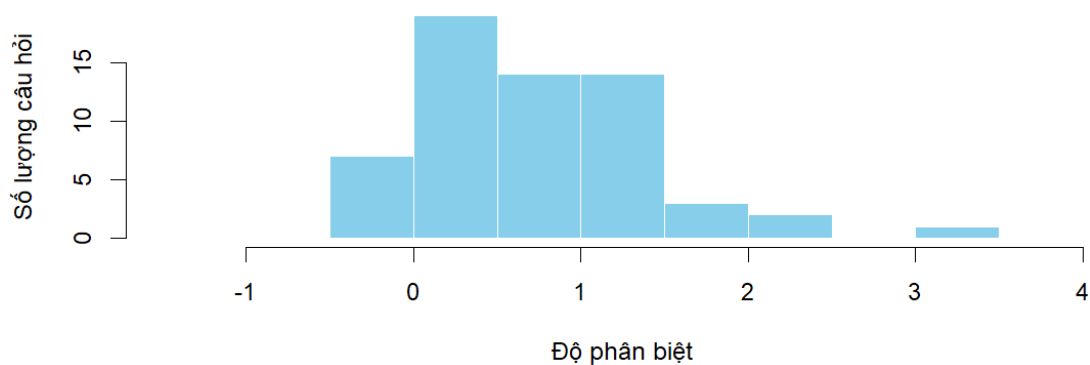
Và các độ khó cho các câu chỉ định ước lượng được là $c_3 = 0.0155683299$, $c_7 = 0.0008448663$, $c_9 = 0.0039931749$, $c_{12} = 0.0039931749$

Biểu đồ Histogram cho độ khó như sau:



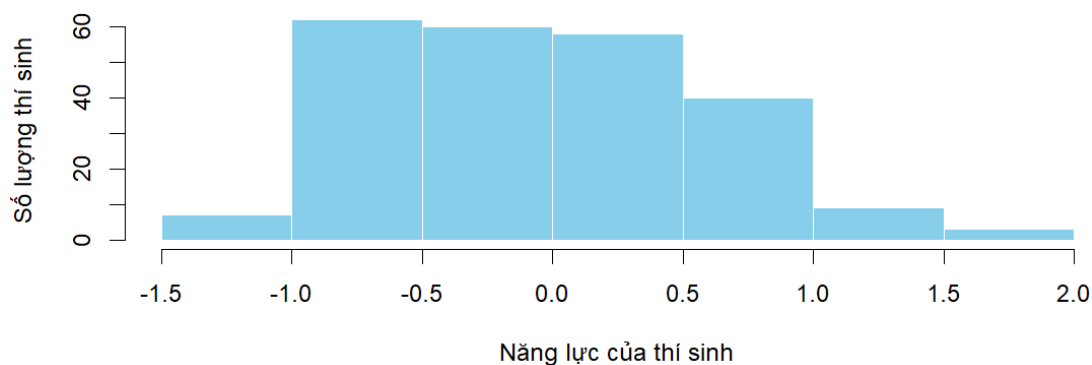
Hình 3.16: Biểu đồ Histogram cho tham số độ khó của mô hình IRT 3 tham số sau khi ước lượng bằng MMLE

và biểu đồ Histogram cho độ phân biệt như sau:



Hình 3.17: Biểu đồ Histogram cho tham số độ phân biệt của mô hình IRT 3 tham số sau khi ước lượng bằng MMLE

Tiếp theo ta ước lượng năng lực của thí sinh bằng phương pháp EAP. Và thu được biểu đồ Histogram:



Hình 3.18: Biểu đồ Histogram cho năng lực thí sinh của mô hình IRT 3 tham số sau khi ước lượng bằng EAP

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI

Trong luận văn này, chúng tôi đã trình bày ba mô hình IRT nhị phân cơ bản gồm IRT 1PL, 2PL và 3PL, cùng với các phương pháp ước lượng tham số như MLE, MMLE, MAP và EAP. Các công thức tính toán và ý tưởng ước lượng tham số câu hỏi dựa trên hướng dẫn của Baker (2001, 2017) cũng đã được phân tích và triển khai cụ thể.

Đối với phương pháp EAP, chúng tôi đề xuất sử dụng phương pháp xấp xỉ tích phân bằng Hermite-Gauss quadrature tương tự như trong phương pháp MMLE, nhằm đảm bảo tính hiệu quả và chính xác trong tính toán. Chúng tôi đã xây dựng và thực nghiệm mã nguồn Python cho các phương pháp MLE và MAP trên tập dữ liệu tự tạo để kiểm tra độ chính xác của mô hình. Với các phương pháp MMLE và EAP, chúng tôi sử dụng gói phần mềm TAM trong môi trường R, thực nghiệm trên tập dữ liệu Dicht_Data2.csv được tải từ nền tảng Kaggle. Kết quả thu được từ cả hai nhóm thực nghiệm đều cho thấy hiệu suất ước lượng tham số khá tốt.

Trong hướng nghiên cứu tiếp theo, chúng tôi sẽ tiếp tục mở rộng triển khai các phương pháp MLE và MAP bằng Python trên các bộ dữ liệu thực tế. Bên cạnh đó, việc nghiên cứu các mô hình IRT đa phân và các phương pháp ước lượng tương ứng cũng là một mục tiêu trọng tâm. Cuối cùng, chúng tôi mong muốn tích hợp lý thuyết IRT và các thuật toán kiểm tra thích ứng (Computerized Adaptive Testing - CAT) vào các hệ thống học tập trực tuyến nhằm cá nhân hóa trải nghiệm học tập và đánh giá.

TÀI LIỆU THAM KHẢO

- [1] Erling B Andersen. The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(1):42–54, 1972.
- [2] Frank B Baker. *The basics of item response theory*. ERIC, 2001.
- [3] Frank B Baker, Seock-Ho Kim, et al. *The basics of item response theory using R*, volume 10. Springer, 2017.
- [4] Ban Chấp hành Trung ương Đảng Cộng sản Việt Nam. Về đổi mới căn bản, toàn diện giáo dục và đào tạo, đáp ứng yêu cầu công nghiệp hóa, hiện đại hóa trong điều kiện kinh tế thị trường định hướng xã hội chủ nghĩa và hội nhập quốc tế, 2013. Truy cập ngày 16/03/2024. URL: <https://moet.gov.vn/tintuc/Pages/doi-moi-can-ban-toan-dien-gd-va-dt.aspx?ItemID=3928>.
- [5] Hoang Tieu Binh and Bui The Duy. Student ability estimation based on irt. In *2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS)*, pages 56–61. IEEE, 2016.
- [6] R Darrell Bock and Murray Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, 46(4):443–459, 1981.
- [7] R Darrell Bock and Robert J Mislevy. Bilog: Maximum likelihood item analysis and test scoring with binary logistic models. *Chicago: International Educational Services*, 1981.
- [8] Anna Brown and Tim J. Croudace. Scoring and estimating score precision using irt. In S.P. Reise and D.A. Revicki, editors, *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, chapter 15. Routledge, New York, NY, 2015.
- [9] Bộ Giáo dục và Đào tạo. Hướng dẫn thực hiện Quy chế thi THPT quốc gia và xét công nhận tốt nghiệp THPT năm 2017, 2017. Truy cập ngày 16/03/2024. URL: <https://moet.gov.vn/tintuc/Pages/>

tin-tong-hop.aspx?ItemID=4479.

- [10] R Philip Chalmers. mirt: A multidimensional item response theory package for the r environment. *Journal of statistical Software*, 48:1–29, 2012.
- [11] Nguyễn Văn Cảnh and Nguyễn Quốc Tuấn. Ứng dụng phần mềm conquest với mô hình irt hai tham số vào việc đánh giá chất lượng đề thi trắc nghiệm khách quan. *HNUE Journal of Science - Educational Sciences*, 65(7):230–242, 2020. doi:10.18173/2354-1075.2020-0093.
- [12] R Darrell Bock and Marcus Lieberman. Fitting a response model for n dichotomously scored items. *Psychometrika*, 35(2):179–197, 1970.
- [13] Ditiiep.com. Đánh giá năng lực người học, 2021. Truy cập: 09-03-2025. URL: <https://ditiiep.com/danh-gia-nang-luc-nguoi-hoc/>.
- [14] Đỗ Anh Dũng. Đổi mới kiểm tra đánh giá theo định hướng tiếp cận năng lực học sinh, 2019. Truy cập: 09-03-2025.
- [15] Susan E Embretson and Steven P Reise. *Item response theory for psychologists*. Psychology Press, 2013.
- [16] Rasch Georg. Probabilistic models for some intelligence and attainment tests. *Copenhagen: Institute of Education Research*, 1960.
- [17] Ronald K Hambleton and Russell W Jones. Comparison of classical test theory and item response theory and their applications to test development. *Educational measurement: issues and practice*, 12(3):38–47, 1993.
- [18] Ronald K Hambleton, Hariharan Swaminathan, and H Jane Rogers. *Fundamentals of item response theory*, volume 2. Sage, 1991.
- [19] Michael R Harwell, Frank B Baker, and Michael Zwarts. Item parameter estimation via marginal maximum likelihood and an em algorithm: A didactic. *Journal of Educational Statistics*, 13(3):243–271, 1988.
- [20] Lê Thị Hoàng Hà. Lý thuyết ứng đáp câu hỏi (item response theory – irt) được sử dụng trong các khâu nào và nhằm mục đích gì trong quy trình thiết kế một bài trắc nghiệm chuẩn hóa?, 2019. Truy cập: 09-03-2025. URL: <http://qm.education.vnu.edu.vn/ly-thuyet-ung-dap-cau-hoi>.
- [21] Lê Thị Bích Hằng, Nguyễn Đình Hưng, Phạm Thị Kim Ngoan, and Bùi

Thị Hồng Minh. Một giải pháp hỗ trợ phân tích, đánh giá câu hỏi trắc nghiệm khách quan trên hệ thống moodle theo mô hình irt. *Tạp chí Giáo dục*, 22(16):17–23, 2022.

- [22] Khoa Toán Tin, Trường Đại học Sư phạm Hà Nội. *Bài giảng Thống kê Xã hội học*. Trường Đại học Sư phạm Hà Nội, 2020.
- [23] KshitizKhanal. Dichotomous data for item response theory, 2022. Truy cập: 19-03-2025. URL: <https://www.kaggle.com/datasets/kshitiz05/dichotomousdataforitemresponsetheroy>.
- [24] Frederic Lord. A theory of test scores. *Psychometric monographs*, 1952.
- [25] Frederic M Lord. *Applications of item response theory to practical testing problems*. Routledge, 1980.
- [26] Nguyễn Thị Hồng Minh and Nguyễn Đức Thiện. Đo lường-Đánh giá trong thi trắc nghiệm khách quan: Độ khó câu hỏi và khả năng của thí sinh. *VNU Journal of Science: Natural Sciences and Technology*, 22(4), 2006.
- [27] Kevin P Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.
- [28] Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: journal of the Econometric Society*, pages 1–32, 1948.
- [29] Simon JD Prince. *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.
- [30] Norman Rose. Maximum likelihood and bayes modal ability estimation in two-parametric irt models: Derivations and implementation. *Schriften zur Bildungsf*, 2010.
- [31] Hariharan Swaminathan and Janice A Gifford. Bayesian estimation in the two-parameter logistic model. *Psychometrika*, 50(3):349–364, 1985.
- [32] Hariharan Swaminathan and Janice A Gifford. Bayesian estimation in the three-parameter logistic model. *Psychometrika*, 51:589–601, 1986.
- [33] Nguyễn Bảo Hoàng Thanh and Nguyễn Thị Thảo. Phân tích câu hỏi trắc nghiệm khách quan dùng đánh giá năng lực vật lí của học sinh trong dạy học

chương “sóng ánh sáng”–vật lí 12 bằng phần mềm quest/conquest. *Tạp chí Khoa học và Công nghệ-Đại học Đà Nẵng*, pages 34–39, 2023.

- [34] Lâm Quang Thiệp. Đo lường trong giáo dục lý thuyết và ứng dụng. *NXB Đại học quốc gia Hà Nội*, 2010.