"Research article" doi: 10.30495/jinev.2022.1956750.2704

Risk of Metropolis-Hastings Robbins-Monro algorithm in multidimensional models of item-response theory in binary data considering test length[1]

Mehdi Molaei Yasavoli[2]*, Ali Delavar[3], Mohammad asgari[4], Jalil yonesi[5], Vahid Rezaei Tabar[6]

(Received: 2022.04.17 - Accepted: 2022.10.31)

1- This article is extracted from the doctoral dissertation of Mehdi Moulai Yesavali, a PhD student in the field of measurement and measurement at Allameh Tabatabai University.

2- Ph. D. Student in Measurement and Assesment (Psychometric), Faculty of Psychology and Educational Sciences, Allameh Tabatabai University, Tehran, Iran
*- Corresponding Author: molaei_m@atu.ac.ir
3- Professor, Department of Measurement and assesment, Faculty of Psychology and Educational Sciences, Allameh Tabatabai University, Tehran, Iran.
4- Associate Professor, Department of Measurement and Assesment, Faculty of Psychology and Educational Sciences, Allameh Tabatabai University, Tehran, Iran.
5- Associate Professor, Department of Measurement and Assesment, Faculty of Psychology and Educational Sciences, Allameh Tabatabai University, Tehran, Iran.
6- Associate Professor, Department of Statistics, Faculty of Statistics, Mathematics and Computer, Allameh Tabatabai University, Tehran, Iran.

Abstract

The present study was conducted with the aim of investigating the risk of MHRM algorithm in multi-dimensional models of item-response theory in binary data, taking into account different test dimensions and lengths. The research method used was a real experiment using a multi-group post-test design. The studied sample was created based on simulation studies under different conditions of independent variables in 27 modes with 100 repetitions for each. The model used was the two-parameter multidimensional model of logistics and the investigated parameters were the slope and difficulty of the items. In order to check the risk of each of the parameters in different experimental conditions, the average squared error index was used. R statistical software packages mirt, interactions, car and psych were used for data generation and analysis. The results of the research showed that the MHRM algorithm has less risk compared to the EM and MCEM algorithms. This issue was especially evident under the conditions of high dimensional data (5 dimensions) and short test length (15 questions). Also, the results of the research showed that when the dimensions of the test increase and the length of the test decreases, the risk of parameter estimation increases significantly. As a result, it can be said that the application of the MHRM algorithm in data with a high number of dimensions and a short test length is necessary, and researchers are advised to use it in the analysis of data with a complex structure such as a high number of dimensions.

***Keywords*:** MHRM algorithm, risk, multidimensional models of item-response theory, binary data, test length

«مقاله پژوهشی» doi: 10.30495/jinev.2022.1956750.2704

مخاطره الگوریتم متروپلیس هستینگز روبینز مونرو در مدل‌های چندبعدی نظریه سؤال پاسخ در داده‌های دو ارزشی با در نظر گرفتن طول آزمون [۱]

مهدی مولایی یساولی[۲*]، علی دلاور[۳]، محمد عسگری[۴]، جلیل یونسی[۵]، وحید رضایی تبار[۶]

(دریافت: ۱۴۰۱/۰۱/۲۸ ـ پذیرش: ۱۴۰۱/۰۸/۰۹)

چکیده

پژوهش حاضر با هدف بررسی مخاطره الگوریتم MHRM در مدل‌های چندبعدی نظریه سؤال پاسخ در داده‌های دو ارزشی با در نظر گرفتن ابعاد و طول آزمون متفاوت مورد بررسی قرار گرفت. روش پژوهش مورد استفاده آزمایشی واقعی و با استفاده از طرح پس‌آزمون چند گروهی بود. نمونه مورد مطالعه براساس مطالعات شبیه‌سازی تحت شرایط مختلف متغیرهای مستقل در ۲۷ حالت با ۱۰۰ تکرار برای هر کدام ایجاد شد. مدل مورد استفاده مدل دو پارامتری چندبعدی لوجستیک و پارامترهای مورد بررسی شیب و دشواری سؤالات بود. جهت بررسی مخاطره هر یک از پارامترها در حالت‌های مختلف آزمایشی شاخص میانگین توان دوم خطاها مورد استفاده قرار گرفت. جهت تولید و تحلیل داده‌ها ار نرم‌افزارهای آماری R بسته‌های mirt، interactions، car و psych استفاده شد. نتایج پژوهش نشان داد الگوریتم MHRM در قیاس با الگوریتم‌های EM و MCEM دارای مخاطره کمتری است. این موضوع بویژه تحت شرایط داده‌هایی با ابعاد بالا (۵ بعد) و طول آزمون کوتاه (۱۵ سؤال) بیشتر مشهود بود. همچنین نتایج پژوهش نشان داد زمانی که ابعاد آزمون افزایش و طول آزمون کاهش می یابد، مخاطره برآورد پارامترها به طور معنی‌داری افزایش می‌یابد. در نتیجه می‌توان گفت کاربرد الگوریتم MHRM در داده‌های با تعداد ابعاد بالا و طول آزمون کوتاه ضروری است و به پژوهشگران توصیه می‌شود که از آن در تحلیل داده‌های با ساختار پیچیده از قبیل تعداد ابعاد بالا بهره گیرند.

واژگان کلیدی: الگوریتم MHRM، مخاطره، مدل‌های چند بعدی نظریه سؤال پاسخ، داده‌های دو ارزشی، طول آزمون

۱- این مقاله مستخرج از رساله دکتری مهدی مولایی یساولی دانشجوی دکتری تخصصی رشته سنجش و اندازه گیری دانشگاه علامه طباطبائی می‌باشد.

۲- دانشجوی دکتری سنجش و اندازه گیری (روانسنجی)، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران
*- نویسنده مسئول: molaei_m@atu.ac.ir
۳- استاد تمام گروه سنجش و اندازه گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران.
۴- دانشیار گروه سنجش و اندازه گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران.
۵- دانشیار گروه سنجش و اندازه گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران.
۶- دانشیار گروه آمار، دانشکده آمار، ریاضی و رایانه، دانشگاه علامه طباطبائی، تهران، ایران.

مقدمه

اندازه‌گیری ویژگی‌های روانشناختی انسانی از قبیل هوش، خلاقیت، پیشرفت تحصیلی، ویژگی‌های شخصیتی طی حدودا دویست سال گذشته مورد توجه علم روانشناسی و بویژه متخصصین روانسنجی بوده است. یکی از مهمترین مسائل در حوزه اندازه‌گیری‌های علوم رفتاری، برآورد دقیق پارامترهای موجود در مدل‌های آماری است. در آمار برآوردی دارای ارزش و اعتبار است که دارای ویژگی ثبات، کارایی، کفایت و نداشتن اریب باشد (دلاور، ۱۴۰۱). یک از شرایطی که برآورد پارامترها را تحت تأثیر قرار می‌دهد، وجود ساختار عاملی است که این وضعیت در سازه‌های انسانی به فراوانی یافت می‌شود. وجود ساختار عاملی بر پیچیدگی‌های محاسباتی در برآورد پارامترها می‌افزاید و پدیده نفرین[1] ابعاد را به دنبال دارد. مسأله نفرین ابعاد توسط ریچارد بلمن[2] در سال ۱۹۵۷ مطرح شد تا رشد فوق‌العاده سریع مشکلات را با افزایش تعداد متغیرها (یا بعد) توصیف کند (کو و اسلوان[3]، ۲۰۰۵). نفرین ابعاد به پدیده-های مختلفی اشاره دارد که هنگام تحلیل و سازماندهی داده‌ها در فضاهای با ابعاد بالا در حوزه‌هایی مانند تجزیه و تحلیل عددی[4]، یادگیری ماشین[5] و داده کاوی[6] بوجود می‌آیند. در نتیجه افزایش تعداد ابعاد آزمون نیز بر پیچیدگی‌های محاسباتی و میزان خطا و دقت برآورد نقش دارد (لیو و پیرس[7]، ۱۹۹۴ و کای، ۲۰۱۰).

چالش دیگری که در اندازه‌گیری‌های علوم رفتاری مشاهده می‌شود، به طول آزمون‌های روانشناختی بر می‌گردد. غالباً زمانی که تعداد سؤالات یک آزمون روانشناختی افزایش می‌یابد، بر آورد پارامترها با دقت بالاتری صورت است و خطای اندازه‌گیری کاهش می‌یابد. از سویی دیگر بدلیل اینکه یکی از اهداف مهم در اندازه‌گیری‌های ویژگی‌های انسانی کیفیت پاسخ‌دهی آزمودنی‌ها است، همواره طول آزمون کوتاه مد نظر متخصصین روانسنجی بوده است. زیرا آزمون‌های طولانی که نیازمند زمان پاسخگویی بالاتر و تمرکز بیشتری است، ممکن است خستگی و عدم تمایل به پاسخگویی را به همراه داشته باشد. ساهین و آنیل (۲۰۱۷) بیان می‌کنند که در مدل‌های نظریه سؤال پاسخ بویژه زمانی که حجم نمونه کم باشد، آزمون‌های با طول کوتاه‌تر دارای دقت اندازه‌گیری پایین‌تر و خطاهای بزرگتری هستند. طول بهینه آزمون یکی از مسائل مهمی است که در زمینه روانسنجی و بویژه برآورد پارامترها اهمیت دارد (پاتسولا[8]،

1- Curse of dimensionality
2- Richard Bellman
3- Kuo & Sloan
4- Numerical Analysis
5- Machine Learning
6- Data Mining
7- Liu & Pierce
8- Patsula

۱۹۹۵ و ساهین و آنیل[1]، ۲۰۱۷). لرد (۱۹۶۸) افزایش طول آزمون را در برآورد دقیق پارامترها مؤثر می-داند و حداقل طول مورد نیاز در نظریه‌های سؤال پاسخ را ۵۰ مورد ذکر می‌کند (به نقل از ساهین و آنیل، ۲۰۱۷).

در پاسخ به مسائل ذکر شده، متخصصین روانسنجی در چارچوب نظریه سؤال پاسخ روش‌های نوینی را ارائه کرده‌اند. نظریه سؤال پاسخ به عنوان یک نظریه روانسنجی جدید با برخورداری از مدل‌های متنوع و پیشرفته و کاربرد الگوریتم‌های متنوع در محاسبات ریاضیاتی، به طور گسترده در آزمون‌های آموزشی و اندازه‌گیری، ارزیابی روانی مورد استفاده قرار گرفته است (بارتولوچی و همکاران[2]، ۲۰۱۵؛ گیبونز و همکاران[3]، ۲۰۱۶). در سال‌های گذشته چندین روش برای ایجاد امکان برآورد مدل‌های نظریه سؤال پاسخ با ابعاد بالا و کارآیی بیشتر در فرآیند تخمین، توسعه یافته‌اند. برای چند دهه، مدل‌های نظریه سؤال پاسخ با استفاده از: (۱) بیشینه احتمال حاشیه‌ای[4] بر اساس الگوریتم بیشینه انتظار[5]، (۲) الگوریتم بیشینه انتظار مونت کارلو (MCEM) و (۳) برآورد بیزین کامل[6] با استفاده از شبیه‌سازی مونت کارلوی زنجیره مارکوفی[7] (MCMC) تخمین زده می‌شود.

الگوریتم EM برای یافتن پارامترهای بیشینه درستمنایی یک مدل آماری در مواردی که معادلات به طور مستقیم حل نمی‌شوند، استفاده می‌شود (دمپستر و همکاران[8]، ۱۹۷۷). الگوریتم EM یکی از روش‌هایی است که براساس وجود متغیر پنهان امکان برآورد پارامترهای مدل آماری را میسر می‌سازد. در این الگوریتم از نقاط کوادراتور گاوسی–هرمیت ثابت، برای برآورد استفاده می‌شود ولیکن زمانی که تعداد ابعاد به چهار یا پنج (یا بیشتر) افزایش یابد، فرایند تخمین با مشکل روبه رو می شود. زیرا تعداد کل نقاط کوادره با قدرتی برابر با تعداد ابعاد افزایش می یابد و ارزیابی انتگرال‌ها را بسیار دشوار می‌کند (لیو و پیرس، ۱۹۹۴؛ نیلور و اسمیت[9]، ۱۹۸۲). الگوریتم MCEM نیز الگوریتم EM است که در آن انتظار در مرحله E از طریق شبیه سازی مونت کارلو به صورت عددی محاسبه می‌شود، توسط منگ و اسچیلینگ[10] (۱۹۹۶) ایجاد شده است. این الگوریتم، انعطاف پذیرترین و عموماً کاربردی ترین روش برای بدست آوردن نمونه مونت کارلو در هر تکرار الگوریتم MCEM از طریق روال‌های زنجیره‌ای مارکوف مونت کارلو مانند نمونه‌بردارهای گیبس و متروپولیس– هستینگز است (لوین و کسلا[11]، ۲۰۰۱).

1- Sahin & Anil
2- Bartolucci, Bacci & Gnaldi
3- Gibbons, Weiss, Frank & Kupfer
4- Marginal Maximum Likelihood (MML)
5- Expectation Maximization (EM)
6- Fully Bayesian estimation
7- Markov Chain Monte Carlo estimation
8- Dempster, Laird & Rubin
9- Naylor & Smith
10- Meng, & Schilling
11- Levine, & Casella

در الگوریتم بیشینه انتظار مونت کارلو (MCEM) برای رسیدن به همگرایی نقطه ای برآورد پارامترها، اندازه شبیه‌سازی (به عنوان مثال، تعداد نمونه‌گیری‌های تصادفی) باید بسیار افزایش یابد، به ویژه در چند تکرار آخر، زیرا تخمین های پارامتر به حداکثر تابع احتمال نزدیکتر می شوند (کای، ۲۰۱۰). علاوه بر این، زمان همگرایی MCEM به این دلیل افزایش می‌یابد که برای هر تکرار مرحله E، نمونه‌گیری مجموعه جدیدی از قرعه کشی‌های تصادفی را ایجاد می‌کند.

رویکرد بعدی، یک رویکرد کاملاً بیزین برای تخمین مدل‌های چندبعدی نظریه سؤال پاسخ است. در تجزیه و تحلیل بیزین، اطلاعات قبلی و کلیه داده‌های موجود در توزیع‌های پسین (با استفاده از قانون بیز[1]) ادغام می‌شوند که می‌توان از آنها در استنتاج‌های خود استفاده کرد. الگوریتم MCMC برای محاسبه تقریب عددی انتگرال‌های چندبعدی استفاده می شود (کاسیم و همکاران[2]، ۲۰۱۹). با این حال، یک چالش بزرگ و اساسی، ویژگی های توزیع پیشنهادی مناسب است، که ممکن است نیاز به تجربه با آزمون و خطا باشد. علاوه بر این، از آنجا که MCMC هنوز هم کل سطح پاسخ را در فضای چندبعدی تقریب می‌زند، ممکن است برای مشکلات چند متغیره هنوز به زمان محاسبات گسترده ای نیاز داشته باشد، که استفاده از آن را در عمل دشوار می کند. سرانجام، ارزیابی همگرایی در MCMC اعمال شده در مدل‌های پیچیده با تعداد زیادی از سؤالات یا تعداد بالایی از ابعاد پنهان، می تواند دشوار باشد و نیاز به قضاوت انسان دارد (باشکوف و دمارس[3]، ۲۰۱۷).

در پاسخ به محدودیت‌های روش‌های محبوب برآورد در مدل‌های چندبعدی نظریه سؤال پاسخ که در بالا توضیح داده شد، روش جدیدی توسط کای (۲۰۱۰) توسعه داده شد. کای (۲۰۱۰) یک الگوریتم متروپلیس–هستینگز روبینز–مونرو[4] (MHRM) تهیه کرد، که نمونه گیری متروپولیس– هستینگز (پاتز و یونکر[5]، ۱۹۹۹) را با نمونه گیری روبینز–مونرو (رابینز و مونرو[6]، ۱۹۵۱) ترکیب می‌کند تا بیشینه درست-نمایی را تسهیل کند. الگوریتم متروپلیس– هستینگز روبینز–مونرو (MHRM) یک الگوریتم ترکیبی مبتنی بر احتمال حاشیه‌ای است که در آن نمونه‌های توزیع شرطی ابعاد از طریق تقریب تصادفی ترکیب می شوند (کای، ۲۰۱۰). MHRM برخلاف MCMC که کل توزیع پسین را تقریب می زند، بر برآورد نقطه ای و خطاهای استاندارد تمرکز دارد و امکان را فراهم می کند که MHRM خیلی سریعتر از MCMC به همگرایی برسد. به طور مشخص، کاربرد الگوریتم MHRM، جهش ها در زنجیره مارکوف به منظور تعیین جهت تغییر در طول تکرارها است (کای، ۲۰۱۰).

1- Bayes
2- Kasim, Tzeferacos, Lamb, Gregori & Vinko
3- Bashkov, B. M., & DeMars
4- Metropolis-Hastings Robbins-Monro
5- Patz & Junker
6- Robbins & Monro

در رابطه با کارامدی الگوریتم MHRM پژوهش‌های محدودی صورت گرفته است. یانگ و کای[1] (۲۰۱۴) در پژوهشی به مطالعه کاربرد الگوریتم MHRM در بهینه کردن برآورد بیشینه برآورد درستنمایی حاشیه‌ای در چارچوب مدل متغیر پنهان چند سطحی غیرخطی پرداختند. نتایج نشان می‌دهد که الگوریتم MHRM می تواند تخمین ها و خطاهای استاندارد را به طور کارآمد تولید کند. کو و شنگ[2] (۲۰۱۶) نیز در پژوهشی با عنوان "مقایسه روش های تخمین برای یک مدل پاسخ درجه بندی چند بعدی چند بعدی" به بررسی توانمندی الگوریتم‌های متفاوت در مدل چندبعدی پاسخ مدرج نظریه سؤال پاسخ پرداختند. در این پژوهش روش های برآورد EM و MHRM تحت شرایط متفاوت کیفیت داده مورد مقایسه قرار گرفتند و توانمندی هر یک از روش های برآورد مورد ارزیابی قرار گرفت. باشکوف و دمارس (۲۰۱۷) در پژوهشی با عنوان "آزمون کارایی الگوریتم متروپلیس هستینگر روبینز مرنرو در برآورد مدل های چند بعدی چند سطحی" کارایی الگوریتم متروپلیس هستینگز روبینز مرنرو را در داده‌های دو ارزشی با استفاده از مدل سه پارامتری مورد ارزیابی قرار دادند و کارایی MHRM در بازیابی پارامترهای سؤال، واریانس و کوواریانس پنهان، و همچنین برآورد توانایی درون خوشه ها و خوشه ها (به عنوان مثال، مدارس) در یک مطالعه شبیه سازی، بررسی کردند که نتایج نشان داد، الگوریتم MHRM عملکرد خوبی در برآورد سؤال، آزمودنی و پارامترهای مدل در سطح گروه داشت.

در کل مطالعات نشان داده‌اند که استفاده از روشهای برآورد ذکر شده باعث می‌شود فرآیند تخمین برای مدل‌های چندبعدی از نظر محاسباتی دشوار[3] و غالباً غیرقابل حل[4] باشد که تعداد زیادی از ابعاد در آن دخیل باشد (کای[5]، ۲۰۱۰؛ چالمرز[6]، ۲۰۱۲). در نتیجه الگوریتم MHRM برای غلبه بر مشکلات برآورد بیشینه احتمال و سایر الگوریتم های مرسوم و ارائه تخمین‌های مفید برای داده‌هایی با تعداد متفاوتی از سؤال، ابعاد مختلف و یا داده‌های گمشده ارائه شد (کو و شنگ، ۲۰۱۶). در نتیجه با عطف به پژوهش‌های انجام شده و نظر به چالش های موجود، در این پژوهش از جهت تجزیه و تحلیل مدل‌های چندبعدی چندسطحی نظریه سؤال پاسخ[7] از MHRM در کنار الگوریتم های EM و MCEM استفاده شد و به بررسی این موضوع اساسی پرداخته شد که آیا الگوریتم MHRM در قیاس با الگوریتم های مرسوم دیگر یعنی الگوریتم‌های EM و MCEM با در نظر گرفتن تعداد ابعاد مختلف و طول متفاوت آزمون عملکرد بهتری دارد؟ و به عبارتی آیا برآوردهای دقیقتری با میزان سوگیری کمتر و کارایی بیشتری به دست می دهد؟.

1- Yang, & Cai
2- Kuo, & Sheng
3- Intensive
4- Intractable
5- Cai
6- Chalmers
7- Item Response Theory Multilevel Multidimensional Models

روش تحقیق

پژوهش حاضر از نظر فلسفی اثبات گرایانه، هدف کاربردی، روش کمی، از نوع آزمایشی واقعی و به دنبال مقایسه پدیده مورد مطالعه خواهد بود. هدف اصلی از کاربرد روش آزمایشی در این پژوهش، تعیین شرایطی که در آن پدیده معینی مبتنی بر مدل های نظریه سؤال پاسخ در مطالعات شبیه سازی اتفاق می‌افتد. طرح پژوهشی مورد استفاده در این پژوهش عاملی و از نوع سه راهه می باشد که متغیرهای مستقل عبارتند از نوع الگوریتم، تعداد ابعاد و طول آزمون.

در این پژوهش به منظور مقایسه میزان مخاطره الگوریتم MHRM از داده‌های شبیه‌سازی شده مونت کارلو استفاده شده است. دلیل استفاده از داده‌های شبیه‌سازی شده مونت کارلو این بود که اساساً دستیابی به داده‌های واقعی آزمون های متفاوت برای مطالعات روش‌شناسی مقایسه الگوریتم ها در شرایط آزمایشی متفاوت بسیار دشوار است. زیرا ایجاد شرایط واقعی آزمون ها برای تعداد زیادی از آزمودنی‌ها با تکرارهای بسیار زیاد و در شرایط یکسان کار بسیار دشواری است و فقط در تخیل می‌تواند، صورت بگیرد اما استفاده از مطالعات مونت‌کارلو این تخیل را در محیطی شبیه سازی شده و تحت کنترل به واقعیت تبدیل می‌کند. این روش سریع‌تر، ارزان‌تر و آسان‌تر از جمع‌آوری اطلاعات از آزمودنی‌های زنده است و به محقق اجازه می‌دهد، مدل‌ها و روش‌های جدید روانسنجی را به سرعت و با هزینه بسیار کمی مورد پژوهش قرار دهند.

مدل مورد استفاده در این پژوهش بدلیل نوع داده‌ها (دو ارزشی)، مدل دو پارامتری چندبعدی لوجستیک بود که دو پارامتر شیب و دشواری را در خود جای دارد. این مدل تعمیم یافته مدل لوجستیک دوپارامتری تک بعدی است که در آن احتمال اینکه فرد j به آیتم i پاسخ مثبت بدهد، توسط:

$$P_{ij} = \frac{\exp(d_i + a_{i1}\theta_{j1} + \cdots + a_{iD}\theta_{jD})}{1 + exp(d_i + a_{i1}\theta_{j1} + \cdots + a_{iD}\theta_{jD})}$$

معین می شود. که در آن D تعداد عامل ها، θj1، ...، θjD پارامترهای مربوط به فرد، di دشواری سؤال، و ai1، ...، aiD شیب سؤالات هستند.

در تولید داده‌ها، شکل توزیع توانایی آزمودنی‌ها براساس توزیع نرمال با میانگین صفر و انحراف معیار یک در نظر گرفته شد. تعیین شکل توزیع توانایی آزمودنی ها معمولاً از توزیع یک جامعه مشخص به صورت تصادفی انتخاب می‌شود که اکثر اوقات توضیح استاندارد است (هان و همیلتون[1]، ۲۰۱۴). به لحاظ نظری هر یک از توزیع های مختلف می توانند به عنوان توزیع پیشین فرض شوند، اما اغلب به نظر می رسد توزیع طبیعی انتخاب درست تری باشد (امبرستون و رایس[2]، ۲۰۱۳). پارامترهای سؤالات (شیب و

1- Han, & Hambleton
2- Embretson, & Reise

دشواری) نیز براساس پیشنهاد بولوت و سونبل[1] (۲۰۱۷) عمل شد. مقادیر پارامتر شیب براساس توزیع لوگ نرمال با میانگین ۰/۳ و انحراف معیار ۰/۲ و پارامتر دشواری براساس توزیع نرمال با میانگین صفر و انحراف معیار یک در نظر گرفته شد.

جامعه آماری و حجم نمونه در این پژوهش مخاطره پارامترهای شیب و دشواری سؤالات شبیه سازی شده ای هستند که در شرایط مختلف با توجه به متغیرهای نوع الگوریتم، طول آزمون و تعداد ابعاد در ۲۷ حالت ایجاد می شود، بود. در این بخش دادها براساس یک طرح متقاطع با طول آزمون ۱۵، ۴۵ و ۱۲۰ و تعداد ابعاد ۱، ۳ و ۵ تولید شدند که در معرض برآورد پارامترها با ۳ روش الگوریتم EM، MCEM و MHRM قرار گرفتند. دلیل انتخاب طول آزمون و تعداد ابعاد مراجعه به ادبیات پژوهش بود. طول آزمون ۱۵ به عنوان نماینده آزمون با طول کوتاه، تعداد ۴۵ نماینده آزمون های با طول متوسط و ۱۲۰ سؤالی نماینده آزمون های با طول بلند بود. از نظر بعد نیز براساس آزمون هایی مانند مقیاس نئو ۵ بعد به عنوان آزمون های با ابعاد بالا در نظر گرفته شد. لازم به ذکر است هر وضعیت، دارای ۱۰۰ تکرار بود که در کل ۲۷۰۰ تکرار صورت گرفت. در ادامه حالت‌های مختلف تولید داده در جدول ۱ ارایه شده است.

جدول ۱: حالت‌های مختلف تولید داده‌های شبیه سازی شده

Table 1

Different modes of generating simulated data

نوع الگوریتم Algorithm type	تعداد ابعاد Dim number	طول آزمون Test lenght	تکرار replication	نوع الگوریتم Algorithm type	تعداد ابعاد Dim number	طول آزمون Test lenght	تکرار replication	نوع الگوریتم Algorithm type	تعداد ابعاد Dim number	طول آزمون Test lenght	تکرار replication
EM	1	15	100	MCEM	1	15	100	MHRM	1	15	100
		45	100			45	100			45	100
		120	100			120	100			120	100
	3	15	100		3	15	100		3	15	100
		45	100			45	100			45	100
		120	100			120	100			120	100
	5	15	100		5	15	100		5	15	100
		45	100			45	100			45	100
		120	100			120	100			120	100

برای پاسخگویی به سؤالات پژوهش با توجه به طرح های تحقیق (عاملی) از تحلیل واریانس عاملی سه راهه در حالت چندمتغیری استفاده شد. به طور خاص، مخاطره[2] پارامترها (شیب و دشواری سؤال ها) ارزیابی شد. جهت تولید و تحلیل داده‌ها از نرم‌افزارهای آماری R بسته‌های mirt، interactions، car و psych استفاده شد.

1- Bulut, & SÜNBÜL
2- Risk

یافته‌ها

ابتدا میانگین و انحراف استاندارد متغیرهای وابسته پژوهش (میانگین توان دوم خطاهای شاخص‌های شیب و دشواری) به تفکیک متغیرهای مستقل در جدول ۲ ارائه شده است.

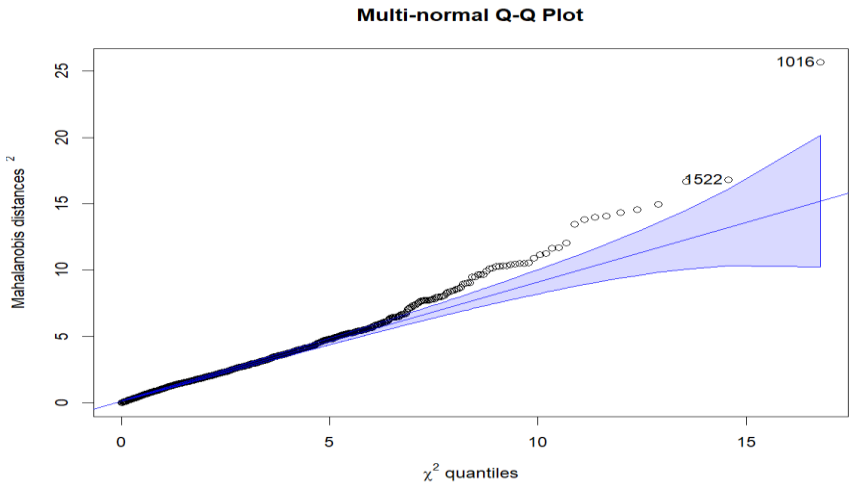
جدول۲: اطلاعات توصیفی میانگین توان دوم خطاها (MSE)

Table 2

Descriptive information of mean square error (MSE)

دشواری (b) Difficulty		شیب (a) slope		تعداد number	طول آزمون Test length	تعداد بعد Number of dimensions	الگوریتم Algorithm
انحراف معیار standard deviation	میانگین mean	انحراف معیار standard deviation	میانگین mean				
0.008	0.058	0.013	0.042	100	14	1	EM
0.007	0.047	0.008	0.044	100	45		
0.009	0.041	0.004	0.023	100	120		
0.011	0.049	0.013	0.037	300	total		
0.017	0.070	0.017	0.037	100	14	3	
0.016	0.061	0.020	0.041	100	45		
0.025	0.052	0.008	0.013	100	120		
0.021	0.061	0.020	0.030	300	total		
0.009	0.060	0.462	0.906	100	14	5	
0.007	0.052	0.492	0.786	100	45		
0.007	0.041	0.527	0.828	100	120		
0.011	0.051	0.495	0.840	300	total		
0.011	0.090	0.012	0.097	100	14	1	MCEM
0.006	0.079	0.008	0.095	100	45		
0.004	0.069	0.006	0.080	100	120		
0.011	0.079	0.012	0.091	300	total		
0.013	0.091	0.013	0.082	100	14	3	
0.015	0.091	0.009	0.087	100	45		
0.014	0.076	0.012	0.070	100	120		
0.016	0.086	0.014	0.080	300	total		
0.035	0.124	0.053	0.130	100	14	5	
0.018	0.093	0.083	0.171	100	45		
0.090	0.166	0.064	0.134	100	120		
0.064	0.128	0.070	0.145	300	total		
0.005	0.032	0.011	0.046	100	14	1	MHRM
0.006	0.031	0.006	0.045	100	45		
0.006	0.018	0.003	0.025	100	120		
0.008	0.027	0.012	0.039	300	total		
0.017	0.052	0.051	0.080	100	14	3	
0.015	0.042	0.029	0.056	100	45		
0.018	0.029	0.016	0.026	100	120		
0.019	0.041	0.041	0.054	300	total		
0.021	0.054	0.036	0.061	100	14	5	
0.042	0.077	0.016	0.040	100	45		
0.014	0.034	0.054	0.076	100	120		
0.033	0.055	0.041	0.059	300	total		

در ادامه جهت پاسخگویی به سؤال پژوهش مبنی بر اینکه آیا الگوریتم MHRM در قیاس با الگوریتم‌های مرسوم دیگر یعنی الگوریتم‌های EM و MCEM با در نظر گرفتن تعداد ابعاد مختلف و تعداد متفاوت سؤالات عملکرد بهتری دارد؟ از روش تحلیل واریانس چندمتغیره عاملی[1] استفاده شد. در این نوع تحلیل باید مفروضه‌هایی رعایت گردند تا بتوان به نتایج بدست آمده اطمینان کرد. مفروضه اول وجود ساختار همبستگی بین متغیرهای وابسته (مخاطره پارامترهای شیب و دشواری) بود که نتایج آزمون کرویت بارتلت نشان داد بین متغیرهای وابسته پژوهش همبستگی لازم جهت تحلیل‌های چندمتغیری وجود دارد (P=۰,۰۱ و Chi-Square=۶۸۶۰,۳۵). یکی دیگر از این مفروضه‌ها، بررسی همسانی ماتریسهای واریانس–کوواریانس می‌باشد که بدین منظور از آزمون باکس[2] استفاده شده است. میزان معناداری آزمون باکس از ۰,۰۵ کوچکتر است (P=۰,۰۱ و F=۱۸۶,۷۵ و Box`sM=۱۴۶۷,۷۶)، لذا نتیجه گرفته می‌شود که ماتریس واریانس–کوواریانس‌ها همگن نیست و لذا برقراری این مفروضه نقض شده است. در ادامه جهت بررسی مفروضه نرمال‌بودن داده‌ها از آزمون نرمال بودن چندمتغیره شاپیرو ویلک[3] و روش نموداری Q-Q استفاده شد. مقدار بدست‌آمده (P=۰,۰۱ و Mvw=۰,۸۱) در آزمون شاپیروویلک نشان از عدم برقراری پیش فرض نرمال بودن داده‌ها دارد. نتایج روش نموداری Q-Q که علاوه بر مفروضه نرمال بودن چندمتغیری، عدم وجود داده‌های پرت چندمتغیری را نیز بررسی می‌کند در شکل ۱ قابل مشاهده است.



شکل ۱: شاخص فاصله ماهالانوبیس جهت بررسی نرمال بودن و عدم وجود داده‌های پرت

Figure 1

Mahalanobis distance index to check the normality and absence of outlier data

1- Factorial Multivariate Analysis of variance
2- Box's Test of Equality of Covariance Matrices
3- Shapiro-Wilk test for Multivariate Normality

همانگونه که نتایج شکل ۱ نشان می‌دهد مشاهده می‌شود بدلیل وجود داده‌های پرت چندمتغیری، حالت نرمال بودن برقرار نیست. با توجه به رد برخی از مفروضه‌ها و البته بدلیل وجود تعداد داده‌های یکسان در هر یک از سطوح طرح عاملی، در بررسی نتایج تحلیل کوواریانس چندمتغیری از آزمون مقاوم اثر پیلایی[1] که نسبت به نقص مفروضه‌ها پایدارتر است، استفاده شد. در جدول ۳ آماره چندمتغیری برای هر یک از متغیرهای مستقل و تعامل بین آنها ارائه شده است.

جدول ۳: نتایج تحلیل کوواریانس چندمتغیری

Table 3

Results of multivariate covariance analysis

متغیر ها Variables	اثر پیلایی Pillais Trace	F Fisher's statistics	درجه آزدی۱ (۲) Degrees of freedom 1(2)	مقدار احتمال P-Value	مجذور اتا Squared ita
الگوریتم Algorithm	0.81	911.73	4 (4346)	0.001	0.40
بعد آزمون test dimension	0.48	421.58	4 (4346)	0.001	0.24
طول آزمون test length	0.04	27.16	4 (4346)	0.001	0.02
الگوریتم*بعد آزمون Algorithm*test dimension	0.59	282.91	4 (4346)	0.001	0.29
الگوریتم*طول آزمون Algorithm*test length	0.08	27.34	4 (4346)	0.001	0.04
بعد آزمون * طول آزمون test dimension*test length	0.03	10.66	4 (4346)	0.001	0.02
الگوریتم*بعد آزمون * طول آزمون Algorithm*test dimension*test length	0.15	27.63	4 (4346)	0.001	0.08

همانگونه که در جدول ۳ مشاهده می شود، آماره چندمتغیری مربوطه یعنی اثر پیلایی در سطح خطای ۰/۰۱۶ (براساس آلفای بونفرونی) برای هر سه متغیر مستقل و حالت های مختلف تعامل بین آنها معنی-دار است (P=۰,۰۰۱). بدین ترتیب ترکیب خطی متغیرهای وابسته (شاخص MSE شیب و دشواری) حداقل یکی از متغیرهای مستقل (نوع الگوریتم، تعداد ابعاد و طول آزمون) تأثیر پذیرفته است. با توجه به اینکه آزمون چندمتغیری مذکور معنادار بوده و ترکیب متغیرهای وابسته از حداقل یکی از متغیرهای مستقل اثر پذیرفته است، لذا بعد از آن به پیگیری وضعیت اثرگذاری هر یک از متغیرهای مستقل بر هر یک از متغیرهای وابسته و نیز تعامل بین آنها به طور مجزا پرداخته شده است. جهت بررسی این موضوع از آزمون تجزیه‌وتحلیل واریانس عاملی تک متغیره استفاده شد که نتایج آن در جدول ۴ ارائه شده است.

1- Pillais Trace

جدول ۴: نتایج تجزیه و تحلیل واریانس تک‌متغیره جهت مقایسه میانگین توان دوم خطاها

Table 4

The results of univariate analysis of variance to compare the mean of squared errors

متغیر وابسته dependent variable	متغیر مستقل independent variable	درجه آزادی Degrees of freedom	میانگین مجذورات mean square	F Fisher's statistics	مقدار احتمال P-Value	مجذور اتا Squared ita
شیب slope	الگوریتم Algorithm	2	15.79	562.90	0.001	0.30
	بعد آزمون test dimension	2	25.75	917.69	0.001	0.41
	طول آزمون test length	2	0.11	4.18	0.01	0.01
	الگوریتم*بعد آزمون Algorithm*test dimension	4	19.88	708.58	0.001	0.51
	الگوریتم*طول آزمون Algorithm* test length	4	0.06	1.97	0.10	0.01
	بعد آزمون * طول آزمون test dimension*test length	4	0.04	1.43	0.22	0.01
	الگوریتم*بعد آزمون*طول آزمون Algorithm*test dimension*test length	8	0.07	2.56	0.01	0.01
	خطا error	2673	0.03	–	–	–
دشواری difficulty	الگوریتم Algorithm	2	0.80	1415.22	0.001	0.51
	بعد آزمون test dimension	2	0.16	276.44	0.001	0.17
	طول آزمون test length	2	0.03	55.10	0.001	0.04
	الگوریتم*بعد آزمون Algorithm*test dimension	4	0.06	108.38	0.001	0.14
	الگوریتم*طول آزمون Algorithm* test length	4	0.03	54.09	0.001	0.07
	بعد آزمون * طول آزمون test dimension*test length	4	0.01	20.46	0.001	0.03
	الگوریتم*بعد آزمون*طول آزمون Algorithm*test dimension*test length	8	0.03	55.88	0.001	0.14
	خطا error	2673	0.001	–	–	–

با توجه به نتایج جدول ۳، F مشاهده شده در سطح خطای ۰/۰۵ تفاوت معناداری را بین میانگین گروه‌های مورد براساس متغیرهای مستقل (نوع الگوریتم، تعداد ابعاد و طول آزمون)، در هر دو متغیر

وابسته شیب و دشواری نشان داد. این نتایج نشان می دهد تفاوت در سطوح هر یک از متغیرهای مستقل، در هر یک از متغیرهای وابسته، میزان میانگین توان دوم خطای متفاوتی را به دنبال دارد. در ادامه با استفاده از آزمون تعقیبی بونفرونی به بررسی تفاوت نمرات متغیرهای وابسته در بین گروه های هر یک از متغیرهای مستقل به صورت زوجی پرداخته شد که نتایج در جدول ۵ ارائه شده است.

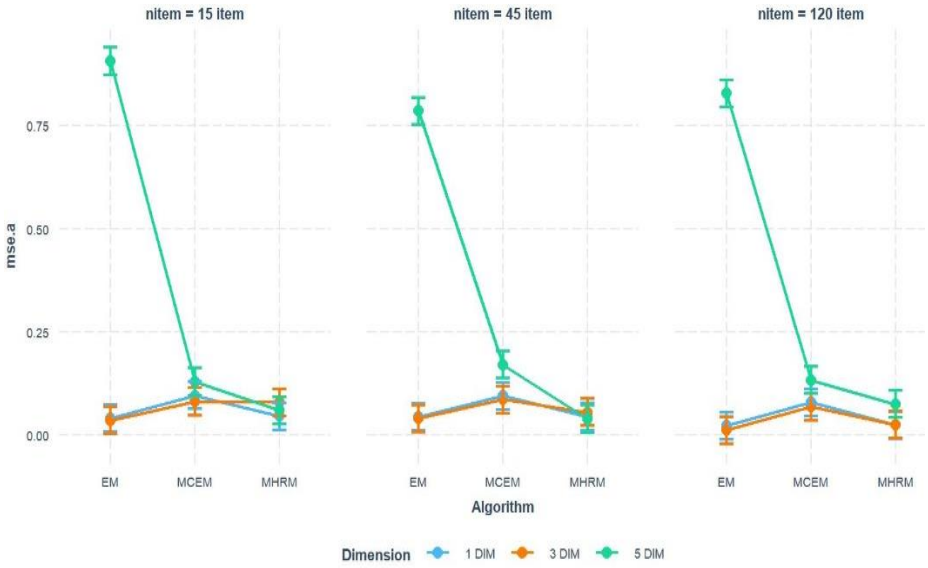
جدول ۵: نتایج آزمون بونفرونی برای مقایسه ی میانگین گروه های متغیرهای مستقل به صورت زوجی

Table 5

The results of Bonferroni test for comparing the mean of groups of independent variables in pairs

متغیر وابسته dependent variable	متغیر مستقل independent variable	گروه group	تفاوت میانگین mean difference	خطای برآورد estimation error	مقداراحتمال P-Value	حدپایین lower	حدبالا upper
شیب slope	الگوریتم Algorithm	EM & MCEM	-0.20	0.01	0.001	0.18	0.22
		EM & MHRM	0.35	0.01	0.001	0.23	0.27
		MCEM & MHRM	0.06	0.01	0.001	0.04	0.07
	تعداد ابعاد dimension	1 & 3	0.001	0.01	0.98	-0.02	0.02
		1 & 5	-0.29	0.01	0.001	-0.31	-0.27
		3 & 5	-0.29	0.01	0.001	-0.31	-0.27
	طول آزمون test length	15 & 45	0.01	0.01	0.31	-0.01	0.03
		15 & 20	0.02	0.01	0.01	0.001	0.04
		45 & 120	0.01	0.01	0.62	-0.01	0.03
دشواری difficulty	الگوریتم Algorithm	EM & MCEM	-0.04	0.001	0.001	-0.05	-0.04
		EM & MHRM	0.01	0.001	0.001	0.01	0.02
		MCEM & MHRM	0.06	0.001	0.001	0.05	0.06
	تعداد ابعاد dimension	1 & 3	-0.01	0.001	0.001	-0.01	-0.01
		1 & 5	-0.03	0.001	0.001	-0.03	-0.02
		3 & 5	-0.02	0.001	0.001	-0.02	-0.01
	طول آزمون test length	15 & 45	0.01	0.001	0.001	0.001	0.01
		15 & 20	0.01	0.001	0.001	0.01	0.01
		45 & 120	0.01	0.001	0.001	0.001	0.01

نتایج جدول ۶ نشان می‌دهد که در مقایسه زوجی سطوح مختلف متغیرهای مستقل (نوع الگوریتم، تعداد ابعاد و طول آزمون) در هر یک از متغیرهای وابسته (شیب و دشواری) در تمامی مقایسه ها بجز دو مورد در متغیر شیب (بین بعد ۱۵ و ۴۵ و نیز طول آزمون ۴۵ و ۱۲۰) در سطح اطمینان ۹۵ درصد تفاوت معنی داری وجود دارد. در ادامه نتایج پژوهش به صورت نمودار در شکل های ۲ و ۳ به تفکیک متغیرهای وابسته پژوهش ارائه شده است.



شکل۲: نمودار مقایسه میانگین توان دوم خطاها (پارامتر شیب)

Figure 2

Comparison diagram of the mean square of errors (slope parameter)

همانگونه که شکل ۲ نشان می دهد که زمانی که تعداد ابعاد بالا در هر سه سطح تعداد سؤالات الگوریتم EM خطای بسیار بالایی را به دنبال دارد و میزان کارایی و سوگیری برآورد به شدت افزایش می یابد ولیکن زمانی که تعداد ابعاد پایین باشد (۱ و ۳ بعد)، این الگوریتم در کنار دو الگوریتم دیگر یعنی MCEM و MHRM عملکرد بسیار مطلوبی دارد و تفاوت قابل توجهی بین آنها وجود ندارد. الگوریتم MCEM نیز در کنار الگوریتم MHRM عملکرد مطلوبی دارد و بیشترین تفاوت بین دو الگوریتم زمانی است که تعداد ابعاد بالا (۵ بعد) می باشد. در کل نتیجه گیری می شود الگوریتم MHRM زمانی که تعداد ابعاد بالاست نسبت به دو الگوریتم دیگر دارای مخاطره کمتری در پارامتر شیب است و برآوردهای دقیقتر و باارزش‌تری ارائه می‌کند.

شکل۳: نمودار مقایسه میانگین توان دوم خطاها (پارامتر دشواری)
Figure 3
Comparison diagram of the mean square of errors (difficulty parameter)

مقایسه مخاطره پارامتر دشواری در شرایط مختلف متغیر مستقل و تعامل بین آنها در نمودار شکل ۳ نشان می‌دهد، به طور کلی الگوریتم MCEM رد مقایسه با دو الگوریتم دیگر یعنی EM و MHRM دارای مخاطره بیشتری است و این امر نشان می‌دهد پارامترهای بدست آمده دارای سوگیری و خطای نمونه‌گیری بیشتری هستند. نکته قابل توجهی که می توان بدان اشاره کرد این است که الگوریتم EM و MHRM زمانی که تعداد سؤالات بیشتر می‌شود و دارای خطای کمتری هستند ولیکن الگوریتم MCEM نتایج ناپایداری را نشان می‌دهد.

بحث و نتیجه‌گیری

هدف پژوهش حاضر بررسی مخاطره الگوریتم متروپلیس هستینگز روبینز مونرو در داده‌های چند بعدی نظریه سؤال پاسخ بود. همچنین در این راستا یکی دیگر از چالش های موجود در داده‌های علوم رفتاری یعنی طول آزمون نیز مورد ارزیابی قرار گرفت. نتایج نشان داد هر سه متغیر نوع الگوریتم، تعداد ابعاد و طول آزمون در میزان میانگین توان دوم خطای برآورد پارامترها اثر گذار هستند.

در بررسی مقایسه الگوریتم MHRM با الگوریتم های EM و MCEM نتایج بدست آمده نشان از برتری ویژه الگوریتم MHRM داشت. بدین معنی که این الگوریتم در مقایسه با دو الگوریتم EM و

MCEM دارای مخاطره کمتری می باشد. این موضوع بویژه زمانی اهمیت پیدا می کند و قابل توجه می شود که ساختار داده‌ها پیچیده می شود. یکی از پیچیدگی هایی که در این پژوهش مورد بررسی قرار گرفت تعداد ابعاد بود. نتایج بررسی نشان داد الگوریتم MHRM در قیاس با الگوریتم های دیگر زمانی که تعداد ابعاد داده‌ها افزایش پیدا می کند، تفوق بیشتری دارد و پارامترهایی که برآورد شدند، دارای میانگین توان دوم خطای کمتری است. نکته مهم دیگر در رابطه با چالش دیگر یعنی طول آزمون بود. در این زمینه نتایج بدست آمده نشان داد زمانی که طول آزمون کاهش می یابد و تعداد سؤالات برای هر عامل کم می شود، الگوریتم های MCEM و بویژه EM با مخاطره بالایی همراه است. هر سه پارامتری که در این پژوهش مورد بررسی قرار گرفتند، این تفاوت ها را نشان دادند. در بررسی پارامتر شیب نتایج نشان از برتری الگوریتم MHRM بر الگوریتم های دیگر در تعداد ابعاد بالا و طول آزمون کوتاه با میزان میانگین توان دوم خطای کمتر داشت. این شرایط در رابطه با پارامتر دشواری نیر صادق بود.

همچنین در اثرات تعاملی بین متغیرها در پارامتر شیب نتایج نشان داد زمانی که تعداد ابعاد بالا در هر سه سطح تعداد سؤالات الگوریتم EM مخاطره بالاتری دارند و براین اساس میزان واریانس نمونه گیری و سوگیری برآورد به شدت افزایش می یابد. از سویی دیگر هنگامی که تعداد ابعاد کاهش می یابد بویژه زمانی که مدل تک عاملی می‌شود، این الگوریتم در کنار دو الگوریتم دیگر یعنی MCEM و MHRM عملکرد بسیار مطلوبی دارد و تفاوت قابل توجهی بین آنها وجود ندارد. لازم به ذکر است در ابعاد پایین، زمان برآورد پارامترها از دو حالت دیگر کمتر می‌شود. در تبیین عملکرد الگوریتم MCEM نیز می‌توان بیان کرد که در کل عملکرد مطلوبی دارد و فقط زمانی که تعداد عوامل به ۵ بعد می رسد، هم زمان بسیار بیشتری برای برآورد نیاز دارد و هم مخاطره برآورد پارامتر شیب افزایش معنی داری پیدا می کند. در نتیجه می‌توان بیان داشت که الگوریتم MHRM زمانی که تعداد ابعاد بالاست نسبت به دو الگوریتم دیگر دارای مخاطره کمتری در پارامتر شیب است و برآوردهای دقیقتر و باارزش‌تری ارائه می‌کند که دارای واریانس نمونه‌گیری کمتر و نیز سوگیری کمتری است.

همچنین در بررسی مخاطره پارامتر دشواری در شرایط مختلف متغیر مستقل و تعامل بین آنها می‌توان ابراز داشت که به طور کلی الگوریتم MCEM در مقایسه با دو الگوریتم دیگر یعنی EM و MHRM دارای مخاطره بیشتری است. این امر نشان می‌دهد پارامترهای بدست آمده زمانی که از الگوریتم MCEM استفاده می‌شود، همراه با خطای بیشتری هستند و برآوردهای سوگیرانه‌تری ارائه می‌دهد. جنبه مهمی که در این بخش می‌توان به آن اشاره کرد اثر تعاملی بین الگوریتم و تعداد سؤالات است. بررسی نتایج نشان می‌دهد که در تعامل بین آنها، هنگامی که تعداد سؤالات افزایش می‌یابد، الگوریتم MCEM نتایج ناپایداری را نشان می‌دهد.

نتایج بدست آمده در پژوهش حاضر با یافته‌های پژوهش‌های کای (۲۰۱۰)، یانگ و کای (۲۰۱۴)، کو و شنگ (۲۰۱۶) و دمارس (۲۰۱۷) همسو بود. کای (۲۰۱۰) عنوان می‌کند الگوریتم MHRM نسبت به الگوریتم‌های مرسوم دیگر از قبیل EM و MCEM دقت بالاتری در برآورد پارامترها در مدل های چند بعدی نظریه سؤال پاسخ دارد. یانگ و کای (۲۰۱۴) نیز بیان می‌دارند الگوریتم MHRM می تواند برآوردها و خطاهای استاندارد را به طور کارآمدی تولید کند. کو و شنگ (۲۰۱۶) نشان می‌دهند، الگوریتم MHRM می‌تواند منجر به بهبود دقت در برآورد پارامترها شود. این نتایج نشان می‌دهد همچنان که در مسأله نفرین ابعاد بیان شد، الگوریتم EM با کاربرد نقاط کوادراتور گاوسی–هرمیت ثابت بار محاسباتی رو به شدت افزایش می‌دهد و حتی تعیین نقاط بهینه با استفاده از نقاط کوادراتور تطبیقی نیز مشکل همچنان باقی می‌ماند. الگوریتم MCEM نیز هر چند با در اختیار قرار دادن توزیع پیشین آگاهی بخش در مرحله E با نمونه‌برداری از نقاط کوادراتور با استفاده از شبیه‌سازی مونت کارلو، به جای استفاده از ربع گاوسی–هرمیتی در برآورد پاراکتر شیب عملکرد EM را بهینه می کند ولیکن در برآورد پارامتر دشواری همراه با خطای بالاتری هست. این مر نشان می‌دهد که احتمالا جهت رسیدن به همگرایی نقطه‌ای برآوردهای پارامتر دشواری، اندازه شبیه‌سازی و نمونه برداری به شدت در تکرارهای آخر افزایش می یابد. نکته دیگیر که می توان به آن اشاره کرد، این است که الگوریتم MCEM برای هر تکرار در مرحله E، نمونبرداری جدیدی از نمونه های تصادفی انجام می دهد، که این امر می‌تواند در طولانی شدن محاسبات و نرسیدن به همگرایی دقیق اثرگذار باشد.

و اما در باره برتری الگوریتم MHRM بر دو الگوریتم دیگر می‌توان به چالمرز(۲۰۱۲) و باشکوف و دمارس (۲۰۱۷) اشاره کرد که بیان می‌دارند استفاده از الگوریتم MHRM باعث می‌شود فرآیند برآورد برای مدل‌های چندبعدی عملکرد مطلوبی داشته باشد. به نظر می‌رسد الگوریتم MHRM با استفاده از ترکیب نمونه‌گیری متروپولیس– هستینگز با نمونه‌گیری روبینز–مونرو بیشینه درست‌نمایی را به خوبی تسهیل می‌کند. از آنجایی که الگوریتم MHRM، بر تخمین نقطه و خطاهای استاندارد تمرکز دارد و همین امر اجازه می دهد الگوریتم MHRM خیلی سریعتر از MCEM و EM به همگرایی برسد. در نتیجه می‌توان گفت از آنجایی که در MHRM، جهش‌ها در زنجیره مارکوف به منظور تعیین جهت تغییر در طول تکرارها است، همگرایی سریعتر و دقیقر اتفاق می‌افتد. در پایان با توجه به یافته‌های پژوهش می‌توان نتیجه‌گیری کرد که الگوریتم MHRM با در نظر گرفتن عوامل مداخله‌گری چون تعداد ابعاد و طول آزمون دارای مخاطره کمتری است و لذا پیشنهاد می شود در تحلیل داده‌های پیچیده که در علوم رفتاری به فراوانی یافت می‌شود، از الگوریتم MHRM استفاده گردد تا برآوردها با خطای کمتری همراه باشند. محدودیت‌هایی در این پژوهش وجود داشت، غالبا بدلیل پیچیدگی مطالعه و گستردگی کار و اصل

امکان پذیری در مطالعات شبیه‌سازی بود. در هر مطالعات شبیه‌سازی بدلیل گستردگی مطالعه در طرح-های متقاطع و ایجاد شرایط قابل مقایسه، امکان در نظر گرفتن و ایجاد شرایط محدودی وجود دارد. در این پژوهش از سه متغیر مستقل استفاده شد که هر یک سه سطح داشتند و امکان در نظر گرفتن سطوح بیشتر در عمل امکان‌پذیر نبود. برای مثال در بررسی الگوریتم‌ها با اضافه شدن فقط یک سطح، شرایط آزمایشی به شدت افزایش می‌یافت و امکان تولید داده‌ها را دشوار می‌نمود. به همین دلیل الگوریتم‌های دیگری از قبیل MCMC و QMCMC می‌توانستند در پژوهش مورد بررسی قرار گیرند که در عمل امکان پذیر نبود. همچنین مولفه‌های مهم دیگری از قبیل چالش حجم داده‌ها و نوع الگوی پاسخ متفاوت نیز می‌توانستند در پژوهش باشند که میسر نبود. لذا در این زمینه به پژوهشگران دیگر پیشنهاد می‌شود، به بررسی چالش‌هایی از قبیل حجم داده‌ها، نوع الگوی پاسخ (دو ارزشی و چند ارزشی و پاسخ پیوسته)، الگوریتم‌های کاربردی دیگر (MCMC و QMCMC)، مدل‌های سؤال پاسخ پرکاربرد دیگر (مانند مدل پاسخ اسمی، مدل های دوارزشی با بیش از دو پارامتر مثل مدل ۴ پارامتری) و ... مورد ارزیابی قرار بگیرند.

منابع References

دلاور، علی (۱۴۰۱). *کتاب احتمالات و آمار کاربردی در روان شناسی و علوم تربیتی*. تهران: نشر رشد.

Asparouhov, T., & Muthén, B. (2012). *General random effect latent variable modeling: Random subjects, items, contexts, and parameters*. In annual meeting of the National Council on Measurement in Education, Vancouver, British Columbia.

Bartolucci, F., Bacci, S., & Gnaldi, M. (2015). *Statistical analysis of questionnaires*: A unified approach based on R and Stata (Vol. 34). CRC Press.

Bashkov, B. M., & DeMars, C. E. (2017). Examining the performance of the Metropolis–Hastings Robbins–Monro algorithm in the estimation of multilevel multidimensional IRT models. *Applied psychological measurement*, 41(5): 323-337.

Bulut, O., & SÜNBÜL, Ö. (2017). Monte Carlo Simulation Studies in Item Response Theory with the R Programming Language R Programlama Dili ile Madde Tepki Kuramında Monte Carlo Simülasyon Çalışmaları. *Journal of Measurement and Evaluation in Education and Psychology*, 8(3): 266-287.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. Psychometrika, 75(1): 33-57.

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahway.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 1-22.

Delavar, A. (2021). *Applied probability and statistics in psychology and educational sciences*. Tehran: Roshd Published [In Persian].

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Gibbons, R. D., Weiss, D. J., Frank, E., & Kupfer, D. (2016). Computerized adaptive diagnosis and testing of mental health disorders. *Annual Review of Clinical Psychology*, 12, 83-104.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1985). *Principles and applications of item response theory*.

Han, K. T., & Hambleton, R. K. (2014). *User's Manual for WinGen 3: Windows Software that Generates IRT Model Parameters and Item Responses (Center for Educational Assessment Report No. 642)*. Amherst, MA: University of Massachusetts.

Liu, Q., & Pierce, D. A. (1994). A note on Gauss—Hermite quadrature. *Biometrika*, 81(3), 624-629.

Kuo, F. Y., & Sloan, I. H. (2005). Lifting the curse of dimensionality. Notices of the AMS, 52(11), 1320-1328.

Kasim, M. F., Bott, A. F. A., Tzeferacos, P., Lamb, D. Q., Gregori, G., & Vinko, S. M. (2019). Retrieving fields from proton radiography without source profiles. *Physical Review E*, 100(3): 033208.

Kuo, T. C., & Sheng, Y. (2016). A comparison of estimation methods for a multi-unidimensional graded response IRT model. *Frontiers in psychology*, 7, 880.), 1-29.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3): 325-335.

Linden, W. J., & van der, & Hambleton, RK (1997*). Handbook of modern item response theory*, 9-39.

Meng, X. L., & Schilling, S. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, 91(435): 1254-1267.

Naylor, J. C., & Smith, A. F. (1982). Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 31(3): 214-225.

Patz, R. J., & Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of educational and behavioral statistics*, 24(4): 342-366.

Robbins, H., & Monro, S. (1951). *A stochastic approximation method. The annals of mathematical statistics*, 400-407.

Yang, J. S., & Cai, L. (2014). Estimation of contextual effects through nonlinear multilevel latent variable modeling with a Metropolis–Hastings Robbins–Monro algorithm. *Journal of Educational and Behavioral Statistics*, 39(6): 550-582.

Sahin, A., & Anil, D. (2017). *The effects of test length and sample size on item parameters in item response theory*.

Patsula, L. (1995). *A comparison of item parameter estimates and ICCs produced with TESTGRAF and BILOG under different test lengths and sample sizes*. University of Ottawa (Canada).