



Comparison of Neyman-Pearson and Evidential Approaches in Separate One-Sided Tests Based on Model Misspecification Error

Dastbaravarde, A. 

Department of Statistics, Yazd University, Yazd, Iran.

Corresponding author: A. Dastbaravarde, dastbaravarde@yazd.ac.ir

Received: 28/3/2024 Revised: 25/11/2024 Accepted and Published Online: 26/11/2024.

Introduction

The Neyman-Pearson hypothesis test is grounded in selecting between a null and alternative hypothesis. This approach leverages probability properties to select among different tests. Evidential inference, on the other hand, is rooted in the law of likelihood. In this approach, the extent to which data supports one model over another is determined by a fixed threshold, unlike the Neyman-Pearson approach, where the error probability sets the threshold. Evidential inference employs the likelihood ratio of hypotheses to measure the evidence in data supporting one composite hypothesis against another; the hypothesis with a higher maximum likelihood is considered better supported by the data. While Neyman-Pearson and evidential approaches offer distinct approaches to hypothesis testing, both are methods for inference and can be compared based on their probability of errors. A less explored issue in hypothesis testing is model misspecification error (which can be termed a Type III error). Model misspecification arises when neither the null nor the alternative hypothesis accurately describes the true data-generating model.

This research compares the Neyman-Pearson and evidential inferential approaches in separate one-sided hypothesis tests, focusing on error probabilities under model misspecification.

Material and Methods

This research employs the Kullback-Leibler divergence, a widely used metric for measuring the difference between distributions. Consequently, under

model misspecification, the preferred (or more accurate) hypothesis is the one that is closer to reality. This implies that the hypothesis should be chosen such that the Kullback-Leibler divergence between the models under that hypothesis and the true model is minimized. Under model misspecification, the probability of hypothesis test errors is redefined as the probability of true errors. Finally, based on the true error probability, the performance of both the Neyman-Pearson and evidential approaches is examined in separate one-sided hypothesis testing.

Results and Discussion

Results indicated that under model misspecification, the true Type II error probability of the Neyman-Pearson test approaches zero as the sample size increases; however, the true Type I error probability may approach one as the sample size increases. In other words, under model misspecification, the Neyman-Pearson test tends to reject the null hypothesis as the sample size grows, regardless of which hypothesis is closer to reality. On the other hand, results showed that in hypothesis testing based on the evidential approach, even under model misspecification, both the true probability of misleadingly strong evidence and the true probability of weak evidence approach zero as the sample size increases. In other words, under model misspecification, in evidential testing, the probability of strong support for the correct hypothesis approaches one as the sample size grows.

Conclusion

The findings of this research suggest that under model misspecification, the evidential approach outperforms the Neyman-Pearson approach in hypothesis testing.

Keywords: Evidential inference, Frequentist inference, Model misspecification, Composite hypothesis.

Mathematics Subject Classification (2010): 62A01, 62B10.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [\(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

مقایسه عملکرد دو رویکرد نیمین-پیرسونی و شواهدی در مساله آزمون فرضیه‌های یک طرفه جدا از هم براساس خطای بدمشخص سازی مدل

علی دست‌برآورده

بخش آمار، دانشکده علوم ریاضی، دانشگاه یزد

نویسنده مسئول: علی دست‌برآورده، dastbaravarde@yazd.ac.ir

تاریخ دریافت: ۱۴۰۳/۱/۹ تاریخ بازنگری: ۱۴۰۳/۹/۵ تاریخ پذیرش و انتشار: ۱۴۰۳/۹/۶

چکیده: در مساله آزمون فرضیه‌های آماری، خطای بدمشخص سازی مدل وضعیتی است که در آن توزیع واقعی داده‌ها هیچ‌کدام از توزیع‌های تحت فرضیه صفر یا فرضیه مقابل نیست. در این پژوهش، احتمال خطاهای بدمشخص سازی مدل برای آزمون فرضیه‌های مرکب از نوع یک طرفه مطالعه شده است. این خطاها در دو رویکرد استنباط آماری نیمین-پیرسونی و شواهدی مقایسه شده است. نتایج نشان می‌دهد که رویکرد شواهدی عملکرد بهتری نسبت به رویکرد نیمین-پیرسونی در این زمینه دارد. واژه‌های کلیدی: استنباط شواهدی، استنباط فراوانی‌گرا، بدمشخص سازی مدل، فرضیه‌های مرکب. کد موضوع بندی ریاضی (۲۰۱۰): 62A01, 62B10.

۱ مقدمه

در دهه ۱۹۳۰ میلادی رویکردی برای آزمون فرضیه‌های آماری با پایه‌گذاری نیمین و پیرسون (۱۹۳۳) معرفی شد که مبتنی بر انتخاب از بین دو فرضیه صفر و فرضیه مقابل بود. در این رویکرد از ویژگی‌های احتمالی (احتمال خطای نوع اول و خطای نوع دوم با نماد به ترتیب α و β) برای انتخاب از بین آزمون‌های مختلف استفاده می‌شود. لم اساسی نیمین-پیرسون نشان می‌دهد که چگونه، به ازای هر مقدار ثابت α ، بهترین آزمون، یعنی با کمترین مقدار β ، را بیابیم. نظریه آزمون فرضیه نیمین-پیرسون با تمرکز عمل‌گرایانه و جذاب آن بر کمیته سازی احتمال‌های ارتکاب خطا، روش علمی توانمندی ایجاد کرده که نظریه آمار در دوران معاصر را تحت سلطه خود قرار داده است



۲۵۶ مقایسه دو رویکرد نیمن-پیرسونی و شواهدی در آزمون فرضیه‌های یک طرفه جدا از هم

(ارقامی و طاهری، ۱۳۹۸). کارلین و روبین (۱۹۵۶) رویکرد نیمن-پیرسون را برای انجام آزمون فرضیه‌های مرکب از نوع یک طرفه تعمیم دادند.

در اواخر قرن بیستم میلادی، ریچارد رویال (۱۹۹۷) براساس پژوهش‌های هکینگ (۱۹۶۵) و ادواردز (۱۹۷۲)، چارچوب رسمی برای استنباط مبتنی بر شواهد در آمار را عرضه کرد و خطاهای این رویکرد استنباطی را مورد بررسی قرار داد. چارچوب استنباط شواهدی مشابه نیمن-پیرسون براساس نسبت درستنمایی پایه‌ریزی شده است. اما نتیجه‌گیری در مورد اینکه داده‌ها (شواهد) از کدام مدل بیشتر پشتیبانی می‌کنند مبتنی بر آستانه‌ای ثابت است؛ برخلاف رویکرد نیمن-پیرسون که آستانه بوسیله احتمال خطا تعیین می‌شود. ایده اصلی نتیجه‌گیری در رویکرد شواهدی عبارت است از وجود شواهد قوی به نفع مدل تحت H_1 وقتی L_1 (میزان درستنمایی تحت H_1)، k برابر L_2 (میزان درستنمایی تحت H_2) و شواهد قوی به نفع مدل تحت H_2 وقتی L_2 ، k برابر L_1 است که k مقداری ثابت و بزرگتر از یک است. در این رویکرد از نسبت ماکسیمم درستنمایی فرضیه‌ها برای سنجش میزان شواهد موجود در داده‌ها در پشتیبانی از یک فرضیه مرکب در مقابل فرضیه مرکب دیگر استفاده می‌شود؛ به‌طوری که داده‌ها از فرضیه‌ای بیشتر پشتیبانی می‌کنند که مقدار ماکسیمم درستنمایی تحت آن فرضیه بیشتر از مقدار متناظر تحت فرضیه رقیب باشد (روبال، ۱۹۹۲، ۱۹۹۷). برای آشنایی بیشتر با استنباط شواهدی به عمادی (۱۳۸۴)، ارقامی و همکاران (۱۳۸۷الف، ب)، دست‌برآورده (۱۳۸۸)، اسماعیلی و همکاران (۱۳۹۰)، هاشم‌پور (۱۳۹۵)، ارقامی و طاهری (۱۳۹۸) و طاهری (۱۳۹۸) مراجعه شود.

دیدگاه‌های نیمن-پیرسونی و شواهدی رویکردهای متفاوتی در انجام آزمون فرضیه‌های آماری دارند. بنابراین مقایسه آنها دشوار است. با این وجود، آنها روش‌هایی برای استنباط هستند و یک ابزار اساسی که بوسیله آن می‌توان آنها را مقایسه کرد احتمال خطا در استنباط است. از طرفی، هر دو رویکرد استنباطی بر پایه نسبت درستنمایی استوارند؛ از این‌رو، با مطالعه رفتار نسبت درستنمایی می‌توان به سوال‌های مربوط به احتمال خطاها در این دو رویکرد پاسخ داد.

یک مساله که در آزمون فرضیه‌ها کمتر به آن توجه شده است خطای بدمشخص‌سازی مدل^۱ است، که آن را خطای نوع سوم می‌توان نامید. خطای بدمشخص‌سازی مدل وضعیتی را مطرح می‌کند که در آن نه مدل تحت فرضیه صفر و نه مدل تحت فرضیه مقابل، مدل واقعی که داده‌ها از آن آمده‌اند را به‌طور دقیق شرح نمی‌دهند. بدمشخص‌سازی می‌تواند دلایل زیادی داشته باشد؛ از جمله: تشخیص نادرست فرم تابعی مدل، حذف متغیرهای کمکی واقعی از مدل، گنجاندن متغیرهای کمکی جعلی در مدل و تشخیص نادرست ساختار خطای اندازه‌گیری (ماستلر، ۱۹۴۸).

دنيس و همکاران (۲۰۱۹) به مقایسه دو رویکرد استنباطی نیمن-پیرسونی و شواهدی در مساله آزمون فرضیه‌های ساده در مقابل ساده براساس احتمال خطاهای بدمشخص‌سازی مدل پرداختند و به این نتیجه رسیدند که رویکرد شواهدی عملکرد بهتری نسبت به رویکرد نیمن-پیرسونی دارد. در این نوشتار قصد داریم دامنه پژوهش را گسترش داده و موضوع موردنظر را در مساله آزمون فرضیه‌های مرکب از نوع یک طرفه جدا از هم نیز مورد بررسی قرار دهیم. از این‌رو، ساختار نوشتار حاضر به این صورت است که: در بخش ۲، روش مقایسه دو رویکرد بیان شده است. در

¹model misspecification error

بخش ۳، احتمال خطای استنباط در دو رویکرد در حالت بدمشخص‌سازی مدل، محاسبه و مقایسه شده است. در نهایت جمع‌بندی و نتیجه‌گیری از این مطالعه ارائه شده است.

۲ روش تحلیل

فرض کنید $\mathcal{P} = \{f_\theta : \theta \in \Theta \subset \mathcal{R}\}$ خانواده‌ای تک پارامتری از توابع جرم یا چگالی احتمال باشد که نمونه تصادفی $\mathbf{x} = (x_1, \dots, x_n)^T$ از آن تولید شده است. همچنین، فرض کنید $\theta^* \in \Theta$ مقدار واقعی پارامتر باشد. تابع درستنمایی برای نمونه مشاهده‌شده $\mathbf{x}$ با $f_\theta(x_i)$ $L(\theta) = \prod_{i=1}^n f_\theta(x_i)$ نمایش داده می‌شود. در این پژوهش برای اندازه‌گیری تفاوت بین دو تابع f_θ از f_{θ^*} از معیار واگرایی کولبک-لایبلر (کولبک، ۱۹۹۷) به صورت

$$K(f_{\theta^*} \| f_\theta) = E_{\theta^*} \left[\log \left(\frac{f_{\theta^*}(X)}{f_\theta(X)} \right) \right] = \int \log \left(\frac{f_{\theta^*}(x)}{f_\theta(x)} \right) dF_{\theta^*}(x),$$

استفاده می‌شود، که در آن F_{θ^*} ، تابع توزیع متناظر با f_{θ^*} و E_{θ^*} امید ریاضی بر حسب F_{θ^*} است. مقدار این معیار همواره نامنفی است و برابر صفر است اگر و فقط اگر f_θ و f_{θ^*} تقریباً همه جا یکسان باشند.

تعریف ۱. (دنيس و همکاران، ۲۰۱۹). فرض کنید مدل واقعی جامعه f_{θ^*} باشد. مدل f_θ را برای f_{θ^*} بدمشخص‌سازی شده گوییم، هرگاه $K(f_{\theta^*} \| f_\theta) > 0$.

بنابراین، از معیار کولبک-لایبلر می‌توان به‌عنوان ابزاری برای سنجش و مقایسه میزان بدمشخص‌سازی توزیع‌های تقریبی از توزیع واقعی بهره برد؛ به‌طوری‌که هرچه مقدار معیار کولبک-لایبلر کوچک‌تر باشد، میزان بدمشخص‌سازی مدل کمتر است. از این‌رو، از بین توزیع‌های تقریبی رقیب، توزیعی که دارای کمترین مقدار واگرایی کولبک-لایبلر از توزیع واقعی باشد ارجح است. فرض کنید H_j ، $j = 1, 2$ ، بیانگر این فرضیه باشد که داده‌ها از توزیعی با تابع چگالی احتمال f_{θ_j} تولید شده‌اند. همچنین، در نظر بگیرید $\Delta K = K(f_{\theta^*} \| f_{\theta_1}) - K(f_{\theta^*} \| f_{\theta_2})$. از آنجا که ΔK می‌تواند مثبت، منفی یا صفر باشد؛ اگر $\Delta K > (<)$ ، آنگاه f_{θ_1} به تابع چگالی احتمال واقعی نزدیک‌تر و فرضیه H_1 (H_2) نسبت به فرضیه H_2 (H_1) ارجح است و اگر $\Delta K = 0$ ، آنگاه هر دو تابع چگالی احتمال به یک اندازه به تابع چگالی احتمال واقعی نزدیک هستند، یا به تعبیر دیگر هر دو فرضیه H_1 و H_2 میزان اطلاعات یکسانی از واقعیت دارند و هیچ کدام بر دیگری ارجحیت ندارد.

تعریف ۲. اگر f_{θ^*} تابع چگالی احتمال واقعی جامعه باشد، کلاس توابع چگالی $\mathcal{P}_1 = \{f_\theta : \theta \in \Theta_1 \subseteq \Theta\}$ را برای f_{θ^*} بدمشخص‌سازی شده گوییم، هرگاه $K(f_{\theta^*} \| \mathcal{P}_1) := \inf_{\theta \in \Theta_1} K(f_{\theta^*} \| f_\theta) > 0$.

به تعبیری، اگر همه اعضای کلاسی از توابع چگالی احتمال برای تابع چگالی احتمال واقعی جامعه بدمشخص‌سازی شده باشند، آنگاه آن کلاس نیز برای مدل واقعی جامعه بدمشخص‌سازی شده است. همچنین، در صورتی که مقدار کولبک-لایبلر دو تابع چگالی احتمال را میزان دوری آنها تعبیر کنیم، معقول است که مقدار کولبک-لایبلر مجموعه‌ای از توابع

چگالی احتمال با تابع چگالی احتمالی مشخص را میزان دور بودن نزدیک‌ترین عضو آن مجموعه از آن تابع چگالی احتمال مشخص در نظر بگیریم.

تذکر ۱. با توجه به تعریف ۲، اگر $\theta^* \in \Theta_1$ ، آنگاه $K(f_{\theta^*} \| \mathcal{P}_1) = 0$.

لم ۱. فرض کنید f_{θ^*} تابع چگالی احتمال واقعی جامعه و $\mathcal{P}_1 = \{f_{\theta} : \theta \in \Theta_1 \subseteq \Theta\}$ کلاسی از توابع چگالی احتمال است. همچنین، فرض کنید θ^* نقطه مرزی^۱ Θ_1 باشد. اگر دنباله‌ای مانند $\{\theta_n\}_{n=1}^{\infty} \subset \Theta_1$ وجود داشته که f_{θ_n} به طور یکنواخت به f_{θ^*} همگرا باشد، آنگاه $K(f_{\theta^*} \| \mathcal{P}_1) = 0$.

برهان: از مساله ۴.۱۷ صفحه ۷۹ کولیک (۱۹۹۷) نتیجه می‌شود $\lim_{n \rightarrow \infty} K(f_{\theta^*} \| f_{\theta_n}) = 0$ و بنابراین نتیجه مطلوب به دست می‌آید.

با توجه به لم ۱ نتیجه می‌شود که اگر $\theta^* \notin \bar{\Theta}_1$ ، که $\bar{\Theta}_1$ بستار^۲ Θ_1 است، آنگاه $K(f_{\theta^*} \| \mathcal{P}_1) > 0$.

تعریف ۳. برای فرضیه مرکب $\theta^* \in \Theta_1$ در مقابل $\theta^* \in \Theta_2$ ، H_2 ، گوئیم فرضیه H_1 (H_2) ارجح (صحیح‌تر) از فرضیه H_2 (H_1) است، هرگاه $0 < K(f_{\theta^*} \| \mathcal{P}_2) - K(f_{\theta^*} \| \mathcal{P}_1) = \Delta K$ ، که در آن $\Delta K = 0$ اگر $j = 1, 2$ ، $\mathcal{P}_j = \{f_{\theta} : \theta \in \Theta_j \subseteq \Theta\}$ ، آنگاه دو فرضیه ارجحیت یکسان دارند.

به بیان دیگر، کلاس توزیع‌های تقریبی‌ای ارجح است که توزیع تقریبی نزدیک‌تری به توزیع واقعی نسبت به کلاس دیگر ارائه می‌دهد.

تعریف ۴. در مسئله آزمون فرضیه $\theta^* \in \Theta_1$ در مقابل $\theta^* \in \Theta_2$ ، H_2 ، آزمون را بدمشخص‌سازی‌شده گوئیم هرگاه $K(f_{\theta^*} \| \mathcal{P}_1) > 0$ و $K(f_{\theta^*} \| \mathcal{P}_2) > 0$.

با توجه به تعریف ۴، مساله آزمون فرضیه، بدمشخص‌سازی‌شده است هرگاه هیچ‌کدام از دو فرضیه گزاره صحیحی به لحاظ منطقی نباشند. اما، از آنجایی که در مسائل کاربردی توزیع واقعی معلوم نیست، انتظار منطقی چنین است که مشاهدات از فرضیه‌ای که به واقعیت نزدیک‌تر است بیشتر پشتیبانی کند. از دیدگاه نظریه تصمیم نیز این مطلب را می‌توان چنین بیان کرد که مشاهدات، منجر به انتخاب فرضیه‌ای شود که به واقعیت نزدیک‌تر باشد.

تذکر ۲. براساس تعریف ۴، تذکر ۱ و لم ۱ می‌توان نتیجه گرفت، اگر $\theta^* \in (\bar{\Theta}_1 \cup \bar{\Theta}_2)^c$ ، که در آن Θ^c متمم مجموعه Θ است، آنگاه مساله آزمون فرضیه $\theta^* \in \Theta_1$ در مقابل $\theta^* \in \Theta_2$ ، H_2 ، بدمشخص‌سازی‌شده است.

مثال ۱. با توجه به تذکر ۲، اگر $\theta^* \in (\theta_1, \theta_2)$ ، آنگاه آزمون فرضیه‌های یک طرفه جدا از هم $H_1 : \theta^* \leq \theta_1$ در مقابل $H_2 : \theta^* > \theta_2$ ، H_2 ، چهار مساله بدمشخص‌سازی مدل است.

¹boundary point

²closure

لم ۲. فرض کنید $\hat{\theta}$ برآوردگر ماکسیمم درستنمایی و سازگار برای θ^* است. اگر θ^* نقطه درونی^۱ $\Theta_1 \subset \Theta$ باشد، آنگاه $\lim_{n \rightarrow \infty} P_{\theta^*}(\hat{\theta} \in \Theta_1) = 1$ که در آن P_{θ^*} احتمال برحسب توزیع واقعی F_{θ^*} است.

برهان: از آنجا که θ^* نقطه درونی Θ_1 است، همسایگی‌ای حول θ^* وجود دارد به‌طوری‌که تماماً داخل Θ_1 است؛ یعنی، $\varepsilon_0 > 0$ وجود دارد که $(\theta^* - \varepsilon_0, \theta^* + \varepsilon_0) \subseteq \Theta_1$. کافی است قرار دهیم $\varepsilon_0 \leq \varepsilon$. در این صورت $\{\hat{\theta} : |\hat{\theta} - \theta^*| \leq \varepsilon\} \subseteq \{\hat{\theta} : \hat{\theta} \in \Theta_1\}$. پس، $P_{\theta^*}(|\hat{\theta} - \theta^*| \leq \varepsilon) \leq P_{\theta^*}(\hat{\theta} \in \Theta_1)$. در نتیجه، بنا به سازگاری $\hat{\theta}$ ، $\lim_{n \rightarrow \infty} P_{\theta^*}(\hat{\theta} \in \Theta_1) = 1$.

۳ احتمال خطا تحت بدمشخص‌سازی مدل

باکس (۱۹۷۹) اظهار کرده‌اند که ”همه مدل‌ها اشتباه هستند، اما برخی از آنها مفیدند“. مساله اساسی برای یک رویکرد استنباطی در انتخاب نزدیک‌ترین تابع چگالی احتمال از بین دو تابع چگالی احتمال f_{θ_1} و f_{θ_2} به تابع چگالی احتمال واقعی f_{θ^*} زمانی‌که هر دو با تابع چگالی احتمال واقعی اختلاف دارند این است که چقدر خوب و کم اشتباه عمل می‌کند. در این بخش، دو رویکرد نیمن-پیرسونی و شواهدی را مقایسه می‌کنیم تا دریابیم چه میزان تحلیل‌ها براساس این دو رویکرد ممکن است به اشتباه انجام شود.

۳.۱ آزمون فرضیه نیمن-پیرسونی

فرض کنید کلاس توزیع‌های $\mathcal{P} = \{f_{\theta} : \theta \in \Theta \subseteq (-\infty, \theta_1] \cup [\theta_2, \infty)\}$ دارای خاصیت نسبت درستنمایی یکنوا^۲ (MLR) در $T(\mathbf{x})$ باشد. در این صورت با استدلالی همانند قضیه کارلین-روبین آزمون $\psi(\mathbf{x}) = I(T(\mathbf{x}) > c)$ که در آن I تابع نشانگر است و نقطه بحرانی c به‌گونه‌ای تعیین می‌شود که اندازه آزمون برابر مقدار ثابت و مشخص α باشد، به‌طور یکنواخت پرتوان‌ترین^۳ (UMP) آزمون در سطح α برای آزمون $H_1 : \theta^* \leq \theta_1$ در مقابل $H_2 : \theta^* \geq \theta_2$ است. با توجه به خاصیت MLR آزمون $\psi(\mathbf{x})$ را می‌توان به‌صورت

$$\psi(\mathbf{x}) = \begin{cases} 1 & \frac{L(\theta_2)}{L(\theta_1)} > c', \\ 0 & \frac{L(\theta_2)}{L(\theta_1)} < c', \end{cases} \quad (1)$$

بازنویسی کرد (لی‌من و رومانو، ۲۰۰۵)، که در آن c' به‌گونه‌ای تعیین می‌شود که $\alpha = P_{\theta_1}(\frac{L(\theta_2)}{L(\theta_1)} > c') = \alpha$. از آنجا که $\log(\frac{L(\theta_1)}{L(\theta_2)}) = \sum_{i=1}^n \log(\frac{f_{\theta_1}(X_i)}{f_{\theta_2}(X_i)})$ بنا به قضیه حد مرکزی، $\log(\frac{L(\theta_1)}{L(\theta_2)})$ وقتی H_1 صحیح است دارای توزیع مجانبی $N(nK_{12}, n\sigma_1^2)$ است، که در آن $K_{12} = K(f_{\theta_1} \| f_{\theta_2})$ ، $\sigma_1^2 = V_{\theta_1}[\log(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)})]$.

^۱interior point

^۲Monotone Likelihood Ratio

^۳Uniformly Most Powerful

۲۶۰ مقایسه دو رویکرد نیمن-پیرسونی و شواهدی در آزمون فرضیه‌های یک طرفه جدا از هم

است. از این‌رو، مقدار c' برای یک اندازه نمونه بزرگ تقریباً برابر است با $c' \approx \exp[\sqrt{n}\sigma_1 Z_{1-\alpha} - nK_{12}]$ که در آن $Z_{1-\alpha}$ چندک $(1-\alpha)$ -ام توزیع نرمال استاندارد است. از طرفی، $\log(\frac{L(\theta_1)}{L(\theta_2)})$ وقتی H_2 صحیح است دارای توزیع مجانبی $N(-nK_{12}, n\sigma_2^2)$ است که در آن $K_{12} = K(f_{\theta_2}|f_{\theta_1})$ ، $\sigma_2^2 = V_{\theta_2}[\log(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)})]$. از این‌رو، حداکثر احتمال خطای نوع دوم آزمون برای یک اندازه نمونه بزرگ در رابطه $\beta \approx \Phi(\frac{\sigma_1}{\sigma_2} Z_{1-\alpha} - \frac{\sqrt{n}}{\sigma_2}(K_{12} + K_{21}))$ صدق می‌کند (دنيس و همکاران، ۲۰۱۹). همان‌طور که مشاهده می‌شود با افزایش اندازه نمونه احتمال خطای نوع دوم به صفر میل می‌کند؛ این در حالی است که خطای نوع اول ثابت است. این مساله اساس برخی انتقادهای از آزمون فرضیه نیمن-پیرسونی است (مایو و اسپانوس، ۲۰۰۶؛ مایکلیدس، ۲۰۲۱).

تعریف ۵ (خطاهای نوع اول و دوم واقعی). فرض کنید قصد داریم از بین دو فرضیه $\theta^* \in \Theta_1$ در مقابل $\theta^* \in \Theta_2$ یکی را انتخاب کنیم. در صورت انتخاب خانواده $\mathcal{P}_2 = \{f_\theta : \theta \in \Theta_2\}$ ، وقتی خانواده $\mathcal{P}_1 = \{f_\theta : \theta \in \Theta_1\}$ به تابع چگالی احتمال واقعی f_{θ^*} نزدیک‌تر است (یعنی، $\Delta K > 0$) مرتکب خطای نوع اول واقعی شده‌ایم. احتمال خطای نوع اول واقعی، برای آزمون نیمن-پیرسونی $\psi(\mathbf{x})$ وقتی $\Delta K > 0$ به‌صورت $\alpha' = E_{\theta^*}(\psi(\mathbf{X}))$ تعریف می‌شود. همچنین، در صورت انتخاب خانواده $\mathcal{P}_1$ ، وقتی خانواده $\mathcal{P}_2$ به تابع چگالی احتمال واقعی f_{θ^*} نزدیک‌تر است (یعنی، $\Delta K < 0$) مرتکب خطای نوع دوم واقعی شده‌ایم. احتمال خطای نوع دوم واقعی، که با β' نشان داده می‌شود، وقتی $\Delta K < 0$ به‌صورت $\beta' = 1 - E_{\theta^*}(\psi(\mathbf{X}))$ تعریف می‌شود.

تذکر ۳. در تعریف ۱، اگر $\theta^* \in \Theta_1$ آنگاه، $\Delta k > 0$ و در نتیجه $\alpha' \leq \alpha$. همچنین، به‌طور مشابه، اگر $\theta^* \in \Theta_2$ آنگاه، $\Delta k < 0$ و در نتیجه $\beta' \leq \beta$.

قضیه ۱. فرض کنید خانواده توزیع‌های تک پارامتری $\mathcal{P} = \{f_\theta : \theta \in (-\infty, \theta_1] \cup [\theta_2, \infty)\}$ دارای خاصیت MLR باشد. در مساله آزمون فرضیه $H_1 : \theta^* \leq \theta_1$ در مقابل $H_2 : \theta^* \geq \theta_2$ که $\theta_1 < \theta_2$ ، برای یک اندازه نمونه بزرگ، مقدار تقریبی احتمال خطای نوع اول و دوم واقعی آزمون در رابطه (۱) زمانی که $\theta^* \in (\theta_1, \theta_2)$ را می‌توان چنین محاسبه کرد:

$$\alpha' \approx \Phi\left(\frac{1}{\sigma_{12}}[-\sigma_1 Z_{1-\alpha} + \sqrt{n}(K_{12} - \Delta K)]\right), \quad \text{وقتی } \Delta K > 0, \quad (2)$$

$$\beta' \approx \Phi\left(\frac{1}{\sigma_{12}}[\sigma_1 Z_{1-\alpha} - \sqrt{n}(K_{12} - \Delta K)]\right), \quad \text{وقتی } \Delta K < 0. \quad (3)$$

برهان: دنيس و همکاران (۲۰۱۹) نشان دادند برای n بزرگ $\log(\frac{L(\theta_1)}{L(\theta_2)})$ دارای توزیع مجانبی $N(n\Delta K, n\sigma_{12}^2)$ است که در آن $\sigma_{12}^2 = V_{\theta^*}[\log(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)})]$. در نتیجه وقتی $\Delta K > 0$ ، $\alpha' = P_{\theta^*}(\frac{L(\theta_1)}{L(\theta_2)} > c') \approx \Phi(\frac{\log c' - n\Delta K}{\sqrt{n\sigma_{12}^2}})$ و وقتی $\Delta K < 0$ ، $\beta' = P_{\theta^*}(\frac{L(\theta_1)}{L(\theta_2)} \leq c') \approx \Phi(\frac{\log c' + n\Delta K}{\sqrt{n\sigma_{12}^2}})$. مقدار c' نتیجه مطلوب به‌دست می‌آید.

از رابطه (۲) نتیجه می‌شود که احتمال خطای نوع اول واقعی تحت بدمشخص‌سازی مدل لزوماً برابر با مقدار اسمی آن نیست (یعنی $\alpha' \neq \alpha$) و زمانی که $\theta_1 = \theta^*$ ، تساوی برقرار است. در واقع، بسته به ترکیب توابع چگالی احتمال f_{θ_1} ، f_{θ^*} و f_{θ_2} ، مقدار احتمال خطای نوع اول واقعی ممکن است بزرگتر، مساوی یا حتی کمتر از مقدار اسمی آن باشد. با توجه به رابطه (۲)، اگر $K_{12} - \Delta K > 0$ ، آنگاه با افزایش n ، α' به مقدار یک میل می‌کند. به عبارت دیگر، تحت شرایط $\Delta K > 0$ و $K_{12} - \Delta K > 0$ ، با افزایش اندازه نمونه، آزمون فرضیه نیمن-پیرسونی به اشتباه تمایل به رد فرضیه منطقی‌تر H_1 دارد.

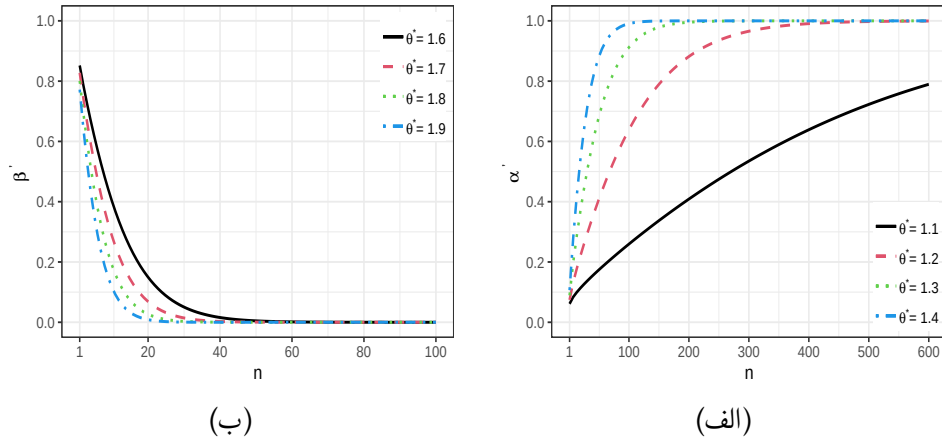
از طرفی، از رابطه (۳) نتیجه می‌شود که احتمال خطای نوع دوم واقعی تحت بدمشخص‌سازی متفاوت با مقدار مورد انتظارش است (یعنی $\beta' \neq \beta$) و زمانی که $\theta_2 = \theta^*$ ، تساوی برقرار است. همچنین با افزایش اندازه نمونه، β' به صفر میل می‌کند؛ زیرا $\Delta K < 0$ و در نتیجه $K_{12} - \Delta K > 0$. به عبارت دیگر، تحت شرط $\Delta K < 0$ ، با افزایش اندازه نمونه، آزمون فرضیه نیمن-پیرسونی به درستی تمایل به پذیرش فرضیه منطقی‌تر H_2 دارد. در نتیجه، در مساله آزمون فرضیه‌های یک‌طرفه جدا از هم و در شرایط بدمشخص‌سازی مدل، آزمون نیمن-پیرسونی با افزایش اندازه نمونه متمایل به رد فرضیه صفر (صحیح یا ناصحیح) است.

مثال ۲. فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی به اندازه n از توزیع نرمال با میانگین واقعی θ^* و واریانس معلوم σ^2 است. برای آزمون فرضیه $H_1: \theta^* \leq \theta_1$ در مقابل $H_2: \theta^* \geq \theta_2$ ، که در آن $\theta_1 < \theta_2$ است، فرض کنید $\theta^* \in (\theta_1, \theta_2)$ ؛ در این صورت با مساله بدمشخص‌سازی مدل روبرو هستیم. در نظر می‌گیریم f_θ نشانگر تابع چگالی نرمال با میانگین و واریانس به ترتیب θ و σ^2 است. همچنین، در نظر می‌گیریم $\mathcal{P}_1 = \{f_\theta: \theta \leq \theta_1\}$ ، $\mathcal{P}_2 = \{f_\theta: \theta \geq \theta_2\}$ و $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$. برای هر $\theta \in \mathcal{R}$ داریم $K(f_{\theta^*} \| f_\theta) = \frac{(\theta^* - \theta)^2}{2\sigma^2}$. توجه کنید که در اینجا معیار کولبک-لایبلر تابعی صعودی از متر اقلیدسی $|\theta^* - \theta|$ است. برای $\theta^* \in (\theta_1, \theta_2)$ داریم $\Delta K = \frac{(\theta^* - \theta_2)^2 - (\theta^* - \theta_1)^2}{2\sigma^2}$. پس، $\Delta K > 0$ اگر و فقط اگر $|\theta^* - \theta_1| > |\theta^* - \theta_2|$. این بدان معنی است که فرضیه‌ای ارجح است که فاصله کمتری از مدل واقعی دارد. بنابراین، در حالت بدمشخص‌سازی مدل، اگر $\bar{\theta} = \frac{\theta_1 + \theta_2}{2}$ ، آنگاه فرضیه H_1 ارجح است؛ اما اگر $\theta^* > \bar{\theta}$ ، آنگاه فرضیه H_2 ارجح است. همچنین، اگر $\theta^* = \bar{\theta}$ ، آنگاه هر دو فرضیه دارای ارجحیت یکسان هستند. برای فرضیه H_1 در مقابل H_2 ، آزمون UMP در سطح α به صورت

$$\psi(\mathbf{x}) = \begin{cases} 1 & \bar{x} \geq \theta_1 + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha}, \\ 0 & \bar{x} < \theta_1 + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha}. \end{cases} \quad (4)$$

به دست می‌آید. احتمال خطاهای نوع اول و دوم واقعی این آزمون برای $\theta_1 < \theta^* < \bar{\theta}$ برابر $\alpha' = \Phi(\sqrt{n} \frac{\theta^* - \theta_1}{\sigma} - Z_{1-\alpha})$ و برای $\theta^* < \theta_2 < \bar{\theta}$ برابر $\beta' = \Phi(-\sqrt{n} \frac{\theta^* - \theta_1}{\sigma} + Z_{1-\alpha})$ است. α' و β' توابعی به ترتیب صعودی و نزولی از اندازه نمونه هستند و زمانی که $n \rightarrow \infty$ ، به ترتیب به یک و صفر میل می‌کنند.

در شکل ۱ روند تغییر احتمال‌های واقعی خطاهای نوع اول و دوم آزمون UMP در رابطه (۴) را برحسب اندازه



شکل ۱. احتمال خطای واقعی الف: نوع اول و ب: نوع دوم، مثال ۲

نمونه به ازای $\theta_1 = 1, \theta_2 = 2, \sigma = 1, \alpha = 0.05$ نشان می‌دهد. برای $\theta^* \in \{1/1, 1/2, 1/3, 1/4\}$ فرضیه H_1 ارجح از H_2 است، اما همان‌طور که در شکل ۱-الف مشاهده می‌شود با افزایش n ، α' نیز افزایش می‌یابد! و هر چه θ^* به $\bar{\theta}$ نزدیک‌تر باشد سریع‌تر (برای اندازه نمونه کمتری) به یک میل می‌کند؛ به‌طوری که، برای مقادیر $1/1, 1/2, 1/3, 1/4$ از میانگین واقعی، مقدار خطای نوع اول واقعی آزمون برای اندازه‌های نمونه به‌ترتیب ۶۱۹، ۱۵۵، ۶۹ و ۳۹ به میزان ۸۰ درصد می‌رسد. این در حالی است که به‌عنوان مثال برای $\theta^* = 1/1$ ، مقدار میانگین واقعی به نسبت ۱ به ۹ به H_1 نزدیک‌تر از H_2 است. برای $\theta^* \in \{1/6, 1/7, 1/8, 1/9\}$ ، فرضیه H_2 ارجح از H_1 است و همان‌طور که در شکل ۱-ب مشاهده می‌شود مطابق انتظار با افزایش n ، β' به‌سرعت به صفر میل می‌کند.

مثال ۳. فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی به اندازه n از توزیع پواسون با میانگین واقعی θ^* است. برای آزمون فرضیه $H_1: \theta^* \leq \theta_1$ در مقابل $H_2: \theta^* \geq \theta_2$ که $\theta_1 < \theta_2$ ، فرض کنید $\theta^* \in (\theta_1, \theta_2)$ ؛ در این صورت با مساله بدمشخص‌سازی مدل روبرو هستیم. با فرض آنکه f_θ تابع جرم احتمال پواسون با میانگین θ باشد، قرار می‌دهیم $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ و $\mathcal{P}_1 = \{f_\theta: 0 < \theta \leq \theta_1\}$ ، $\mathcal{P}_2 = \{f_\theta: \theta \geq \theta_2\}$ ، برای هر $\theta \in (0, \infty)$ داریم $K(f_{\theta^*} \| f_\theta) = \theta^* \log \frac{\theta^*}{\theta} + \theta - \theta^*$. برای $\theta^* \in (\theta_1, \theta_2)$ داریم $\Delta K = \theta^* \log \frac{\theta_1}{\theta_2} + \theta_2 - \theta_1$. پس، اگر $\Delta K > 0$ اگر و فقط اگر $\theta^* < \theta^\dagger$ که در آن $\theta^\dagger = \frac{\theta_2 - \theta_1}{\log \theta_2 - \log \theta_1}$. بنابراین، در حالت بدمشخص‌سازی مدل، اگر $\theta^\dagger < \theta^*$ ، آنگاه فرضیه H_1 ارجح است؛ اما اگر $\theta^* > \theta^\dagger$ ، آنگاه فرضیه H_2 ارجح است. همچنین، اگر $\theta^* = \bar{\theta}$ ، آنگاه هر دو فرضیه دارای ارجحیت یکسان هستند. برای فرضیه H_1 در مقابل H_2 ، آزمون UMP در

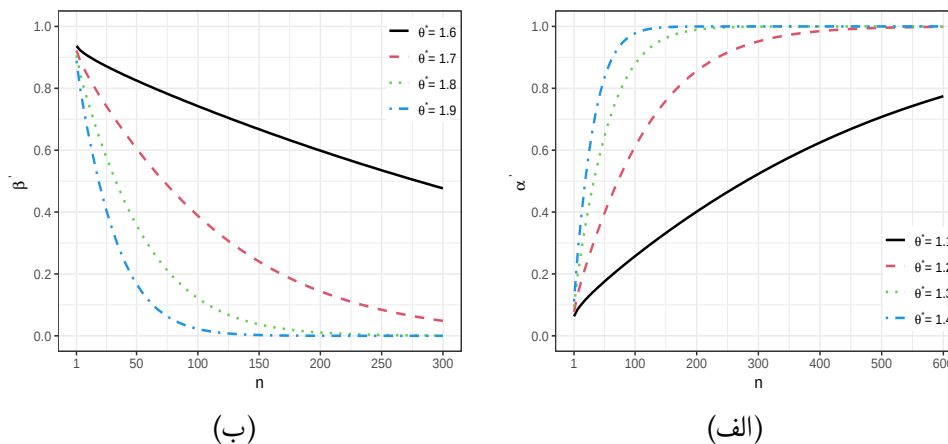
اندازه α به صورت

$$\psi(\mathbf{x}) = \begin{cases} 1 & \sum_{i=1}^n X_i > c, \\ \gamma & \sum_{i=1}^n X_i = c, \\ 0 & \sum_{i=1}^n X_i < c, \end{cases} \quad (5)$$

به دست می‌آید، که در آن c و γ به گونه‌ای تعیین می‌شوند که

$$P_{\theta_1}(\sum_{i=1}^n X_i > c) + \gamma P_{\theta_1}(\sum_{i=1}^n X_i = c) = \alpha.$$

در واقع، c چندک $(1 - \alpha)$ ام توزیع پواسون با میانگین $n\theta_1$ است و $\gamma = \frac{F_{n\theta_1}(c) - (1 - \alpha)}{f_{n\theta_1}(c)}$ برای n بزرگ، از قضیه حد مرکزی داریم $c \approx \sqrt{n\theta_1} Z_{1-\alpha} + n\theta_1$ و $\gamma P_{\theta_1}(\sum_{i=1}^n X_i = c) \approx 0$. اکنون احتمال خطای نوع اول و دوم واقعی این آزمون را محاسبه می‌کنیم. برای $\theta_1 < \theta^* < \theta^\dagger$ ، $\alpha' \approx \Phi(\sqrt{n} \frac{\theta^* - \theta_1}{\sqrt{\theta^*}} - Z_{1-\alpha} \sqrt{\frac{\theta_1}{\theta^*}})$ و برای $\theta^\dagger < \theta^* < \theta_2$ ، $\beta' \approx \Phi(\sqrt{n} \frac{\theta_1 - \theta^*}{\sqrt{\theta^*}} + Z_{1-\alpha} \sqrt{\frac{\theta_1}{\theta^*}})$ ، پس، β' و α' توابعی به ترتیب صعودی و نزولی از اندازه نمونه هستند و زمانی که $n \rightarrow \infty$ ، به ترتیب به یک و صفر میل می‌کنند.



شکل ۲. احتمال خطای واقعی الف: نوع اول و ب: نوع دوم، مثال ۳

شکل ۲ روند تغییر احتمال‌های خطای واقعی نوع اول و دوم آزمون UMP در رابطه (۵) را برحسب اندازه نمونه برای $\theta_1 = 1$ ، $\theta_2 = 2$ ، نشان می‌دهد. برای رسم نمودارها مقادیر دقیق α' و β' محاسبه شده است. مقدار θ^* تقریباً برابر ۱٫۴۴۳ است؛ از این رو، برای $\theta^* \in \{1/8, 1/2, 1/3, 1/4\}$ ، فرضیه H_1 ارجح از H_2 است، اما از

۲۶۴ مقایسه دو رویکرد نیمن-پیرسونی و شواهدی در آزمون فرضیه‌های یک طرفه جدا از هم

شکل ۲- الف مشاهده می‌شود با افزایش n ، α' نیز افزایش می‌یابد! و هر چه θ^* به $\theta^\dagger$ نزدیک‌تر باشد سریع‌تر به یک میل می‌کند. برای $\theta^* \in \{1/6, 1/8, 1/4, 1/2\}$ ، فرضیه H_2 ارجح از H_1 است و از شکل ۲-ب مشاهده می‌شود با افزایش n ، β' به صفر میل می‌کند.

۳.۲ آزمون فرضیه شواهدی

ساختار آزمون شواهدی $H_1 : \theta^* \in \Theta_1$ در مقابل $H_2 : \theta^* \in \Theta_2$ را می‌توان به صورت

$$\Psi(\mathbf{x}) = \begin{cases} \text{شواهد قوی در پشتیبانی از } H_1 & \Lambda(\mathbf{x}) \geq \log k, \\ \text{شواهد ضعیف} & |\Lambda(\mathbf{x})| < \log k, \\ \text{شواهد قوی در پشتیبانی از } H_2 & \Lambda(\mathbf{x}) \leq -\log k, \end{cases} \quad (۶)$$

بیان کرد، که در آن $\Lambda(\mathbf{x}) = \log \left[\frac{\sup_{H_1} L(\theta)}{\sup_{H_2} L(\theta)} \right]$. مقدار k (> 1) توسط محقق انتخاب می‌شود و به احتمال خطاها و اندازه نمونه بستگی ندارد. این احتمال خطاها هستند که به k بستگی دارند و نه برعکس.

علاوه براین، رویال (۲۰۰۰) احتمال خطاهای بالقوه رویکرد شواهدی در مرحله پیش از جمع‌آوری داده‌ها را تعریف کرد که در طراحی آزمایش و تعیین اندازه نمونه کاربرد دارد و به‌عنوان یک گواه، این اطمینان را به محقق می‌دهد که نتایج آزمون اغلب او را گمراه نمی‌کند. رویکرد شواهدی ممکن است منجر به خطا در استنباط شود که بر دو گونه شواهد قوی گمراه‌کننده و شواهد ضعیف است که به ترتیب با نمادهای M و W نمایش داده می‌شوند. احتمال خطاهای رویکرد شواهدی وقتی داده‌ها از مدل f_{θ^*} ، $\theta^* \in \Theta_1$ ، باشند $M_1 = P_{\theta^*}(\Lambda(\mathbf{x}) \leq -\log k | H_1)$ و $W_1 = P_{\theta^*}(-\log k < \Lambda(\mathbf{x}) < \log k | H_1)$ و احتمال خطاها وقتی داده‌ها از مدل f_{θ^*} ، $\theta^* \in \Theta_2$ ، باشند $M_2 = P_{\theta^*}(\Lambda(\mathbf{x}) \geq \log k | H_2)$ و $W_2 = P_{\theta^*}(-\log k < \Lambda(\mathbf{x}) < \log k | H_2)$ است.

تعریف ۶ (شواهد قوی گمراه‌کننده واقعی و شواهد ضعیف واقعی). شواهد قوی گمراه‌کننده واقعی، شواهد قوی در پشتیبانی از یک فرضیه است در حالی که فرضیه دیگر به واقعیت نزدیک‌تر است و شواهد ضعیف واقعی، عدم مشاهده شواهدی قوی است در حالی که یک فرضیه نسبت به فرضیه دیگر به واقعیت نزدیک‌تر است. احتمال شواهد قوی گمراه‌کننده واقعی و احتمال شواهد ضعیف واقعی در حالت $\Delta K > 0$ برابر $M'_1 = P_{\theta^*}(\Lambda(\mathbf{X}) \leq -\log k)$ و $W'_1 = P_{\theta^*}(|\Lambda(\mathbf{X})| < \log k)$ و در حالت $\Delta K < 0$ برابر $M'_2 = P_{\theta^*}(\Lambda(\mathbf{X}) \geq \log k)$ و $W'_2 = P_{\theta^*}(|\Lambda(\mathbf{X})| < \log k)$ هستند.

تذکر ۴. در تعریف ۶، اگر $\theta^* \in \Theta_1$ ، آنگاه $\Delta K > 0$ و در نتیجه $M'_1 = M_1$ و $W'_1 = W_1$. همچنین، اگر $\theta^* \in \Theta_2$ ، آنگاه $\Delta K < 0$ و در نتیجه $M'_2 = M_2$ و $W'_2 = W_2$.

قضیه ۲. فرض کنید $f_{\theta^*} \stackrel{iid}{\sim} X_1, \dots, X_n$ و $\mathcal{P} = \{f_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ یک خانواده تک پارامتری باشد به‌طوری که $\theta^* \in \Theta$ و $L(\theta)$ روی فضای پارامتری Θ تک‌مدی باشد. برای آزمون $\theta_1 \leq \theta^* : H_1$ در مقابل

شواهد قوی گمراه‌کننده واقعی و کل خطای واقعی برای $\Delta K > 0$ ، به‌ترتیب در روابط

$$M'_1 \leq P_{\theta^*}(\hat{\theta} \geq \theta_r) + P_{\theta^*}\left(\frac{L(\theta_1)}{L(\theta_r)} \leq \frac{1}{k}\right),$$

$$W'_1 + M'_1 \leq P_{\theta^*}(\hat{\theta} \geq \theta_r) + P_{\theta^*}\left(\frac{L(\theta_1)}{L(\theta_r)} \leq k\right) + P_{\theta^*}\left(\frac{L(\theta^*)}{L(\theta_r)} \leq k\right),$$

و برای $\Delta K < 0$ ، به‌ترتیب در روابط

$$M'_r \leq P_{\theta^*}(\hat{\theta} \leq \theta_1) + P_{\theta^*}\left(\frac{L(\theta_1)}{L(\theta_r)} \geq k\right),$$

$$W'_r + M'_r \leq P_{\theta^*}(\hat{\theta} \leq \theta_1) + P_{\theta^*}\left(\frac{L(\theta_1)}{L(\theta_r)} \geq \frac{1}{k}\right) + P_{\theta^*}\left(\frac{L(\theta_1)}{L(\theta^*)} \geq \frac{1}{k}\right),$$

صدق می‌کنند که در آن‌ها $\hat{\theta}$ ، برآوردگر ماکسیمم درست‌نمایی θ است.

برهان: با توجه تک‌مدی بودن تابع درست‌نمایی داریم:

$$\Lambda(\mathbf{x}) = \log \begin{cases} \frac{L(\hat{\theta})}{L(\theta_r)} & \hat{\theta} \leq \theta_1, \\ \frac{L(\theta_1)}{L(\theta_r)} & \theta_1 < \hat{\theta} < \theta_r, \\ \frac{L(\theta_1)}{L(\hat{\theta})} & \hat{\theta} \geq \theta_r. \end{cases}$$

بنابراین، برای $\Delta K > 0$ ،

$$M'_1 = P_{\theta^*}\left[\frac{L(\hat{\theta})}{L(\theta_r)} \leq \frac{1}{k}, \hat{\theta} \leq \theta_1\right] + P_{\theta^*}\left[\frac{L(\theta_1)}{L(\theta_r)} \leq \frac{1}{k}, \theta_1 < \hat{\theta} < \theta_r\right]$$

$$+ P_{\theta^*}\left[\frac{L(\theta_1)}{L(\hat{\theta})} \leq \frac{1}{k}, \hat{\theta} \geq \theta_r\right].$$

بنابراین $P_{\theta^*}\left[\frac{L(\hat{\theta})}{L(\theta_r)} \leq \frac{1}{k}, \hat{\theta} \leq \theta_1\right] = 0$ داریم $L(\hat{\theta}) \geq L(\theta_r)$ به توجه به

$$M'_1 \leq P_{\theta^*} \left[\frac{L(\theta_1)}{L(\theta_r)} \leq \frac{1}{k} \right] + P_{\theta^*} [\hat{\theta} \geq \theta_r] \text{ از طرفی}$$

$$\begin{aligned} W'_1 + M'_1 &= P_{\theta^*} \left[\frac{L(\hat{\theta})}{L(\theta_r)} \leq k, \hat{\theta} \leq \theta_1 \right] + P_{\theta^*} \left[\frac{L(\theta_1)}{L(\theta_r)} \leq k, \theta_1 < \hat{\theta} < \theta_r \right] \\ &+ P_{\theta^*} \left[\frac{L(\theta_1)}{L(\hat{\theta})} \leq k, \hat{\theta} \geq \theta_r \right]. \end{aligned}$$

با توجه به $k \leq \frac{L(\hat{\theta})}{L(\theta_r)} \frac{L(\theta^*)}{L(\theta_r)} = \frac{L(\hat{\theta})}{L(\theta_r)} \leq k$ داریم $\frac{L(\theta^*)}{L(\theta_r)} \leq k$ از این رو،

$$W'_1 + M'_1 \leq P_{\theta^*} \left[\frac{L(\theta^*)}{L(\theta_r)} \leq k \right] + P_{\theta^*} \left[\frac{L(\theta_1)}{L(\theta_r)} \leq k \right] + P_{\theta^*} [\hat{\theta} \geq \theta_r].$$

برای M'_1 و $W'_1 + M'_1$ نیز نتایج به‌طور مشابه به‌دست می‌آید.

قضیه ۳. فرض کنید شرایط قضیه ۲ برقرار باشد. تحت شرایط نظم، M'_j و $W'_j + M'_j$ ، $j = 1, 2$ ، با افزایش اندازه نمونه به صفر میل می‌کنند.

برهان: از دنیس و همکاران (۲۰۱۹) نتیجه می‌شود که برای اندازه نمونه بزرگ

$$P_{\theta^*} \left[\log \left(\frac{L(\theta_1)}{L(\theta_r)} \right) \leq \log k \right] \approx \Phi \left(\frac{\log k}{\sqrt{n}\sigma_{1r}} - \sqrt{n} \frac{\Delta K}{\sigma_{1r}} \right), \quad (۷)$$

$$P_{\theta^*} \left[\log \left(\frac{L(\theta^*)}{L(\theta_r)} \right) \leq \log k \right] \approx \Phi \left(\frac{\log k}{\sqrt{n}\sigma_1} - \sqrt{n} \frac{K_{1r}}{\sigma_1} \right). \quad (۸)$$

از آنجا که $\Delta K > 0$ و $K_{1r} > 0$ ، زمانی که $n \rightarrow \infty$ ، احتمال‌های در روابط (۷) و (۸) به صفر میل می‌کنند. از طرفی بنا به اصل عکس نقیض، گزاره منطقی $\{\hat{\theta} \geq \theta_r \implies \Delta K = -K_{r1} < 0\}$ گزاره منطقی $\{\Delta K > 0 \implies \hat{\theta} < \theta_r\}$ را نتیجه می‌دهد. بنابراین، برای $\Delta K > 0$ بنا به لم ۲، $\lim_{n \rightarrow \infty} P_{\theta^*} [\hat{\theta} \geq \theta_r] = 0$. در نتیجه ثابت می‌شود که M'_j و $W'_j + M'_j$ با افزایش اندازه نمونه به صفر میل می‌کنند. برای M'_1 و $W'_1 + M'_1$ نیز نتایج به‌طور مشابه به‌دست می‌آید.

از قضیه ۳ نتیجه می‌شود که احتمال شواهد واقعی گمراه‌کننده قوی و احتمال کل خطای واقعی تحت بدمشخص‌سازی مدل، با افزایش اندازه نمونه کاهش یافته و به صفر میل می‌کند؛ به عبارت دیگر، احتمال واقعی شواهد قوی در پشتیبانی از مدلی که به مدل واقعی نزدیک‌تر است یعنی $(W'_j + M'_j) - 1$ ، $j = 1, 2$ ، با افزایش اندازه نمونه افزایش یافته و به یک میل می‌کند.

مثال ۴. در ادامه مثال ۲، برای آزمون H_1 در مقابل H_2 ، به روش شواهدی و آزمون (۶) داریم:

$$\Lambda(\mathbf{x}) = \log \begin{cases} \frac{L(\bar{x})}{L(\theta_2)} & \bar{x} \leq \theta_1 \\ \frac{L(\theta_1)}{L(\theta_2)} & \theta_1 < \bar{x} < \theta_2 \\ \frac{L(\theta_1)}{L(\bar{x})} & \bar{x} \geq \theta_2 \end{cases} = \begin{cases} \frac{n}{\sigma^2} (\bar{x} - \theta_2)^2 & \bar{x} \leq \theta_1, \\ \frac{n}{\sigma^2} (\theta_2 - \theta_1) (\bar{\theta} - \bar{x}) & \theta_1 < \bar{x} < \theta_2, \\ -\frac{n}{\sigma^2} (\bar{x} - \theta_1)^2 & \bar{x} \geq \theta_2. \end{cases}$$

توجه کنید که $\Lambda(\mathbf{x})$ تابعی پیوسته و اکیداً نزولی از $\bar{x}$ است. پس،

$$\Psi(\mathbf{x}) = \begin{cases} H_1 \text{ شواهد قوی در پشتیبانی از} & \bar{x} \leq a, \\ \text{شواهد ضعیف} & \text{سایر جاها،} \\ H_2 \text{ شواهد قوی در پشتیبانی از} & \bar{x} \geq b, \end{cases} \quad (9)$$

که در آن برای $n \leq \frac{2\sigma^2}{(\theta_2 - \theta_1)^2} \log k$ و $a = \theta_2 - \frac{\sigma}{\sqrt{n}} \sqrt{2 \log k}$ و $b = \theta_1 + \frac{\sigma}{\sqrt{n}} \sqrt{2 \log k}$ ، و برای $n \geq \frac{2\sigma^2}{(\theta_2 - \theta_1)^2} \log k$ و $a = \bar{\theta} - \frac{\sigma}{\sqrt{n}} \frac{\log k}{\theta_2 - \theta_1}$ و $b = \bar{\theta} + \frac{\sigma}{\sqrt{n}} \frac{\log k}{\theta_2 - \theta_1}$ ، برای $\theta^* \in (\theta_1, \theta_2)$ داریم:

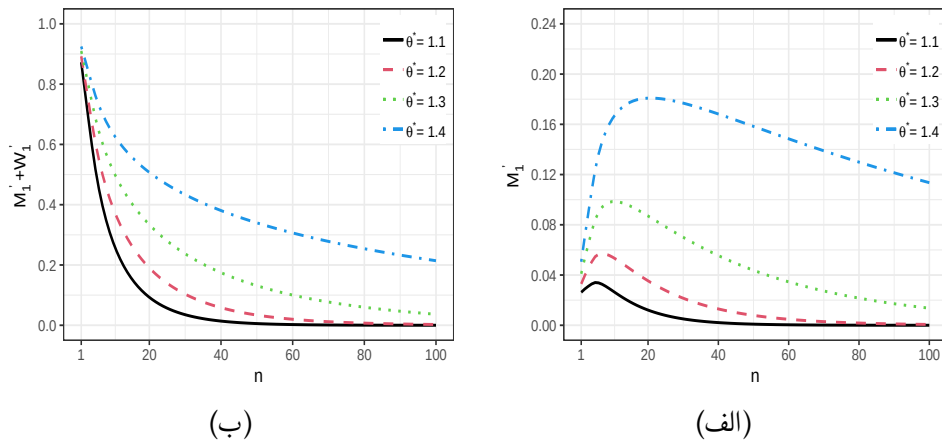
$$M'_j = \begin{cases} \Phi \left(\sqrt{n} \frac{|\theta^* - \theta_j|}{\sigma} - \sqrt{2 \log k} \right) & n \leq \frac{2\sigma^2}{(\theta_2 - \theta_1)^2} \log k, \\ \Phi \left(-\sqrt{n} \frac{|\theta^* - \bar{\theta}|}{\sigma} - \frac{\sigma}{\sqrt{n}} \frac{\log k}{\theta_2 - \theta_1} \right) & n \geq \frac{2\sigma^2}{(\theta_2 - \theta_1)^2} \log k, \end{cases}$$

$$M'_j + W'_j = \begin{cases} \Phi \left(-\sqrt{n} \frac{|\theta^* - \theta_{2-j}|}{\sigma} + \sqrt{2 \log k} \right) & n \leq \frac{2\sigma^2}{(\theta_2 - \theta_1)^2} \log k, \\ \Phi \left(-\sqrt{n} \frac{|\theta^* - \bar{\theta}|}{\sigma} + \frac{\sigma}{\sqrt{n}} \frac{\log k}{\theta_2 - \theta_1} \right) & n \geq \frac{2\sigma^2}{(\theta_2 - \theta_1)^2} \log k. \end{cases}$$

توجه کنید که همواره $M'_j + W'_j > M'_j$ ، $j = 1, 2$. به عبارت دیگر، در این آزمون شواهدی همواره احتمال شواهد قوی درست بیشتر از احتمال شواهد قوی نادرست است. همچنین، مشاهده می‌شود که با افزایش اندازه نمونه، احتمال شواهد قوی گمراه‌کننده واقعی و احتمال کل خطای واقعی به صفر میل می‌کنند. اگر $\theta^* = \bar{\theta}$ ، آنگاه با افزایش اندازه نمونه، $M'_j + W'_j$ به ۰.۵ میل می‌کنند. به عبارت دیگر، در حالتی که هر دو فرضیه دارای ارجحیت یکسان هستند احتمال شواهد قوی در پشتیبانی هر دو فرضیه، $1 - (M'_j + W'_j)$ ، برابر است.

با توجه به تذکر ۴، اگر $\theta^* \leq \theta_1$ ، آنگاه $M'_1 = M_1$ و $M'_2 + W'_2 = M_2 + W_2$ ، و اگر $\theta^* \geq \theta_2$ ، آنگاه $M'_1 + W'_1 = M_1 + W_1$ و $M'_2 = M_2$.

شکل ۳-الف روند تغییر احتمال شواهد قوی گمراه‌کننده واقعی و احتمال کل خطای واقعی (شکل ۳-ب) آزمون شواهدی در رابطه (۹) را برحسب اندازه نمونه برای $\theta_1 = 1$ ، $\theta_2 = 2$ ، $\sigma = 1$ ، $k = 8$ و مقادیر $\theta^* \in \{1/8, 1/4, 1/2, 3/4, 1\}$ نشان داده شده است. از شکل ۳-الف مشاهده می‌شود که روند تغییر احتمال شواهد واقعی قوی گمراه‌کننده با افزایش اندازه نمونه در ابتدا افزایشی و سپس کاهش‌ی است. حداکثر مقدار احتمال شواهد قوی گمراه‌کننده واقعی برابر $\Phi(-2\sqrt{\frac{\bar{\theta} - \theta^*}{\theta_2 - \theta_1}} \log k)$ است که در اندازه نمونه نزدیک به $\frac{\sigma^2 \log k}{(\theta_2 - \theta_1)(\bar{\theta} - \theta^*)}$ رخ می‌دهد.



شکل ۳. الف: احتمال شواهد قوی گمراه‌کننده واقعی و ب: احتمال کل خطای واقعی، مثال ۴

برای مقادیر $1/1$ ، $1/2$ ، $1/3$ و $1/4$ از میانگین واقعی، حداکثر مقدار احتمال شواهد قوی گمراه‌کننده واقعی به ترتیب برابر $0/34$ ، $0/57$ ، $0/99$ و $1/81$ است که به ترتیب در اندازه نمونه 5 ، 7 ، 10 و 21 رخ می‌دهد. در شکل ۳-ب مشاهده می‌شود که روند تغییر احتمال کل خطای واقعی با افزایش اندازه نمونه اکیداً کاهشی است. به عبارت دیگر، با افزایش اندازه نمونه احتمال شواهد قوی واقعی در پشتیبانی فرضیه ارجح نیز افزایش می‌یابد. با توجه به نتایج، در حالت $n = 100$ و برای مقادیر $1/1$ ، $1/2$ ، $1/3$ و $1/4$ از میانگین واقعی، احتمال شواهد قوی واقعی در پشتیبانی فرضیه ارجح H_1 ، یعنی $1 - (M'_1 + W'_1)$ ، به ترتیب برابر $1/00$ ، $0/997$ ، $0/963$ و $0/786$ است. در هر دو نمودار شکل ۳، احتمال‌های خطا نسبت به θ^* اکیداً نزولی هستند.

مثال ۵. در ادامه مثال ۳، برای آزمون H_1 در مقابل H_2 ، به روش شواهدی و آزمون (۶) داریم:

$$\Lambda(\mathbf{x}) = \log \begin{cases} \frac{L(\bar{x})}{L(\theta_1)} & \bar{x} \leq \theta_1 \\ \frac{L(\theta_1)}{L(\theta_2)} & \theta_1 < \bar{x} < \theta_2 \\ \frac{L(\theta_2)}{L(\bar{x})} & \bar{x} \geq \theta_2 \end{cases} = \begin{cases} n(\theta_2 - \bar{x}) - n\bar{x} \log \frac{\theta_2}{\bar{x}} & \bar{x} \leq \theta_1, \\ n(\theta_2 - \theta_1) - n\bar{x} \log \frac{\theta_2}{\theta_1} & \theta_1 < \bar{x} < \theta_2, \\ n(\bar{x} - \theta_1) - n\bar{x} \log \frac{\bar{x}}{\theta_1} & \bar{x} \geq \theta_2. \end{cases}$$

توجه کنید که $\Lambda(\mathbf{x})$ تابعی پیوسته و اکیداً نزولی از $\bar{x}$ است. پس،

$$\Psi(\mathbf{x}) = \begin{cases} \text{شواهد قوی در پشتیبانی از } H_1 & \bar{x} \leq a, \\ \text{شواهد ضعیف} & \text{سایر جاها،} \\ \text{شواهد قوی در پشتیبانی از } H_2 & \bar{x} \geq b, \end{cases} \quad (10)$$

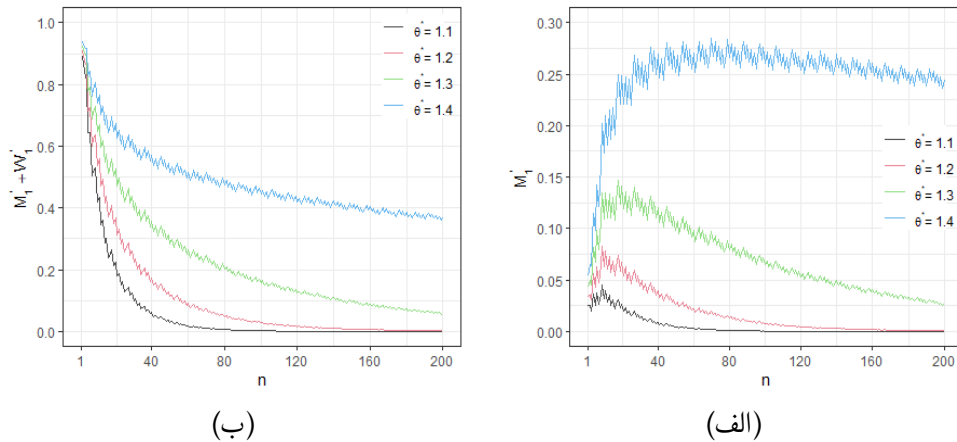
که در آن برای $n > \frac{\log k}{\theta_1 \log \frac{\theta_1}{\theta_1 + \theta_2 - \theta_1}}$ و برای $a = \frac{(\theta_1 - \theta_1) - \frac{\log k}{n}}{\log \frac{\theta_1}{\theta_1}}$ ریشه معادله $a \log \frac{a}{\theta_1} + \theta_2 - a - \frac{\log k}{n} = 0$ است؛ همچنین، برای $n > \frac{\log k}{\theta_2 \log \frac{\theta_2}{\theta_1 + \theta_2 - \theta_2}}$ و برای $b = \frac{(\theta_2 - \theta_2) + \frac{\log k}{n}}{\log \frac{\theta_2}{\theta_1}}$ ریشه معادله $b \log \frac{b}{\theta_1} + \theta_1 - b - \frac{\log k}{n} = 0$ است. از این‌رو، برای $\theta^* < \theta^\dagger$ احتمال واقعی خطاها به صورت $M'_1 + W'_1 = P_{\theta^*}(\sum_{i=1}^n X_i > na)$ و $M'_1 = P_{\theta^*}(\sum_{i=1}^n X_i \geq nb)$ و برای $\theta^* > \theta^\dagger$ نیز $M'_1 + W'_1 = P_{\theta^*}(\sum_{i=1}^n X_i < nb)$ و $M'_1 = P_{\theta^*}(\sum_{i=1}^n X_i \leq na)$ به دست می‌آید. برای n بزرگ و $j = 1, 2$ ، احتمال واقعی خطاها را می‌توان به صورت

$$M'_j \approx \Phi\left(\frac{1}{\log \frac{\theta_1}{\theta_1} \sqrt{\theta^*}} \left[-\sqrt{n}|\Delta K| - \frac{\log k}{\sqrt{n}} \right]\right),$$

$$M'_j + W'_j \approx \Phi\left(\frac{1}{\log \frac{\theta_1}{\theta_1} \sqrt{\theta^*}} \left[-\sqrt{n}|\Delta K| + \frac{\log k}{\sqrt{n}} \right]\right),$$

تقریب زد، که در آن برای $M'_1 + W'_1$ و $M'_1 = M_1 + W_1$ و برای $\Delta K > 0$ ، $M'_1 + W'_1 = M_1 + W_1$ و $M'_1 = M_1$ ، بنابراین، مشاهده می‌شود که با افزایش اندازه نمونه، احتمال شواهد قوی گمراه‌کننده واقعی و احتمال کل خطای واقعی به صفر میل می‌کنند. اگر $\theta^* = \theta^\dagger$ ، آنگاه با افزایش اندازه نمونه، $M'_1 + W'_1$ به 0.5 میل می‌کند. به عبارت دیگر، در حالتی که هر دو فرضیه دارای ارجحیت یکسان هستند احتمال شواهد قوی در پشتیبانی هر دو فرضیه، $1 - (M'_1 + W'_1)$ برابر است.

با توجه به تذکر ۴، اگر $\theta^* \leq \theta_1$ ، آنگاه $M'_1 = M_1$ و $M'_1 + W'_1 = M_1 + W_1$ و اگر $\theta^* \geq \theta_2$ ، آنگاه $M'_2 + W'_2 = M_2 + W_2$ و $M'_2 = M_2$.



شکل ۴. الف: احتمال شواهد قوی گمراه‌کننده واقعی و ب: احتمال کل خطای واقعی، مثال ۵

شکل ۴- الف روند تغییر احتمال شواهد قوی گمراه‌کننده واقعی و احتمال کل خطای واقعی (شکل ۴-ب) آزمون شواهدی در رابطه (۱۰) را برحسب اندازه نمونه برای $\theta_1 = 1, \theta_2 = 2, k = 8$ و مقادیر $\theta^* \in \{1/4, 1/3, 1/2, 1/1\}$ نشان می‌دهد. توجه شود که برای رسم نمودارها مقادیر دقیق M'_j و $W'_j + M'_j$ محاسبه شده است و نه مقادیر تقریبی. همچنین، نوسان مشاهده شده در منحنی‌ها به این دلیل است که $\sum_{i=1}^n X_i$ یک متغیر تصادفی گسسته است؛ از این رو، برای محاسبه احتمال‌ها تنها مقادیر صحیح بزرگتر از nb و na لحاظ می‌شود؛ این در حالی است که $\sum_{i=1}^n X_i$ دارای توزیع پواسون با میانگین $n\theta^*$ است و $n\theta^*$ می‌تواند مقادیر ناصحیح را نیز اختیار کند. در شکل ۴- الف مشاهده می‌شود که روند تغییر احتمال شواهد واقعی قوی گمراه‌کننده با افزایش اندازه نمونه در ابتدا افزایشی و سپس کاهشی است. برای مقادیر $1/8, 1/4, 1/3, 1/2$ از میانگین واقعی، حداکثر مقدار احتمال شواهد قوی گمراه‌کننده واقعی به ترتیب برابر $0.45, 0.82, 0.46$ و 0.285 است که به ترتیب در اندازه نمونه $9, 18$ و 70 رخ می‌دهد. در شکل ۴- ب مشاهده می‌شود که روند تغییر احتمال کل خطای واقعی با افزایش اندازه نمونه کاهشی است. به عبارت دیگر، با افزایش اندازه نمونه احتمال شواهد قوی واقعی در پشتیبانی فرضیه ارجح نیز افزایش می‌یابد. با توجه به نتایج، در حالت $n = 200$ و برای مقادیر $1/8, 1/4, 1/3$ از میانگین واقعی، احتمال شواهد قوی واقعی در پشتیبانی فرضیه ارجح H_1 ، یعنی $1 - (M'_j + W'_j)$ ، به ترتیب برابر $0.998, 0.999$ و 0.941 است. در هر دو نمودار شکل ۴، احتمال‌های خطا نسبت به θ^* اکیداً نزولی هستند.

قضیه ۴. فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی به اندازه n از توزیعی متعلق به خانواده نمایی تک پارامتری با فرم کانونی $\{h(x) = \exp\{T(x)\theta^* - A(\theta^*)\} / \exp\{T(x)\theta_1 - A(\theta_1)\}\}$ است. قصد داریم آزمون $H_0: \theta^* \leq \theta_1$ مقابل $H_1: \theta^* \geq \theta_2$ که $\theta_1 < \theta^* < \theta_2$ را به روش شواهدی و آزمون در رابطه (۶) انجام دهیم. در این صورت، برای یک اندازه نمونه بزرگ داریم:

$$M'_j \approx \Phi\left(-\sqrt{n} \frac{|A'(\theta^*) - A'(\theta_1)|}{(A''(\theta^*))^{3/2}} - \frac{1}{\sqrt{n}} \frac{\log k}{(\theta_2 - \theta_1)(A''(\theta^*))^{3/2}}\right),$$

$$M'_j + W'_j \approx \Phi\left(-\sqrt{n} \frac{|A'(\theta^*) - A'(\theta_1)|}{(A''(\theta^*))^{3/2}} + \frac{1}{\sqrt{n}} \frac{\log k}{(\theta_2 - \theta_1)(A''(\theta^*))^{3/2}}\right),$$

که در آن A' و A'' به ترتیب مشتقات مرتبه اول و دوم A ، θ_0 ریشه معادله $\Delta K = 0$ ، و $j = 1, 2$.

برهان: با توجه به برقراری شرایط نظم برای خانواده توزیع‌های نمایی داریم $E_{\theta^*}(T(X)) = A'(\theta^*)$ و $I(\theta^*) = V_{\theta^*}(T(X)) = A''(\theta^*)$ که در آن $I(\theta^*) = \text{اطلاع فیشر پارامتر } \theta^*$ است. پس، از آنجا که $A'(\theta^*) \geq A'(\theta_1)$ ، تابعی صعودی از θ^* است. اکنون، معیار کولبک-لایبلر f_{θ} از f_{θ^*} چنین $K(f_{\theta^*} \| f_{\theta}) = A'(\theta^*)(\theta^* - \theta) - (A(\theta^*) - A(\theta))$ به دست می‌آید. معیار $K(f_{\theta^*} \| f_{\theta})$ برای $\theta < \theta^*$ تابعی نزولی و برای $\theta > \theta^*$ تابعی صعودی از θ است. بنابراین، $K(f_{\theta^*} \| \mathcal{P}_j) = K(f_{\theta^*} \| f_{\theta_j})$ و $\Delta K = -A'(\theta^*)(\theta_2 - \theta_1) + (A(\theta_2) - A(\theta_1))$ ، ریشه معادله $\Delta K = 0$ ، یعنی θ_0 ، در رابطه

همچنین، $\Delta K > (<)$ اگر و فقط اگر $\theta_0 < (>)\theta^*$. از طرفی، $A'(\theta_0) = \frac{A(\theta_0)-A(\theta_1)}{\theta_0-\theta_1}$ صدق می‌کند. همچنین، $L(\theta) = n\bar{T}(x)\theta - nA(\theta) + \prod_{i=1}^n h(x_i)$ و برآورد ماکسیمم درستنمایی θ^* ، یعنی $\hat{\theta}$ ، در رابطه $A'(\hat{\theta}) = \bar{T}(x)$ صدق می‌کند. برای اندازه نمونه‌های بزرگ که $n > \frac{\log k}{\min(K_{12}, K_{21})}$ ، تابع آزمون شواهدی به صورت

$$\Psi(x) = \begin{cases} \text{شواهد قوی در پشتیبانی از } H_1 & A'(\hat{\theta}) \leq \frac{A(\theta_2)-A(\theta_1)}{\theta_2-\theta_1} - \frac{1}{n} \frac{\log k}{\theta_2-\theta_1}, \\ \text{شواهد ضعیف} & |A'(\hat{\theta}) - \frac{A(\theta_2)-A(\theta_1)}{\theta_2-\theta_1}| < \frac{1}{n} \frac{\log k}{\theta_2-\theta_1}, \\ \text{شواهد قوی در پشتیبانی از } H_2 & A'(\hat{\theta}) \geq \frac{A(\theta_2)-A(\theta_1)}{\theta_2-\theta_1} + \frac{1}{n} \frac{\log k}{\theta_2-\theta_1}. \end{cases}$$

به دست می‌آید. از طرفی، با توجه به ویژگی‌های حدی برآورد ماکسیمم درستنمایی در خانواده توزیع‌های نمایی و با استفاده از روش دلتا، برای اندازه نمونه بزرگ می‌توان توزیع $A'(\hat{\theta})$ را توزیع $N(A'(\theta^*), \frac{(A''(\theta^*))^r}{n})$ تقریب زد. بنابراین، برای اندازه نمونه‌های بزرگ که $n > \frac{\log k}{\min(K_{12}, K_{21})}$ ، نتیجه مطلوب به دست می‌آید.

تذکر ۵. فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی از $\{T(x)\xi(\eta^*) - B(\eta^*)\}$ است. $f_{\eta^*}(x) \propto \exp\{T(x)\xi(\eta^*) - B(\eta^*)\}$ است. قصد داریم آزمون $H_1: \eta^* \leq \eta_1$ در مقابل $H_2: \eta^* \geq \eta_2$ که $\eta_1 < \eta^* < \eta_2$ را به روش شواهدی انجام دهیم. در صورتی که $\xi(\eta)$ تابعی صعودی از η باشد، کافی است قرار دهیم $\theta^* = \xi(\eta^*)$. همچنین، در صورتی که $\xi(\eta)$ تابعی نزولی از η باشد، کافی است قرار دهیم $\theta^* = -\xi(\eta^*)$ و $T(x)$ را با $-T(x)$ جایگزین کنیم. در هر دو صورت، آزمون H_1 در مقابل H_2 معادل آزمون $H_1: \theta^* \leq \theta_1$ در مقابل $H_2: \theta^* \geq \theta_2$ است و با استفاده از قضیه ۴ می‌توان تقریبی از احتمال خطاها به دست آورد.

از قضیه ۴، احتمال خطاها در نمونه‌های بزرگ رفتاری مشابه آنها برای توزیع نرمال در مثال ۴ است.

بحث و نتیجه‌گیری

هدف از مطالعه حاضر بررسی عملکرد دو رویکرد نیمن-پیرسونی و شواهدی در مساله آزمون آزمون فرضیه‌های یک‌طرفه جدا از هم را براساس احتمال خطای استنباط تحت بدمشخص سازی مدل بود. برای نیل به چنین هدفی از ابزاری به نام معیار واگرایی کولبک-لایبلر استفاده شد. نتایج نشان داد که رویکرد نیمن-پیرسونی، تحت شرایطی که امکان رخ داد آنها در واقعیت وجود دارد، با افزایش اندازه نمونه، آزمون مبتنی بر این رویکرد به اشتباه تمایل به رد فرضیه صحیح‌تر صفر دارد. اما در مقابل، در رویکرد شواهدی با افزایش حجم نمونه، احتمال شواهد قوی درست به یک میل می‌کند. در نتیجه، به لحاظ خطاها در استنباط، رویکرد شواهدی عملکرد بهتری نسبت به رویکرد نیمن-پیرسونی دارد.

تقدیر و تشکر

از داوران، سردبیر و ویراستار محترم مجله که موجبات ارتقای کیفیت مقاله را فراهم آوردند سپاسگزاری می‌نمایم.

مراجع

ارقامی، ن.ر.، زمان‌زاده، ا.، و راحت، س. (۱۳۸۷)، شواهد آماری: تفسیر نتایج آزمون‌های آماری کلاسیک، اندیشه آماری، ۱۳(۱)، ۳-۱۰.

ارقامی، ن.ر.، دست‌برآورده، ع.، و زمان‌زاده، ا. (۱۳۸۷)، شواهد آماری، قسمت دوم: قانون درست‌نمایی و معایب p -مقدار به‌عنوان پشتیبانی داده‌ها از فرضیه صفر، اندیشه آماری، ۱۳(۲)، ۳-۱۶.

اسماعیلی، ح.، توحیدی، م.، روزگار، س.ر.، و امیری، م. (۱۳۹۰)، p -مقدار معمولی و p -مقدار اصلاح شده، چگونه بهتر قضاوت کنیم؟، مجله علوم آماری، ۵(۱)، ۱-۲۲.

دست‌برآورده، ع. (۱۳۸۸)، نظریه نیمین-پیرسون از دیدگاه شواهدی. پایان‌نامه کارشناسی ارشد، مشهد، دانشگاه فردوسی مشهد، دانشکده علوم ریاضی، گروه آمار.

رویال، ر.م. (۱۳۹۸)، شواهد آماری: رویکردی مبتنی بر درست‌نمایی. ترجمه: ارقامی، ن.ر. و طاهری، س.م. چاپ اول، مرکز نشر دانشگاهی.

طاهری، س.م. (۱۳۹۸)، جایگاه استنباط شواهدی در فلسفه آمار، اولین سمینار استنباط شواهدی، مشهد، دانشگاه فردوسی مشهد، ۱۰-۱۶.

عمادی، م. (۱۳۸۴)، استنباط شواهدگرا. رساله دکتری، مشهد، دانشگاه فردوسی مشهد، دانشکده علوم ریاضی، گروه آمار.

هاشم‌پور، م. (۱۳۹۵)، استنباط کلاسیک، بیزی و شواهدی براساس آماره‌های ترتیبی دنباله‌ای. رساله دکتری، مشهد، دانشگاه فردوسی مشهد، دانشکده علوم ریاضی، گروه آمار.

Box, G. E. (1979), Robustness in the Strategy of Scientific Model Building, *Robustness in Statistics*, ed G. Wilkinson (New York, NY: Academic Press), 201–236.

Dennis, B., Ponciano, J. M., Taper, M. L., and Lele, S. R. (2019), Errors in Statistical Inference Under Model Misspecification: Evidence, Hypothesis Testing, and AIC, *Frontiers in Ecology and Evolution*, 7, 372.

- Edwards, A. (1972). *Likelihood*. Cambridge: Cambridge University Press.
- Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge: Cambridge University Press.
- Karlin, S. and Rubin, H. (1956), The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio, *The Annals of Mathematical Statistics*, **27**(2), 272-299.
- Kullback, S. (1997), *Information Theory and Statistics*, Courier Corporation.
- Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses*, Vol. 3, New York: springer.
- Mayo, D. G., and Spanos, A. (2006), Severe Testing as a Basic Concept in a Neyman–Pearson Philosophy of Induction. *The British Journal for the Philosophy of Science*, **57**(2), 323-357.
- Michaelides, M. (2021), Large Sample Size Bias in Empirical Finance. *Finance Research Letters*, **41**, 101835.
- Mosteller, F. (1948), A k-sample slippage test for an extreme population. *Annals of Mathematical Statistics*, **19**, 58–65.
- Neyman, J., and Pearson, E. S. (1933), IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **231**(694-706), 289-337.
- Royall, R. M. (1992), The Elusive Concept of Statistical Evidence, *Bayesian statistics*, **4**, 405-413.
- Royall, R. M. (1997), *Statistical Evidence: A Likelihood Paradigm*, London, UK: Chapman and Hall.
- Royall, R. M. (2000), On the Probability of Observing Misleading Statistical Evidence, *Journal of the American Statistical Association*, **95**, 760-768.