Journal of Statistical Sciences, Spring and Summer, 2022
Vol. 16, No. 1, pp 127-148
DOI: 10.29252/jss.16.1.127

Modified Two-Stage Sampling Around the Mean of the First Order Autoregressive Model

Sajjadipanah, S., Mahmoudi, E., Zamani, M.[3]
Department of Statistics, Yazd University, Yazd , Iran.

Corresponding author: E. Mahmoudi, emahmoudi@yazd.ac.ir

Received: 19 October 2021 **Revised:** 4 February 2022 **Accepted and Published Online:** 12 February 2022.

Introduction

Due to its simplicity, the autoregressive process is widely used in practice. In some cases, we face restrictions in estimating the parameter for predicting the expected value of the time series models and cannot use an optimal fixed sample size procedure. One appropriate approach to tackle this problem is to use sequential methods. Sequential procedures are distinguished by the definition of the stopping rule and are stopped by a pre-defined stopping rule as soon as asymptotic properties of the procedure are observed. Many authors have researched sequential sampling procedures, namely purely sequential and two-stage procedures.

We are also interested in investigating the performance of the modified two-stage procedure because of the most important and widely used sequential methods and the operational savings of these procedures. The modified two-stage approach proposes a situation where we can provide a strategy for determining the initial sample size in the two-stage process that, in many cases, prevents overestimation of the final sample size. The advantages of the modified two-stage include the simplicity of implementation and reduced weakness of the two-stage procedure in estimating. The point estimation is studied based on the least-squares estimator as the reciprocal of the cost per observation tends to infinity and interval estimation when the width of the confidence interval goes to zero. We presented the performance of the procedure under the theorems that demonstrate the asymptotic properties of the procedures, including asymptotic risk efficiency, asymptotic efficiency,

and asymptotic consistency.

Material and Methods

We conduct Monte Carlo simulation studies to investigate the performance of procedures based on least-squares estimators. The performance of estimators and confidence intervals are evaluated utilizing a simulation study. We report the results in terms of the stopping variables, the ratio of the average stopping variable to the optimal fixed sample size, the root of mean square error (RMSE) of the estimators, and the proportion of risk efficiency functions. Furthermore, real-time series data is considered to illustrate the applicability of the modified two-stage procedure.

Results and Discussion

The simulation results confirm the theoretical results and show the procedure's effectiveness compared to the optimal fixed sample size. Also, the actual data results show the excellent performance of this procedure in practice.

Conclusion

The modified two-stage procedure and operational savings are more accessible to implement than the most commonly used purely sequential procedure. It is also more accurate than the two-stage procedure. Also, the procedure performs well and is an excellent candidate for analyzing time series models.

Keywords: Modified two-stage procedure, Autoregressive model, least-squares estimator, Monte Carlo simulation.

Mathematics Subject Classification (2010): 62L12, 62M10.

مجله علوم آماری، بهار و تابستان ۱۴۰۱
جلد ۱۶، شماره ۱، ص ۱۲۷ – ۱۴۸
DOI: 10.29252/jss.16.1.127
مقاله پژوهشی

نمونه‌گیری دو مرحله‌ای بهبودیافته پیرامون میانگین مدل خودبازگشتی مرتبه اول

سودابه سجادی‌پناه، عیسی محمودی و محمدصادق زمانی
گروه آمار، دانشگاه یزد

نویسنده مسئول : عیسی محمودی، emahmoudi@yazd.ac.ir

تاریخ دریافت: ۱۴۰۰/۰۷/۲۷ تاریخ بازنگری: ۱۴۰۰/۱۱/۱۵ تاریخ پذیرش و انتشار: ۱۴۰۰/۱۱/۲۱

چکیده: در این مقاله، روش نمونه‌گیری دومرحله‌ای بهبودیافته پیرامون میانگین مدل خودبازگشتی مرتبه اول مطالعه شده است. برآورد نقطه‌ای و فاصله‌ای میانگین مدل بر اساس برآوردگرهای کمترین توان‌های دوم با شرط مینیمم‌سازی تابع مخاطره بررسی شده است. توزیع مجانبی برآوردگر میانگین نیز بر اساس قاعده توقف نقطه‌ای ارائه شده است. هم‌چنین مطالعه شبیه‌سازی مونت کارلویی برای بررسی کارایی روش پیشنهادی نسبت به روش اندازه نمونه ثابت بهینه بر اساس متغیر توقف، نسبت متغیر به اندازه نمونه ثابت بهینه، برآورد میانگین، ریشه دوم میانگین توان‌های دوم خطا، نسبت تابع مخاطره حاصل از روش ارائه شده به مخاطرهٔ اندازه نمونه ثابت بهینه و احتمال پوشش بازهٔ اطمینان طراحی و اجرا شده است. در انتها، با به‌کارگیری داده واقعی کاربرد روش ارائه شده مورد بررسی قرار گرفته است.

واژه‌های کلیدی: نمونه‌گیری دومرحله‌ای بهبودیافته، مدل خودبازگشتی، برآوردگر کمترین توان‌های دوم، شبیه‌سازی مونت کارلویی.
کد موضوع‌بندی ریاضی (۲۰۱۰): 62M10، 62L12.

۱ مقدمه

کاربردهای متعدد داده‌های سری زمانی در زمینه‌های گوناگون محققان را ترغیب به مطالعه در این زمینه کرده است. اما گاهی اندازه نمونه از پیش مشخص نیست و نتایج آزمایش تصادفی تعیین‌کننده اندازه نمونه هستند. یکی از اهداف در تحلیل مدل‌های سری زمانی، به‌کارگیری مدل برازش‌شده در پیش‌بینی مقادیر مورد انتظار فرایند است که با

محدودیت مجهول بودن اندازه نمونه نمی‌توان به استنباط در مورد مدل پرداخت. یکی از راهکارهای ارائه‌شده در این شرایط، استفاده از روش‌های نمونه‌گیری دنباله‌ای است. ایده استفاده از روش‌های دنباله‌ای جهت تعیین اندازه دقیق نمونه با استفاده از قاعده توقف مناسب به منظور برآورد نقطه‌ای و فاصله‌ای به صورت گسترده از مدت‌ها مورد توجه بوده است. در سال ۱۹۴۰ ماهالانوبیس طی فعالیت‌های علمی خود برای برآورد میزان محصول کنف در مزارع بنگال به اهمیت نمونه‌گیری مرحله‌ای پی‌برد و در مطالعات میدانی خود به توسیع طرح‌های نمونه‌ای برای برآورد وسعت مزارع کنف پرداخت. این توسیع بنیادین در سایر پژوهش‌های نمونه‌ای بزرگ–مقیاس توسط برخی از پژوهشگران از جمله آبراهام والد نیز مورد توجه قرار گرفت. آبراهام والد به عنوان پیشتاز تحلیل‌های دنباله‌ای در سال ۱۹۴۰ با همکاری دستیاران خود به توسعه نظری و روش‌شناسی آزمون‌های دنباله‌ای پرداخت. سپس، انسکامب (۱۹۵۳)، رابینز (۱۹۵۹) و چائو و رابینز (۱۹۶۵) قاعده توقف دنباله‌ای را برای برآورد نقطه‌ای و فاصله‌ای میانگین جامعه ارائه دادند.

مسئله اصلی در تحلیل دنباله‌ای، تعیین بهترین تصمیم یا قاعده برای توقف نمونه‌گیری و تعیین اندازه نمونه است. نمونه‌گیری دنباله‌ای بر اساس تعریف قاعده توقف مشتمل بر چندین روش است که از پرکاربردترین آن‌ها می‌توان به روش دنباله‌ای محض، روش دومرحله‌ای و دومرحله‌ای بهبودیافته اشاره کرد. محققان بسیاری در زمینه مدل‌های سری زمانی با رویکرد دنباله‌ای مطالعه و پژوهش انجام داده‌اند. روش دومرحله‌ای اولین بار توسط اشتاین (۱۹۴۵، ۱۹۴۹) برای بررسی برآورد فاصله‌ای پارامترها پیشنهاد شد که این روش با نام خود او نیز شناخته می‌شود. ماخوپادهای (۱۹۸۰) برآورد فاصله‌ای میانگین بر اساس روش دومرحله‌ای را در توزیع نرمال با واریانس مجهول مطالعه کردند. سیرام (۱۹۸۷، ۱۹۸۸) روش نمونه‌گیری دنباله‌ای محض را برای برآورد پارامتر و میانگین مدل خودبازگشتی مرتبه اول مورد بررسی قرار داد. باساوا و همکاران (۱۹۹۰) روش دنباله‌ای محض در داده‌های وابسته را بررسی کردند. فخری–ذاکری و لی (۱۹۹۲) روش نمونه‌گیری دنباله‌ای محض را برای برآورد پارامتر میانگین فرایند خطی بررسی کردند. آن‌ها در این مقاله، ویژگی‌های کارایی مجانبی، کارایی مجانبی ریسک و توزیع مجانبی برآوردگرها را مورد مطالعه قرار دادند. فخری–ذاکری و لی (۱۹۹۳) روش نمونه‌گیری دنباله‌ای محض را برای برآورد بردار میانگین در فرایند خطی چندمتغیره بررسی نمودند که تعمیمی بر مطالعه این نویسندگان در سال قبل بود. لی (۱۹۹۴) در بررسی روش نمونه‌گیری دنباله‌ای محض در مدل خودبازگشتی مرتبه اول شروط مطرح‌شده در مقاله سیرام (۱۹۸۸) را تعدیل نمود و ویژگی‌های کارایی مجانبی و کارایی مجانبی ریسک را ارائه نمود. ماخوپادهای (۱۹۹۵) نیز مروری بر روش دنباله‌ای محض برای برآورد فاصله‌ای پارامتر میانگین فرایند خطی داشت. باسو و داس (۱۹۹۷) روش نمونه‌گیری دنباله‌ای محض را برای برآورد بردار پارامتر مدل خودبازگشتی مرتبه p مطالعه کردند.

روش نمونه‌گیری دومرحله‌ای با توجه به استراتژی توقفی که برای تعیین اندازه نمونه ارائه می‌دهد، در مقایسه با سایر روش‌های دنباله‌ای در اجرا ساده‌تر است اما در برآورد فاصله‌ای، متغیر این روش بیش‌برآورد دارد. برای رفع این مشکل، ماخوپادهای و داگان (۱۹۹۷) روش دومرحله‌ای را به منظور برآورد فاصله‌ای میانگین توزیع نرمال با در نظر گرفتن کرانی برای واریانس جامعه بهبود بخشید. سپس ماخوپادهای و داگان (۱۹۹۹) مطالعه خود را با

مفروضات جدید تعمیم دادند. لی و سیرام (۱۹۹۹) مدل ناخطی خودبازگشتی آستانه‌ای[1] ایستا را در نظر گرفتند. آن‌ها برآورد پارامترهای مدل بر اساس برآوردگرهای کمترین توان‌های دوم را با روش نمونه‌گیری دنباله‌ای محض بررسی کردند. ماخوپادهای و ایبد (۱۹۹۹) برای مدل رگرسیون خطی چندگانه با فرض نرمال و نانرمال بودن خطاها و مدل خودبازگشتی مرتبه p با اغتشاش‌های نانرمال روش نمونه‌گیری دنباله‌ای تسریع‌یافته کلی را بررسی کردند. سیرام (۲۰۰۱) برآورد فاصله‌ای پارامترهای مدل خودبازگشتی آستانه‌ای مرتبه اول با طول ثابت را با روش نمونه‌گیری دنباله‌ای محض بررسی نمود. سپس نتایج را به مدل خودبازگشتی آستانه‌ای مرتبه اول چندگانه تعمیم داد. مارتینسک (۲۰۰۱) برآورد پارامتر میانگین مدل خودبازگشتی با ضرایب تصادفی و توزیع حاشیه‌ای بتا را با روش نمونه‌گیری دنباله‌ای محض بررسی نمود. لی (۲۰۰۳) روش نمونه‌گیری دنباله‌ای محض را در مدل رگرسیون تصادفی‌شده در نظر گرفت. کشکوفسکی و کونو (۲۰۰۸) نیز روش نمونه‌گیری دنباله‌ای محض را برای برآورد نقطه‌ای در مدل خودبازگشتی مرتبه اول با ضرایب تصادفی بررسی کردند.

هم‌چنین فوردینر و همکاران (۲۰۰۹) برآورد پارامتر مدل خودبازگشتی مرتبه اول با اغتشاش‌های وابسته را با استفاده از روش نمونه‌گیری دنباله‌ای محض مطالعه نمودند. گمبی (۲۰۱۰) برآورد فاصله‌ای پارامتر میانگین فرایند خطی را با روش نمونه‌گیری دنباله‌ای محض بررسی نمود. سیرام و لسی (۲۰۱۴) با استفاده از شبیه‌سازی مونت کارلویی عملکرد روش نمونه‌گیری دنباله‌ای محض در مدل خودبازگشتی مرتبه اول و مدل خودبازگشتی آستانه‌ای مرتبه اول بر اساس ویژگی‌های کارایی مجانبی، کارایی مجانبی ریسک و سازگاری مجانبی برآوردگرهای نقطه‌ای و فاصله‌ای را به ترتیب بررسی کردند. لای (۲۰۱۴) بر برآورد فاصله‌ای با طول ثابت در مدل خودبازگشتی مرتبه p و به صورت کلی در مدل رگرسیون تصادفی با روش نمونه‌گیری دنباله‌ای محض مروری انجام داد.

کوزائینوف (۲۰۱۵) روش نمونه‌گیری دنباله‌ای محض را برای مدل خودبازگشتی مرتبه اول مقیاسی با پارامتر رانش بررسی نمود. محمودی و همکاران (۱۳۹۶) روش دنباله‌ای محض را برای برآورد پارامتر مقیاس توزیع نمایی وقتی که تابع ریسک توسط مقدار ثابت از پیش تعیین شده‌ای کران دار شده است، مطالعه کردند. در سال‌های اخیر، کارماکار و ماخوپادهای (۲۰۱۷، ۲۰۱۹) مدل‌های خودبازگشتی با ضرایب تصادفی از مرتبه p تک‌متغیره و چندمتغیره را در نظر گرفتند و روش نمونه‌گیری دنباله‌ای محض را بررسی نمودند. هم‌چنین، سیرام و صمدی (۲۰۱۹) بر روش نمونه‌گیری دنباله‌ای محض مروری داشتند که توسط سیرام (۱۹۸۸) ارائه شده بود. خلیفه و همکاران (۲۰۲۰) عملکرد روش دومرحله‌ای را برای برآورد فاصله‌ای در توزیع نمایی بررسی کردند. جان و ژانگ (۲۰۲۰) نیز رده‌ای از قاعده‌های توقف بر اساس روش دومرحله‌ای بهبودیافته تحت شرایط مینیمم‌سازی تابع زیان ارائه نمودند. اخیرا نیز سجادی پناه و همکاران (۲۰۲۱) روش نمونه‌گیری دومرحله‌ای را در مدل خودبازگشتی مرتبه اول در برآورد نقطه‌ای و فاصله‌ای بررسی کردند. سپس به مقایسه روش پیشنهادی با روش دنباله‌ای محض پرداختند.

روش دنباله‌ای محض کمترین اندازه نمونه را ارائه می‌دهد اما نمونه‌گیری تحت این روش هزینه بالایی دارد و از لحاظ اجرا بسیار زمان‌بر است. محققان برای رفع این محدودیت‌ها روش‌های نمونه‌گیری مختلفی را با ارائه قاعده توقف و استراتژی برای تعیین اندازه نمونه معرفی کردند که از جمله می‌توان به روش‌های دومرحله‌ای و دومرحله‌ای

[1] Threshold Autoregressive Process

بهبودیافته اشاره کرد. یکی از معایب روش نمونه‌گیری دومرحله‌ای در بعضی مسائل تصمیم‌گیری، مقدار حداقل اندازه نمونه اولیه است که انتخاب مقدار نمونه نامناسب هزینه بالای نمونه‌گیری را به ما تحمیل می‌کند. قابل ذکر است، سادگی محاسبات و سادگی در اجرا یکی از مزایای روش نمونه‌گیری دومرحله‌ای است. یک راهکار ارائه‌شده برای حل این مشکل نمونه‌گیری تحت این روش، ارائه قاعده توقفی برای تعیین اندازه نمونه اولیه است که استراتژی توقف جدید، روش نمونه‌گیری دومرحله‌ای بهبودیافته نامیده می‌شود. هم‌چنین از مزیت‌های دیگر این روش نمونه‌گیری در مقایسه با روش دومرحله‌ای، عملکرد بهتر روش در برآورد فاصله‌ای است. در واقع در این روش نمونه‌گیری ابتدا محقق با قاعده ارائه‌شده اندازه نمونه اولیه را تعیین می‌کند، سپس با استفاده از اندازه نمونه اولیه به تعیین اندازه نمونه نهایی می‌پردازد. همان‌طور که اشاره شد، نویسندگان این مقاله پیشتر به بررسی عملکرد روش دومرحله‌ای و مقایسه آن با روش نمونه‌گیری دنباله‌ای محض پرداخته‌اند. اما ضعف در انتخاب مقدار مناسب نمونه اولیه در دسترس در اجرا آن‌ها را بر آن داشت که روش نمونه‌گیری دومرحله‌ای بهبودیافته را برای پارامتر میانگین در مدل خودبازگشتی مرتبه اول مورد مطالعه قرار دهند.

در ادامه به بررسی عملکرد روش با استفاده از معیارهای سنجش کارایی روش پیشنهادی پرداخته شده است که در واقع این معیارها، کارایی روش پیشنهادی را در مقایسه با روش اندازه نمونه ثابت بهینه مورد ارزیابی قرار می‌دهند (گوش و همکاران، ۱۹۹۷). هم‌چنین روش دومرحله‌ای بهبودیافته با وجود تغییر اندک در تعریف قاعده توقف دارای دقت بیشتر در مقایسه با روش دومرحله‌ای و نرخ همگرایی سریعتری نیز در معیارهای سنجش کارایی روش در مقایسه با روش اندازه نمونه ثابت بهینه است، هنگامی که هزینه نمونه‌گیری یا طول فاصله به سمت مقادیر کوچک همگرا می‌شوند. معیارهای سنجش کارایی روش در برآورد نقطه‌ای و فاصله‌ای با ارائه قاعده توقف مناسب اثبات شده است که شامل کارایی مجانبی، کارایی مجانبی ریسک و سازگاری مجانبی می‌شوند. سپس در ادامه با استفاده از شبیه‌سازی به بررسی نتایج نظری پرداخته شده است که نرخ همگرایی روش پیشنهادی نسبت به روش با اندازه نمونه ثابت بهینه سنجیده می‌شود. هم‌چنین در انتها با مقایسه هر دو روش دومرحله‌ای بهبودیافته و دومرحله‌ای کاربرد روش پیشنهادی در پیاده‌سازی و اجرا مورد بررسی و مطالعه قرار گرفته است.

در این راستا، در بخش ۲ برآوردهای نقطه‌ای و فاصله‌ای میانگین حاصل از روش نمونه‌گیری دومرحله‌ای بهبودیافته به‌دست آمده است. در بخش ۳ عملکرد برآوردگرهای حاصل از روش نمونه‌گیری دومرحله‌ای بهبودیافته با مطالعه شبیه‌سازی بررسی شده است. در انتها در بخش ۴، کاربرد روش دومرحله‌ای در برآورد نقطه‌ای و فاصله‌ای با استفاده از داده واقعی ارائه شده است.

۲ روش دومرحله‌ای بهبودیافته

فرایند خودبازگشتی مرتبه اول $\mathrm{AR}(۱)$ با میانگین به صورت

$$X_i = \mu + \beta(X_{i-۱} - \mu) + \varepsilon_i, \quad i = ۱, ۲, \ldots,$$

در نظر گرفته می‌شود، که در آن $\{\varepsilon_i, i \geq ۱\}$ دنباله مستقل و هم‌توزیع (i.i.d) با توزیع مجهول F، میانگین صفر و واریانس $\sigma^۲$ هستند. هم‌چنین از دنباله $\{X_j, j < i\}$ ناهمبسته و از متغیر اولیه $X_۰$ با فرض $\mathbf{E}[X_۰^۲] < \infty$ مستقل هستند. برآوردگرهای کمترین توان‌های دوم β و μ به صورت

$$\hat{\beta}_n = \frac{\sum_{i=۱}^{n-۱}(X_i - \bar{X}_n)(X_{i+۱} - \bar{X}_n)}{\sum_{i=۱}^{n}(X_i - \bar{X}_n)^۲}, \qquad \hat{\mu} = \frac{\sum_{i=۲}^{n} X_i - \hat{\beta}\sum_{i=۲}^{n} X_{i-۱}}{(n-۱)(۱-\hat{\beta})}.$$

حاصل می‌شوند. از آنجا که $\bar{X}_n \simeq \frac{\sum_{i=۲}^{n} X_{i-۱}}{n-۱} \simeq \frac{\sum_{i=۲}^{n} X_i}{n-۱}$، برآوردگر پارامتر μ به صورت $\hat{\mu} \simeq \bar{X}_n = \frac{۱}{n}\sum_{i=۱}^{n} X_i$ حاصل می‌شود. هم‌چنین این برآوردگر وقتی $n \to \infty$ دارای توزیع مجانبی

$$\sqrt{n}\big(\bar{X}_n - \mu\big) \xrightarrow{D} N\big(۰, \frac{\sigma^۲}{(۱-\beta)^۲}\big). \tag{۱}$$

است (اندرسون، ۱۹۷۱). تابع زیان برای برآورد به صورت $L_n(\bar{X}_n, \mu) = a(\bar{X}_n - \mu)^۲ + cn$ در نظر گرفته می‌شود، که در آن $a > ۰$ وزن معلوم و $c > ۰$ هزینه هر واحد نمونه‌گیری است. هنگامی که β و σ معلوم هستند، تابع مخاطره به صورت،

$$R_n = \mathbf{E}\big[L_n(\bar{X}_n, \mu)\big] = a\mathbf{E}\big[\bar{X}_n - \mu\big]^۲ + cn = \frac{a\sigma^۲}{n(۱-\beta)^۲} + cn + o(n^{-۱}),$$

به‌دست می‌آید (سیرام، ۱۹۸۷). همانطور که اشاره شد، در بررسی روش پیشنهادی هدف مینیمم‌سازی تابع مخاطره است. با فرض نادیده گرفتن $o(n^{-۱})$ اندازه نمونه ثابت بهینه تقریبا برابر $n_c \simeq \left(\frac{a}{c}\right)^{۱/۲}\frac{\sigma}{(۱-\beta)}$ است. تابع مخاطره متناظر با اندازه نمونه ثابت بهینه نیز به صورت

$$R_{n_c} \simeq \frac{\sigma}{(۱-\beta)}\big(\frac{a}{c}\big)^{۱/۲}\big(۲\big(\frac{c^۲a}{a}\big)^{۱/۲}\big) \simeq ۲cn_c.$$

حاصل می‌شود. در بسیاری از شرایط در عمل σ یا β مجهول هستند و روش اندازه نمونه ثابت بهینه کاربرد ندارد. بر اساس آنچه گفته شد، راهکار پیشنهادی استفاده از روش‌های دنباله‌ای است.

تعریف ۱. در روش نمونه‌گیری دومرحله‌ای اشتاین[1]، با فرض محدودیتی از پیش تعیین‌شده که با b نشان داده می‌شود، برای برآورد θ، یک نمونه مقدماتی به حجم m از توزیع انتخاب می‌شود. در این روش قاعده توقف $T_m = \max\{m, \lfloor g(S,b)\rfloor + ۱\}$، است، که در آن $\lfloor x\rfloor$ بزرگترین عدد کوچکتر از x است. آماره S مبتنی بر

[1] Stein Two-Stage Procedure

نمونه تصادفی مقدماتی $X_1, \ldots, X_m$ و g تابعی مناسب از آماره $S = s(X_1, \ldots, X_m)$ و محدودیت b است. اگر $T_m = m$، نمونه مقدماتی m کافی بوده و پارامتر جامعه توسط $W_m = t(X_1, \ldots, X_m)$ برآورد می‌شود. در غیر این صورت نمونه تصادفی به حجم $T_m - m$ مستقل از نمونه مقدماتی از توزیع انتخاب شده و به نمونه اولیه اضافه می‌گردد. در نهایت پارامتر جامعه توسط آماره $W_{T_m} = t(X_1, \ldots, X_m, \ldots, X_{T_m})$ برآورد می‌شود (گوش و همکاران، ۱۹۹۷).

تعریف ۲. روش نمونه‌گیری دومرحله‌ای بهبودیافته[1] با فرض محدودیتی از پیش تعیین‌شده که با b نشان داده می‌شود، جهت برآورد پارامتر θ از جامعه، ابتدا با توجه به نمونه مقدماتی به حجم m_0، اندازه نمونه اولیه به صورت $m = \max\{m_0, [A^{1/2(1+\gamma)}]\}$ تعیین می‌شود، که در آن $\gamma \in (1/2, \infty)$، $m_0 \geq 2$ مقادیر صحیح ثابت و A نشان‌دهنده عکس هزینه نمونه‌گیری است. سپس قاعده توقف برای تعیین اندازه نمونه به صورت $T_{2m} = \max\{m, \lfloor g(S, b) \rfloor + 1\}$ تعیین می‌شود (گوش و همکاران، ۱۹۹۷).

اینک با توجه به استراتژی قاعده توقف روش دومرحله‌ای بهبودیافته به تعیین اندازه نمونه نهایی برای مدل پیشنهادی پرداخته می‌شود. ابتدا اندازه نمونه اولیه به صورت $m = \max\{m_0, [(\frac{a}{c})^{1/2(1+\gamma)}]\}$ تعریف می‌شود، که در آن $\gamma \in (1/2, \infty)$ و $m_0 \geq 2$ مقادیر صحیح ثابت هستند. اندازه نمونه نهایی با توجه به اندازه نمونه ثابت بهینه و اندازه نمونه اولیه به صورت

$$N_m = \max\{m, \lfloor (\frac{a}{c})^{1/2} (\frac{\hat{\sigma}_m}{|1 - \hat{\beta}_m|} + m^{-h}) \rfloor + 1\}, \tag{2}$$

ارائه می‌شود، که در آن $h > 0$ ثابت مناسب است. اکنون با توجه به تعریف قاعده توقف روش نمونه‌گیری دومرحله‌ای، متغیر توقف مبنی بر اندازه نمونه ثابت بهینه برای میانگین مدل به صورت

$$N_{m1} = \max\{m, \lfloor (\frac{a}{c})^{1/2} (\frac{\hat{\sigma}_m}{|1 - \hat{\beta}_m|} + m^{-h}) \rfloor + 1\},$$

به‌دست می‌آید، که در آن m عدد ثابت دلخواه است. همان‌طور که مشاهده می‌شود فقط مرحله دوم از استراتژی توقف روش نمونه‌گیری بهبودیافته یعنی N_m است و می‌بایست محقق در مورد انتخاب مقدار مناسب اندازه نمونه اولیه (m) تصمیم‌گیری نماید. در ادامه قضیه اصلی این بخش که نشان‌دهنده کارا بودن روش دومرحله‌ای بهبودیافته نسبت به روش اندازه نمونه ثابت بهینه است، ارائه می‌شود. برای این منظور به ازای $0 < \varepsilon < 1$ قرار داده می‌شود $n_1 = [n_c(1 - \varepsilon)]$ و $n_2 = [n_c(1 + \varepsilon)]$.

لم ۱. فرض کنید برای $p > 2$، $\mathbf{E}[|\varepsilon_1|^{2p}] < \infty$ و $h \in (0, \frac{p-2}{4})$ و داشته باشیم $(a/c)^{1/2(1+h)} \leqslant m = o(c^{-1/2(1+\gamma)})$. فرض کنید $s = p/2$ آن‌گاه برای هر $\varepsilon > 0$ هنگامی‌که $c \to 0$

[1] Modified Two-Stage Procedure

نتایج زیر حاصل می‌شوند

$$P(N_m \leq n_۱) = O\big(c^{s/۲(۱+h)}\big), \tag{۳}$$
$$P(N_m \geq n_۲) = O\big(c^{s/۲(۱+h)}\big). \tag{۴}$$

برهان: برای اثبات (۳) می‌توان نوشت:

$$\begin{aligned}
P(N_m \leq n_۱) &= P\big(N_m \leq (۱-\varepsilon)(a/c)^{۱/۲}\frac{\sigma}{(۱-\beta)}\big)\\
&\leq P\big(\frac{\hat{\sigma}^۲_m}{(۱-\hat{\beta}_m)^۲} \leq (۱-\varepsilon)^۲\frac{\sigma^۲}{(۱-\beta)^۲}\big)\\
&\leq P\big(\frac{\hat{\sigma}^۲_m}{(۱-\hat{\beta}_m)^۲} - \frac{\sigma^۲}{(۱-\beta)^۲} \leq (\varepsilon^۲ - ۲\varepsilon)\frac{\sigma^۲}{(۱-\beta)^۲}\big)\\
&\leq P\big(\big|\frac{\hat{\sigma}^۲_m}{(۱-\hat{\beta}_m)^۲} - \frac{\sigma^۲}{(۱-\beta)^۲}\big| \geq \frac{\varepsilon(۲-\varepsilon)\sigma^۲}{(۱-\beta)^۲}\big)\\
&= O\big(m^{-p/۲}\big) = O\big((c/a)^{s/۲(۱+h)}\big) = O\big(c^{s/۲(۱+h)}\big).
\end{aligned}$$

تساوی قبل با توجه به لم ۳ از سیرام (۱۹۸۷) نتیجه می‌شود. برای اثبات (۴)، نوشته می‌شود:

$$\begin{aligned}
P(N_m \geq n_۲) &\leq P\big(\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} > (۱+\varepsilon)\frac{\sigma}{(۱-\beta)} - (a/c)^{-۱/۲}۲ - m^{-h}\big)\\
&\leq P\big(\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} > (۱+\varepsilon)\frac{\sigma}{(۱-\beta)} - (a/c)^{-۱/۲}۲ - m^{-h} - \frac{\sigma}{|۱-\beta|} + \frac{\sigma}{|۱-\beta|}\big)\\
&= P\big(\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} - \frac{\sigma}{|۱-\beta|} > (a/c)^{-۱/۲}\big(n_۲ - (a/c)^{۱/۲}\frac{\sigma}{|۱-\beta|}\big) - m^{-h}\big)\\
&= P\big(\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} - \frac{\sigma}{|۱-\beta|} > (a/c)^{-۱/۲}(n_۲ - n_c) - (a/c)^{-۱/۲}۲ - m^{-h}\big)
\end{aligned}$$

با انتخاب cهای به اندازه کافی کوچک، داریم

$$(a/c)^{-۱/۲}(n_۲ - n_c) - (a/c)^{-۱/۲}۲ - m^{-h} > \frac{\varepsilon\sigma}{۲(۱-\beta)}$$

بنابراین با توجه به لم ۳ از سیرام (۱۹۸۷) داریم

$$\begin{aligned}P(N_m \geq n_۲) &\leq P\Big(\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} - \frac{\sigma}{|۱-\beta|} > \frac{\varepsilon\sigma}{۲(۱-\beta)}\Big)\\ &\leq P\Big(\Big|\frac{\hat{\sigma}^۲_m}{|۱-\hat{\beta}_m|^۲} - \frac{\sigma^۲}{(۱-\beta)^۲}\Big| > \frac{\varepsilon^۲\sigma^۲}{۴(۱-\beta)^۲}\Big)\\ &= O\big(m^{-p/۲}\big) = O\big((c/a)^{s/۲(۱+h)}\big) = O\big(c^{s/۲(۱+h)}\big).\end{aligned}$$

قضیه ۱. فرض کنید ۲ $p >$، $\mathbf{E}[|\varepsilon_۱|^{۲p}] < \infty$. همچنین برای h در عبارت (۲) داشته باشیم $h \in \big(\circ, \frac{p-۲}{۴}\big)$ و $(a/c)^{۱/۲(۱+h)} \leqslant m = o(c^{-۱/۲})$. آن‌گاه هنگامی‌که $c \to \circ$ داریم

$$\frac{N_m}{n_c} \xrightarrow{a.s} ۱, \tag{۵}$$

$$\mathbf{E}\Big[\frac{N_m}{n_c}\Big] \to ۱, \tag{۶}$$

$$\frac{R_{N_m}}{R_{n_c}} \to ۱. \tag{۷}$$

برهان: با توجه به (۲) و با فرض $I_{(N_m=m)}$ به عنوان تابع نشانگر در $N_m = m$ می‌توان نوشت:

$$\Big(\frac{a}{c}\Big)^{۱/۲}\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} + m^{-h} \leq N_m \leq \Big(\frac{a}{c}\Big)^{۱/۲}\frac{\hat{\sigma}_m}{|۱-\hat{\beta}_m|} + m^{-h} + mI_{(N_m=m)}.$$

با توجه به $\frac{m}{n_c} \xrightarrow{a.s} \circ$، لم ۱ و $I_{(N_m=m)} \xrightarrow{a.s} \circ$ با گرفتن حد و امید از رابطه بالا هنگامی‌که $c \to \circ$، عبارات (۵) و (۶) به ترتیب نتیجه می‌شوند. جهت دستیابی به رابطه (۷) می‌توان نوشت:

$$R_{N_m} = \mathbf{E}\big[L_{N_m}(\bar{X}_{N_m}, \mu)\big] = a\mathbf{E}[\bar{X}^۲_{N_m}] + c\mathbf{E}\big[N_m\big].$$

آن‌گاه با تقسیم بر تابع مخاطره متناظر با اندازه نمونه ثابت بهینه نتیجه می‌شود

$$\frac{R_{N_m}}{R_{n_c}} = \frac{a\mathbf{E}[\bar{X}^۲_{N_m}] + c\mathbf{E}[N_m]}{۲cn_c}.$$

با توجه به عبارت (۶)، کافی است نشان داده شود که عبارت دوم سمت راست تساوی بالا به سمت ۱/۲ همگرا است. بدون از دست دادن کلیت مسئله فرض می‌شود $\mu = \circ$ و $\sigma^۲ = ۱$. به‌علاوه $A = [n_۲ < N_{m_c} < n_۳]$،

$B=[N_{m_c}\leq n_۲]$ و $D=[N_{m_c}\geq n_۳]$ در نظر گرفته می‌شود. اکنون می‌بایست روابط

$$\frac{a\mathbf{E}[\bar{X}^{۲}_{N_m}]I_{\bar{A}}}{cn_c}\to ۰, \tag{۸}$$

$$\frac{a\mathbf{E}[(\bar{X}_{N_m}-\bar{X}_{n_c})^{۲}]I_{A}}{cn_c}\to ۰. \tag{۹}$$

اثبات شود. با توجه به عبارت (۳) و با استدلال مشابه در اثبات رابطه (۳۱.۲) از سیرام (۱۹۸۷) معادله (۸) نتیجه می‌شود که به طور مشابه می‌توان نشان داد

$$\frac{۱}{cn_c}\mathbf{E}[\bar{X}^{۲}_{N_m}I_B]\to ۰,\qquad \frac{۱}{cn_c}\mathbf{E}[\bar{X}^{۲}_{N_m}I_D]\to ۰.$$

با به‌کارگیری نامساوی‌های مثلثی، قضیه فوبینی و استدلال مشابه در اثبات رابطه (۳۲.۲) از سیرام (۱۹۸۷) معادله (۹) را نتیجه می‌دهد و اثبات کامل می‌شود.

قضیه ۲. به‌ازای $p>۲$، فرض کنید $\mathbf{E}[|\varepsilon_۱|^{۲p}]<\infty$ و به ازای $h\in\left(۰,\frac{p-۲}{۴}\right)$ داشته باشیم $(a/c)^{۱/۲(۱+h)}\leqslant m=o(c^{-۱/۲(۱+\gamma)})$، آن‌گاه وقتی $c\to ۰$ داریم

$$\sqrt{N_m}\big(\bar{X}_{N_m}-\mu\big)\xrightarrow{D} N\big(۰,\frac{\sigma^۲}{(۱-\beta)^۲}\big).$$

برهان: با توجه به (۸) و (۹) و تساوی

$$\sqrt{N_m}(\bar{X}_{N_m}-\mu)=\sqrt{\frac{N_m}{n_c}}\sqrt{n_c}(\bar{X}_{N_m}-\bar{X}_{n_c})+\sqrt{\frac{N_m}{n_c}}\sqrt{n_c}(\bar{X}_{n_c}-\mu).$$

داریم $\sqrt{n_c}|\bar{X}_{N_m}-\bar{X}_{n_c}|\xrightarrow{P}۰$ و بنابر (۱) و (۵) و قضیه اسلاتسکی نتیجه حاصل می‌شود.

در ادامه بازهٔ اطمینان برای μ با طول ثابت $(۲d)$ و ضریب اطمینان $۱-\alpha$ با استفاده از روش دومرحله‌ای بهبودیافته به دست آورده می‌شود. بازهٔ اطمینان بر اساس نمونه ثابت n، $\left(\bar{X}_n-d,\bar{X}_n+d\right)$ با احتمال پوشش $۱-\alpha$ در نظر گرفته می‌شود که اندازه نمونه ثابت بهینه نیز تقریبا برابر با

$k_d \simeq [d^{-۲}(۱-\beta)^{-۲}\sigma^۲ z^۲_{(۱-\alpha)/۲}]$ به‌دست می‌آید. هنگامی‌که $d \to ۰$ نتیجه می‌شود

$$\begin{aligned} P\big(\mu \epsilon C_{k_d}\big) &= P\big(\big|\bar{X}_{k_d} - \mu\big| < d\big) \\ &= P\big(\frac{\sqrt{k_d}\big|\bar{X}_{k_d} - \mu\big|}{\sqrt{\sigma^۲/(۱-\beta)^۲}} < \frac{\sqrt{k_d}d}{\sqrt{\sigma^۲/(۱-\beta)^۲}}\big) \longrightarrow ۱ - \alpha. \end{aligned}$$

همانطور که در بخش قبل نیز اشاره شد، هنگامی‌که σ یا β مجهول هستند، نمی‌توان از روش اندازه نمونه ثابت بهینه استفاده کرد. در این شرایط راهکار مناسب پیشنهادی برای تعیین اندازه نمونه، استفاده از روش نمونه‌گیری دومرحله‌ای بهبودیافته است. در این حالت نیز مشابه برآورد نقطه‌ای، ابتدا قاعده توقف برای تعیین اندازه نمونه اولیه به صورت $m^d = \max\{m_۰, [(z_{(۱-\alpha)/۲}/d)^{۲/(۱+\gamma)}] + ۱\}$ ارائه می‌شود. اندازه نمونه نهایی نیز با توجه به طول فاصله، اندازه نمونه اولیه و اندازه نمونه ثابت بهینه به صورت

$$N^d_m = \max\{m^d, \lfloor d^{-۲} z^۲_{(۱-\alpha)/۲}(\frac{\hat{\sigma}^۲_m}{|۱-\hat{\beta}_m|^۲} + m^{-h})\rfloor + ۱\}. \tag{۱۰}$$

تعریف می‌شود. نتیجه اصلی این بخش با ارائه یک قضیه مطرح می‌شود که نشان‌دهنده ویژگی‌های کارایی مجانبی مرتبه اول و سازگاری مجانبی برآوردگر بر اساس روش نمونه‌گیری دومرحله‌ای بهبودیافته است.

لم ۲. فرض کنید برای $p > ۲$، $\mathbf{E}[|\varepsilon_۱|^{۲p}] < \infty$، $h \in \big(۰, \frac{p-۲}{۴}\big)$ و $m^d = o(d^{-۲/(۱+\gamma)})$. فرض کنید $s = p/۲$ آن‌گاه برای هر $\varepsilon > ۰$ وقتی $c \to ۰$ آنگاه $P(N_m = m^d) = O\big(d^{۲\gamma/۲(۱+\gamma)}\big)$. **برهان:**

$$\begin{aligned} P(N^d_{m^d} = m^d) &\le\ P\big(\frac{\hat{\sigma}^۲_m}{(۱-\hat{\beta}_m)^۲} \le m^d d^۲ z^{-۲}_{(۱-\alpha)/۲}\big) \\ &\le\ P\big(\frac{\hat{\sigma}^۲_m}{(۱-\hat{\beta}_m)^۲} - \frac{\sigma^۲}{(۱-\beta)^۲} \le m^d d^۲ z^{-۲}_{(۱-\alpha)/۲} - \frac{\sigma^۲}{(۱-\beta)^۲}\big) \\ &\le\ P\big(\big|\frac{\hat{\sigma}^۲_m}{(۱-\hat{\beta}_m)^۲} - \frac{\sigma^۲}{(۱-\beta)^۲}\big| \ge m^d d^۲ z^{-۲}_{(۱-\alpha)/۲}\big) \\ &=\ O\big(d^{۲\gamma/۲(۱+\gamma)}\big). \end{aligned}$$

با توجه به لم ۳ از سیرام (۱۹۸۷) نتیجه قبل به‌سادگی به‌دست می‌آید و اثبات کامل می‌شود.

قضیه ۳. فرض کنید برای $p > ۲$، $\mathbf{E}[|\varepsilon_۱|^{۲p}] < \infty$ برقرار باشد. به‌علاوه $h \in \big(۰, \frac{p-۲}{۴}\big)$ که h در (۱۰)

تعریف شده است. آن‌گاه هنگامی‌که $d \to 0$،

$$\frac{N_m^d}{k_d} \xrightarrow{a.s} 1, \tag{۱۱}$$

$$\mathbf{E}\Big[\frac{N_m^d}{k_d}\Big] \to 1, \tag{۱۲}$$

$$P\big(\mu \in C_{N_m^d}\big) \to 1-\alpha. \tag{۱۳}$$

برهان: از تعریف (۱۰) نامساوی زیر حاصل می‌شود

$$d^{-2} z^2_{(1-\alpha)/2} \frac{\hat{\sigma}^2_m}{|1-\hat{\beta}_m|^2} + m^{-h} \leq N_m^d \leq d^{-2} z^2_{(1-\alpha)/2} \frac{\hat{\sigma}^2_m}{|1-\hat{\beta}_m|^2} + m^{-h} + m I_{(N_m^d = m^d)}.$$

با به‌کارگیری لم ۲ و با توجه به نکات $\frac{m}{k_d} \xrightarrow{a.s} 0$ و $I_{(N_m^d=m^d)} \xrightarrow{a.s} 0$، به‌سادگی با گرفتن حد و امید از طرفین نامساوی قبل هنگامی‌که $d \to 0$، عبارات (۱۱) و (۱۲) به‌دست می‌آیند. استدلال رابطه (۱۳) به صورت زیر دنبال می‌شود

$$\sqrt{N_m^d}(\bar{X}_{N_m^d} - \mu) = \sqrt{\frac{N_m^d}{k_d}} \sqrt{k_d} (\bar{X}_{N_m^d} - \bar{X}_{k_d}) + \sqrt{\frac{N_m^d}{k_d}} \sqrt{k_d} (\bar{X}_{k_d} - \mu). \tag{۱۴}$$

از عبارات (۸) و (۹)، داریم $k_d E(\bar{X}_{N_m^d} - \bar{X}_{k_d})^2 \to 0$ و $\sqrt{k_d}|\bar{X}_{N_m^d} - \bar{X}_{k_d}| \to 0$. پس

$$\begin{aligned} P(\bar{X}_{N_m^d} - d < \mu < \bar{X}_{N_m^d} + d) \;&=\; P\Big(\frac{N_d(\bar{X}_{N_m^d} - \mu)^2}{\frac{\sigma^2}{(1-\beta)^2}} < \frac{N_m^d d^2}{\frac{\sigma^2}{(1-\beta)^2}}\Big) \\ &\geq\; P\Big(\frac{N_d(\bar{X}_{N_m^d} - \mu)^2}{\frac{\sigma^2}{(1-\beta)^2}} < \frac{\frac{\hat{\sigma}^2_m}{(1-\hat{\beta}_m)^2}}{\frac{\sigma^2}{(1-\beta)^2}} z^2_{(1-\alpha)/2}\Big) \\ &\to\; 1-\alpha. \end{aligned}$$

با توجه به عبارت (۱۴)، (۱۱)، (۱) و قضیه اسلاتسکی اثبات حاصل می‌شود.

۳ مطالعات شبیه‌سازی

در این بخش با استفاده از شبیه‌سازی مونت کارلویی عملکرد روش نمونه‌گیری دومرحله‌ای بهبودیافته در مدل خودبازگشتی مرتبه اول با میانگین بررسی می‌شود. مدل خودبازگشتی مرتبه اول با فرض $\beta = ۰/۳$، توزیع اغتشاش‌ها $\varepsilon_i \sim N(۰,۱)$ و مقادیر میانگین از -۶ تا ۶ با افزایش ۲ به بازه مقادیر پیشنهادی در نظر گرفته شده است. به منظور بررسی عملکرد برآورد نقطه‌ای با توجه به روابط (۵)-(۷)، متغیرهای توقف حاصل از روش دومرحله‌ای بهبودیافته $(\widehat{EN_m})$، نسبت متغیر توقف روش به اندازه نمونه ثابت بهینه $(\widehat{EN_m}/n_c)$ و تابع مخاطره حاصل از دو روش دومرحله‌ای بهبودیافته و اندازه نمونه ثابت بهینه $(R_{\bar{X}_{N_m}}/R_{n_c})$ به‌دست آمده‌اند. هم‌چنین ریشه دوم میانگین توان‌های دوم خطای میانگین $(RMSE\bar{X}_{N_m})$ برای بررسی دقت روش در برآورد نقطه‌ای محاسبه شده‌اند. نتایج حاصل از برآورد نقطه‌ای با در نظر گرفتن توان ۰/۲۸ و $a = ۱$ با ۱۰۰۰۰ تکرار در فرایند شبیه‌سازی حاصل شده است که برای تعیین مقدار اندازه نمونه اولیه، مقادیر $m_۰$ و هزینه متناظر با آن $(m_۰, c) = (۱۰, ۰/۰۵), (۳۵, ۰/۰۰۵), (۵۰, ۰/۰۰۰۵)$ در نظر گرفته شده است. با توجه به مقادیر مفروض ابتدا اندازه نمونه اولیه (m) به‌دست می‌آید و سپس اندازه نمونه نهایی حاصل می‌شود که با استفاده از آن، متوسط مقادیر برآوردگرها در هر تکرار به‌دست می‌آید. هم‌چنین برای بررسی عملکرد روش در برآورد فاصله‌ای با توجه به روابط (۱۱)-(۱۳)، متغیر توقف فاصله‌ای $(\widehat{EN^d_m})$، نسبت متغیر توقف به اندازه نمونه ثابت بهینه $(\widehat{EN^d_m}/k_d)$ و احتمال‌های پوشش $(CP(C_{\bar{X}_{N^d_m}}))$ با فرض ضریب اطمینان ۰/۹۵ محاسبه شده‌اند. برای دستیابی به نتایج و تعیین اندازه نمونه اولیه، $m_۰ = ۲۵$، توان ۱/۲۵ و طول فاصله مقادیر $d = ۱/۲, ۰/۵$ در نظر گرفته شده‌اند. هم‌چنین توزیع مجانبی برآوردگر میانگین نیز با انجام یک آزمون ناپارامتری کولموگروف-اسمیرنوف بررسی و به منظور درک شهودی بهتر، هیستوگرام آن نیز با برازش توزیع نرمال رسم شده است. شکل‌ها نیز با فرض $\beta = ۰/۳$، $m_۰ = ۲۰, ۵۰$، $a = ۱$ و $c = ۰/۰۰۰۵$ با توان ۱/۲۵ رسم شده‌اند. نتایج حاصل از برآورد نقطه‌ای، فاصله‌ای، آزمون کلموگروف-اسمیرنوف و نمودارهای مربوط به هیستوگرام توزیع مجانبی در جدول‌های ۱، ۲، ۳ و در شکل ۱ به ترتیب گزارش و رسم شده‌اند. قابل ذکر است، همه محاسبات با استفاده از نرم‌افزار R انجام شده است.

همانطور که در جدول ۱ ملاحظه می‌شود، متغیر توقف روش دومرحله‌ای با کاهش هزینه افزایش می‌یابد. نسبت متغیر توقف به اندازه نمونه ثابت بهینه با کاهش هزینه به مقدار یک نزدیک می‌شود که تایید کننده ویژگی کارایی مجانبی در قضیه ۱ است. هم‌چنین نسبت تابع مخاطره حاصل از روش دومرحله‌ای بهبودیافته به تابع مخاطره متناظر با اندازه نمونه ثابت بهینه با کاهش هزینه به سمت مقدار یک نزدیک می‌شود که موید ویژگی کارایی مجانبی ریسک قضیه ۱ است. ریشه دوم میانگین توان‌های دوم خطای برآوردگر نیز با کاهش هزینه و افزایش اندازه نمونه، دارای مقداری ناچیز است که بر دقت قاعده توقف پیشنهادی در برآورد نقطه‌ای تاکید دارد. هم‌چنین همان‌گونه که از نتایج حاصل مشاهده می‌شود، برای مقادیر کوچک $m_۰$، ویژگی‌ها با نرخ همگرایی سریعی به مقدار ۱ همگرا می‌شوند که نشان‌دهنده عملکرد خوب روش در برآورد نقطه‌ای است. طبق نتایج جدول ۳، آزمون در سطح $\alpha = ۰/۰۵$ معنی‌دار نیست که تایید کننده توزیع مجانبی نرمال برآوردگر میانگین و قضیه ۲ است. در شکل ۱ با برازش منحنی نرمال

جدول ۱. برآورد نقطه‌ای بر اساس متغیر توقف N_m.

$R_{\bar{X}_{N_m}}/R_{n_c}$	$RMSE\bar{X}_{N_m}$	$\widehat{EN_m}/n_c$	$\widehat{EN_m}$	$(m_\circ, c, n_c, \mu)$
۱٫۴۳۷۴	۰٫۷۱۱۸	۲٫۴۷۴۸	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, −۶)
۱٫۴۳۴۳	۰٫۷۰۷۰	۲٫۴۷۴۹	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, −۴)
۱٫۴۳۲۸	۰٫۷۰۸۴	۲٫۴۷۴۹	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, −۲)
۱٫۴۳۷۱	۰٫۷۱۳۹	۲٫۴۷۵۰	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, ۰)
۱٫۴۳۵۴	۰٫۷۱۶۲	۲٫۴۷۵۰	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, ۲)
۱٫۴۳۵۵	۰٫۷۰۹۳	۲٫۴۷۴۸	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, ۴)
۱٫۴۳۴۵	۰٫۷۱۱۱	۲٫۴۷۴۹	۵٫۰۰۰۰	(۱۰, ۰٫۰۵, ۲٫۱۲۹۵, ۶)
۱٫۱۵۰۰	۰٫۲۴۳۷	۱٫۷۳۳۱	۳۵٫۰۱۴۵	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, −۶)
۱٫۱۵۵۴	۰٫۲۴۵۲	۱٫۷۳۳۶	۳۵٫۰۲۴۳	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, −۴)
۱٫۱۴۹۵	۰٫۲۴۳۹	۱٫۷۳۳۲	۳۵٫۰۱۷۱	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, −۲)
۱٫۱۴۷۵	۰٫۲۴۲۵	۱٫۷۳۳۵	۳۵٫۰۲۲۰	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, ۰)
۱٫۱۵۵۶	۰٫۲۴۶۹	۱٫۷۳۳۲	۳۵٫۰۱۷۸	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, ۲)
۱٫۱۵۸۷	۰٫۲۴۶۸	۱٫۷۳۳۳	۳۵٫۰۱۸۵	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, ۴)
۱٫۱۵۹۳	۰٫۲۴۷۹	۱٫۷۳۳۰	۳۵٫۰۱۳۴	(۳۵, ۰٫۰۰۵, ۶٫۷۳۴۳, ۶)
۱٫۰۱۹۱	۰٫۱۸۵۴	۰٫۹۸۱۲	۶۲٫۶۸۸۷	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, −۶)
۱٫۰۲۵۰	۰٫۱۸۶۹	۰٫۹۸۰۹	۶۲٫۶۷۰۵	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, −۴)
۱٫۰۱۳۳	۰٫۱۸۴۹	۰٫۹۷۸۲	۶۲٫۴۹۸۵	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, −۲)
۱٫۰۰۷۶	۰٫۱۸۴۲	۰٫۹۷۸۹	۶۲٫۵۴۴۶	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, ۰)
۱٫۰۲۱۴	۰٫۱۸۶۱	۰٫۹۷۹۳	۶۲٫۵۶۸۸	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, ۲)
۱٫۰۱۵۳	۰٫۱۸۵۱	۰٫۹۷۹۴	۶۲٫۵۷۷۶	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, ۴)
۱٫۰۱۸۰	۰٫۱۸۵۰	۰٫۹۸۳۵	۶۲٫۸۳۴۹	(۵۰, ۰٫۰۰۰۵, ۲۱٫۲۹۵۸, ۶)

جدول ۲. برآورد فاصله‌ای بر اساس متغیر توقف N_m^d.

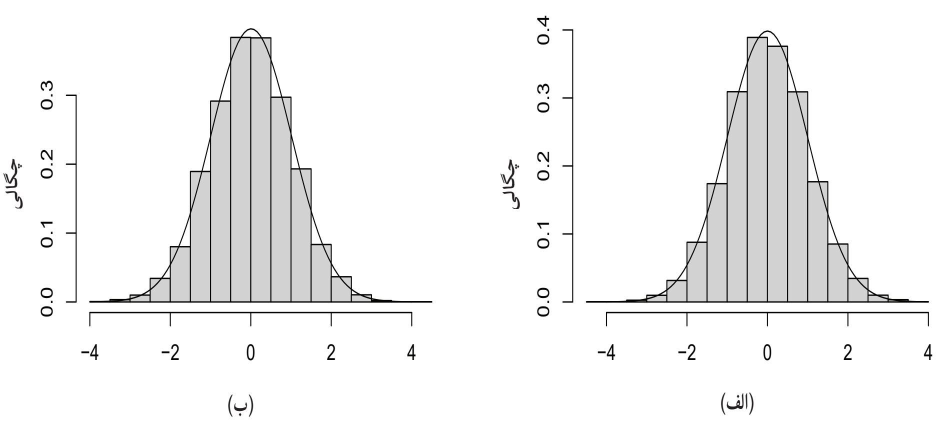
$CP(C_{\bar{X}_{N_m^d}})$	$\widehat{EN_m^d}/k_d$	$\widehat{EN_m^d}$	(m, d, k_d, μ)
۰٫۹۹۹۹	۴٫۵۹۳۸	۲۵٫۰۱۰۲	(۲۵, ۱٫۲, ۵٫۴۴۴۲, −۶)
۰٫۹۹۸۹	۴٫۵۹۳۶	۲۵٫۰۰۹۱	(۲۵, ۱٫۲, ۵٫۴۴۴۲, −۴)
۰٫۹۹۹۸	۴٫۵۹۳۸	۲۵٫۰۱۰۱	(۲۵, ۱٫۲, ۵٫۴۴۴۲, −۲)
۰٫۹۹۹۶	۴٫۵۹۳۸	۲۵٫۰۱۴۰	(۲۵, ۱٫۲, ۵٫۴۴۴۲, ۰)
۰٫۹۹۹۸	۴٫۵۹۳۰	۲۵٫۰۰۵۵	(۲۵, ۱٫۲, ۵٫۴۴۴۲, ۲)
۰٫۹۹۹۸	۴٫۵۹۳۵	۲۵٫۰۰۸۴	(۲۵, ۱٫۲, ۵٫۴۴۴۲, ۴)
۰٫۹۹۹۵	۴٫۵۹۵۳	۲۵٫۰۱۸۳	(۲۵, ۱٫۲, ۵٫۴۴۴۲, ۶)
۰٫۹۴۸۲	۱٫۲۳۱۴	۳۸٫۶۱۶۵	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, −۶)
۰٫۹۴۷۱	۱٫۱۸۹۵	۳۷٫۳۰۳۴	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, −۴)
۰٫۹۴۱۱	۱٫۱۳۸۸	۳۵٫۷۱۴۳	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, −۲)
۰٫۹۴۰۹	۱٫۱۳۱۳	۳۵٫۴۷۷۲	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, ۰)
۰٫۹۳۹۴	۱٫۱۴۲۸	۳۵٫۸۳۸۴	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, ۲)
۰٫۹۴۵۹	۱٫۱۸۱۹	۳۷٫۰۶۵۱	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, ۴)
۰٫۹۴۷۶	۱٫۲۲۴۵	۳۸٫۳۹۹۶	(۲۵, ۰٫۵, ۲۱٫۷۷۶۹, ۶)

به هسیتوگرام توزیع مجانبی، مشاهده می‌شود که برای مقادیر مختلف $(m_\circ, c)$ دارای توزیع نرمال است. متغیر توقف فاصله‌ای با کاهش طول فاصله افزایش می‌یابد و نسبت متغیر به اندازه نمونه ثابت بهینه نیز به مقدار یک

نزدیک می‌شود. نزدیکی نسبت متغیر توقف به اندازه نمونه ثابت بهینه در برآورد فاصله‌ای نیز تایید کننده ویژگی کارایی مجانبی قضیه ۳ است. احتمال پوشش نیز با کاهش طول فاصله به مقدار ۰/۹۵ نزدیک می‌شود که موید ویژگی سازگاری مجانبی در قضیه ۳ است.

جدول ۳. نتایج آزمون کلموگروف-اسمیرنوف.

$p-value$	D	(m_0, c)
۰/۹۳۰۹	۰/۰۰۵۴	(۲۰, ۰/۰۰۵)
۰/۹۹۴۳	۰/۰۰۴۲	(۲۰, ۰/۰۰۰۵)
۰/۰۰۶۱	۰/۸۳۸	(۵۰, ۰/۰۰۵)
۰/۸۵۳۴	۰/۰۰۶۰	(۵۰, ۰/۰۰۰۵)



شکل ۱. برازش توزیع نرمال به توزیع مجانبی بر اساس $m_0 =$ ۲۰, ۵۰ به ترتیب در (الف) و (ب) با فرض $\mu =$ ۵

۴ مثال کاربردی

مجموعه داده‌های یک فرآیند شیمیایی شامل قرائت ساعتی از گرانروی فرآیند در نظر گرفته شده است. مجموعه داده‌ها توسط شرفی و نعمت الهی (۲۰۱۶) مورد مطالعه قرار گرفته که مدل خودبازگشتی مرتبه اول با میانگین و توزیع اغتشاش نرمال چوله به آن برازش شده است. در این مطالعه عملکرد روش پیشنهادی با روش دومرحله‌ای در برآورد نقطه‌ای و فاصله‌ای بر اساس متغیر توقف، برآوردگرها و واریانس مدل با فرض $h =$ ۱، توان‌های ۰/۲۸ و

۱/۲۵ مقایسه شده است. متغیرهای توقف روش دومرحله‌ای برای برآورد نقطه‌ای و فاصله‌ای با توجه به اندازه نمونه ثابت بهینه به ترتیب به صورت

$$N_{m1} = \max\{m, \lfloor \Big(\frac{a}{c}\Big)^{1/2} (\frac{\hat{\sigma}_m}{|1-\hat{\beta}_m|} + m^{-h})\rfloor + 1\},$$
$$N^d_{m1} = \max\{m, \lfloor d^{-2} z^2_{(1-\alpha)/2} (\frac{\hat{\sigma}^2_m}{|1-\hat{\beta}_m|^2} + m^{-h})\rfloor + 1\},$$

در نظر گرفته می‌شوند، که در آن m اندازه نمونه اولیه است. نتایج برای متغیرهای توقف برای برآورد نقطه‌ای و فاصله‌ای حاصل از هر دو روش به ترتیب در جدول‌های ۴ و ۵ گزارش شده است. همان‌گونه که در جدول ۴ مشاهده می‌شود، متغیر توقف برای هزینه کم دقیقا با روش دومرحله‌ای یکسان حاصل می‌شود. هم‌چنین متغیر توقف روش دومرحله‌ای بهبودیافته مقادیر کمتر از روش دومرحله‌ای را با کاهش هزینه به‌دست می‌آورد. برآوردگرها با کاهش هزینه دارای مقادیری نزدیک به‌هم هستند. طبق نتایج جدول ۵ نیز، مقادیر متغیرهای توقف و برآوردگرهای حاصل از هر دو روش نزدیکی بسیاری به یکدیگر دارند. طبق دستاوردها هر دو روش عملکرد بسیار مشابهی دارند.

جدول ۴. برآورد نقطه‌ای بر اساس متغیرهای توقف N_m و N_{m1}.

$\widehat{\sigma}_{N_{m1}}$	$\widehat{\sigma}_{N_m}$	$\widehat{\beta}_{N_{m1}}$	$\widehat{\beta}_{N_m}$	$\bar{X}_{N_{m1}}$	$\bar{X}_{N_m}$	N_{m1}	N_m	(m, c)
۰/۰۵۵۷	۰/۰۵۵۲	−۰/۴۱۶۶	−۰/۴۱۶۰	۷/۸۵۰۰	۷/۷۹۰۰	۴	۴	(۲, ۰/۰۵)
۰/۱۱۹۵	۰/۱۲۹۳	۰/۳۳۱۶	۰/۲۷۸۴	۸/۰۶۶۶	۸/۰۲۵۰	۹	۸	(۸, ۰/۰۰۵)
۰/۰۸۰۸	۰/۰۹۵۳	۰/۷۱۱۱	۰/۶۹۵۹	۸/۵۱۰۶	۸/۴۷۰۱	۴۷	۴۰	(۴۰, ۰/۰۰۰۵)

جدول ۵. برآورد فاصله‌ای بر اساس متغیرهای توقف N^d_m و N^d_{m1}.

$\widehat{\sigma}_{N^d_{m1}}$	$\widehat{\sigma}_{N^d_m}$	$\widehat{\beta}_{N^d_{m1}}$	$\widehat{\beta}_{N^d_m}$	$\bar{X}_{N^d_{m1}}$	$\bar{X}_{N^d_m}$	N^d_{m1}	N^d_m	(m, d)
۰/۰۵۵۷	۰/۰۵۷۲	−۰/۴۱۶۶	−۰/۴۱۰۰	۷/۸۵	۷/۸۰	۴	۴	(۲, ۰/۸)
۰/۱۱۹۵	۰/۱۱۰۳	۰/۳۳۱۶	۰/۳۳۳۳	۸/۰۶۶۶	۸/۰۶۰۵	۹	۹	(۸, ۰/۴)
۰/۰۸۷۰	۰/۰۸۷۱	۰/۲۹۹۶	۰/۲۹۹۶	۸/۱۰۷۶	۸/۰۰۱۲	۱۳	۱۳	(۱۱, ۰/۳)

بحث و نتیجه‌گیری

در این مقاله، روش نمونه‌گیری دومرحله‌ای بهبودیافته در مدل‌های خودبازگشتی مرتبه اول با میانگین بررسی شده است. در این مطالعه، ویژگی‌های کارایی مجانبی، کارایی مجانبی ریسک، سازگاری مجانبی و توزیع مجانبی برآوردگر حاصل از روش نمونه‌گیری دومرحله‌ای بهبودیافته اثبات شده که نشان‌دهنده عملکرد نزدیک روش دومرحله‌ای بهبودیافته برای مقادیر کوچک هزینه و طول فاصله در مقایسه با روش اندازه نمونه ثابت بهینه است. برای بررسی عملکرد روش

پیشنهادی در براورد نقطه‌ای و فاصله‌ای مطالعه شبیه‌سازی مونت کارلویی انجام گرفته است که دستاوردهای شبیه‌سازی موید نتایج نظری هستند. هم‌چنین نتایج نشان‌دهنده نرخ همگرایی سریع روش پیشنهادی برای مقادیر کوچک نمونه انتخاب‌شده در دسترس است. در بررسی داده واقعی نیز روش دومرحله‌ای بهبودیافته در مقایسه با روش دومرحله‌ای، عملکرد بسیار نزدیک در برآورد نقطه‌ای و فاصله‌ای برای مقادیر مختلف نمونه داشته است. با توجه به نتایج حاصل، این روش عملکردی نزدیک به روش اندازه نمونه ثابت بهینه دارد و هم‌چنین در عمل با روش دومرحله‌ای تفاوت چندانی ندارد. با توجه به سادگی استراتژی توقف تحت این روش که تفاوت زیادی در مقایسه با روش دومرحله‌ای ندارد، کاهش هزینه و افزایش دقت نسبت به روش دومرحله‌ای برای تعیین اندازه نمونه روشی بسیار مناسب است. هم‌چنین نتایج نشان می‌دهد که با افزایش دقت با ارائه قاعده توقف، کارایی نیز افزایش می‌یابد. کارا بودن روش پیشنهادی نسبت به اندازه نمونه ثابت بهینه و روش دومرحله‌ای و رفع محدودیت‌های تحت روش دومرحله‌ای از مزیت‌های این روش است. با توجه به کاربرد بسیار داده‌های سری زمانی به عنوان راهکاری برای بررسی ویژگی‌ها و پیش‌بینی داده‌ها به‌کارگیری این روش پیشنهاد می‌شود.

تقدیر و تشکر

نویسندگان مقاله بر خود لازم می‌دانند از سردبیر محترم، داوران گرامی، ویراستار مجله و دکتر حسین نادب که نظراتشان موجب بهبودی مقاله گردید، تشکر کنند.

مراجع

محمودی، ع.، لاله‌زاری، ر. و روغنی، ق. (۱۳۹۶)، برآورد دنباله ای پارامتر مقیاس توزیع نمایی با محدودیت کران داری تابع مخاطره، مجله علوم آماری، ۱، ۱۳۳-۱۴۸.

Anscombe, F. (1953), Sequential Estimation, *Journal of Royal Statistical Society*, **35**, 1-21.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, Wiley, New York.

Basawa, I. V., McCormick, W. P. and Sriram, T. N. (1990), Sequential Estimation for Dependent Observations with an Application to Non-Standard Autoregressive Processes, *Stochastic processes and their applications*, **35**, 149-168.

Basu, A. K. and Das, J. K. (1997), Sequential Estimation of the Autoregressive Parameters in Ar(p) Model, *Sequential Analysis*, **16**, 1-24.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series:Theory and Methods*, 2nd ed, Springer Texts in Statistics.

Chow, Y. S. and Robbins, H. E. (1965), On the Asymptotic Theory of Fixed Width Sequential Confidence Interval for Mean, *Annals of Mathematical Statistics*, **36**, 457-462.

Fakhre-Zakeri, I. and Lee, S. (1992), Sequential Estimation of the Mean of a Linear Process, *Sequential analysis*, **11**, 181-197.

Fakhre-Zakeri, I. and Lee, S. (1993), Sequential Estimation of the Mean Vector of a Multivariate Linear Process, *Multivariate Analysis*, **47**, 196-209.

Fourdrinier, D., Konev, V. and Pergamenshchikov, S. (2009), Truncated Sequential Estimation of the parameter of a First Order Autoregressive Process with Dependent Noises, *Mathematical Methods of Statistics*, **18**, 43-58.

Ghosh, M., Mukhopadhyay, N. and Sen, P. K. (1997), *Sequential Estimation*, New York: Wiley.

Gombay, E. (2010), Sequential Confidence Intervals for Time Series, *Periodica Mathematica Hungarica*, **61**, 183-193.

Jun, H. and Zhuang. Y. (2020), A broader Class of Modified Two-Stage Minimum Risk point Estimation Procedures for a Normal Mean, *Communications in Statistics-Simulation and Computation*, 1-15.

Karmakar, B. and Mukhopadhyay, I. (2018), Risk Efficient Estimation of Fully Dependent Random Coefficient Autoregressive Models of General Order, *Communications in Statistics-Theory and Methods*, **47**, 4242-4253.

Karmakar, B. and Mukhopadhyay, I. (2019), Risk-Efficient Sequential Estimation of Multivariate Random Coefficient Autoregressive Process, *Sequential Analysis*, **38**, 26-45.

Kashkovsky, D. V. and Konev, V. V. (2008), Sequential Estimates of the Parameters in a Random Coefficient Autoregressive Process, *Optoelectronics, Instrumentation and Data Processing*, **44**, 52-61.

Khalifeh, A., Mahmoudi, E. and Chaturvedi, A. (2020), Sequential Fixed-accuracy Confidence Intervals for the Stress–Strength Reliability Parameter for The Exponential Distribution:Two-Stage Sampling Procedure, *Computational Statistics*, **35**, 1553-1575.

Kusainov, M. I. (2015), Risk Effectiveness of Adaptive One-Step Predictions of Autoregression with a Noisy Parameter, *Bulletin of Tomsk State University*, **32**, 33-43.

Lai, T. L. (2014), Discussion on Sequential Estimation for Time Series Models by T. N. Sriram and Ross Laci, *Sequential Analysis*, **33**, 169-173.

Lee, S. (1994), Sequential Estimation for the Parameters of a Stationary Autoregressive Model, *Sequential Analysis*, **13**, 301-317.

Lee, S. (2003), The sequential Estimation in Stochastic Regression Model with Random Coefficients, *Statistics and Probability Letters*, **61**, 71–81.

Lee, S. and Sriram, T. N. (1999), Sequential Point Estimation of Parameters in a Threshold AR(1) Model, *Stochastic Processes and Their Applications*, **84**, 343-355.

Martinsek, A. T. (2001), Sequential Estimation of the Mean in a Random Coefficient Autoregressive Model with Beta Marginals, *Statistics and Probability Letters*, **51**, 53-61.

Mukhopadhyay, N. (1995), Sequential Estimation of Means of Linear Processes, *Metrika*, **42**, 279-290.

Mukhopadhyay, N. and Abid, A. D. (1999), Accelerated Sequential Sampling for Generalized Linear and Ar(p) Models, *Stochastic Modelling and Applications*, **2**, 1-15.

Mukhopadhyay, N. and Duggan, W. T. (1997), Can a Two-Stage Procedure Enjoy Second-Order Properties, *Sankhyā: The Indian Journal of Statistics, Series A*, **59**, 435-448.

Mukhopadhyay, N. and Duggan, W. T. (1999), On a Two-Stage Procedure Having Second-Order Properties with Applications. *Annals of the Institute of Statistical Mathematics*, **51**, 621-636.

Mukhopadhyay, N. and Sriram, T. N. (1992), On Sequential Comparisons of Means of First-Order Autoregressive Models. *Metrika*, **39**, 155-164.

Mukhopadhyay, N. (1980), A Consistent and Asymptotically Efficient Two-Stage Procedure to Construct Fixed Width Confidence Intervals for the Mean, *Metrika* , **27**, 281-284.

Robbins, H. E. (1959), Sequential Estimation of the Mean of a Normal Population, *in Probability and Statistics*, U. Grenander, ed., pp. 235-245, Uppsala: Almquist Wiksell.

Sajjadipanah, S., Mahmoudi. E. and Zamani. M. (2021), Two-Stage Procedure in a First-Order Autoregressive Process and Comparison with a Purely Sequential Procedure, *Sequential Analysis*, **40**, 466-481.

Sharafi, M. and Nematollahi, A. R. (2016), AR(1) Model with Skew-Normal Innovations, *Metrika*, **79**, 1011-1029.

SriramSriram, T. N. (1987), Sequential Estimation of the Mean of a First-Order Stationary Autoregressive Process, *The Annals of Statistics*, **15**, 1079-1090.

Sriram, T. N. (1988), Sequential Estimation of the Autoregressive Parameter in a First Order Autoregressive Process, *Sequential Analysis*, **7**, 53-74.

Sriram, T. N. (2001), Fixed Size Confidence Regions for Parameters of Threshold AR(1) Models, *Journal of Statistical Planning and Inference*, **97**, 293-304.

Sriram, T. N. and Laci, R. (2014), Sequential Estimator for Time Series Models, *Sequential Analysis*, **33**, 136-157.

Sriram, T. N. and Samadi, S. Y. (2019), Second-Order Analysis of Regret for Sequential Estimation of the Autoregressive Parameter in a First-Order Autoregressive Model, *Sequential Analysis*, **38**, 411-435.

Stein, C. (1945), A Two-Sample Test for a Linear Hypothesis Whose Power Is Independent of the Variance, *The Annals of Mathematical Statistics*, **16**, 243-258.

Stein, C. (1949), Some Problems in Sequential Estimation (Abstract), *Econometrica*, **17**, 77-78.