مجله علوم آماری، بهار و تابستان ۱۳۹۰
جلد ۵، شماره ۱، ص ۱-۲۲

p-مقدار معمولی و p-مقدار اصلاح شده، چگونه بهتر قضاوت کنیم؟

حمید اسماعیلی[۱]، مینا توحیدی[۲]، سیدروح‌الله روزگار[۲]، مهدی امیری[۲]
[۱] معاونت اجتماعی و پیشگیری از وقوع جرم دادگستری کل استان بوشهر
[۲]گروه آمار، دانشگاه شیراز

تاریخ دریافت: ۱۳۸۹/۱/۲۲ تاریخ آخرین بازنگری: ۱۳۸۹/۵/۱۶

چکیده : اغلب در آزمون فرض از p-مقدار برای تصمیم‌گیری استفاده می‌شود. آیا p-مقدار بهترین معیار برای رد یا تأیید فرضیه صفر است؟ آیا می‌توان معیاری بهتر از آن در اختیار داشت؟ در این مقاله مسأله آزمون فرضیه نه به‌عنوان یک تصمیم بلکه به‌عنوان یک مسأله برآوردیابی برای احتمال رخ دادن مجموعه مشخص شده با Θ_0 در نظر گرفته می‌شود و از p-مقدار به‌عنوان برآوردگری برای احتمال رخ دادن Θ_0 استفاده خواهد شد. از طرفی در نظر گرفتن اعداد حقیقی به‌عنوان فضای پارامتری همواره مورد تأیید محققان بوده است. در حالی‌که در بسیاری از کاربردها فضای پارامتری محدود شده است. برای حالتی که فضای پارامتری کراندار باشد معیاری به نام p-مقدار اصلاح شده در توزیع نرمال برای آزمون‌های یک و دو طرفه ارائه خواهد شد که نسبت به p-مقدار معمولی عملکرد بهتری دارد.

واژه‌های کلیدی : p- مقدار اصلاح شده، قانون درستنمایی، فضای پارامتری محدود شده، استنباط شواهدی.

آدرس الکترونیک مسئول مقاله: حمید اسماعیلی، esmailihamid@ymail.com
کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲F۰۳

۱ مقدمه

روش‌های معمول آزمون فرض یک مسأله انتخاب تصمیم بین $H_۰$ و $H_۱$ است، که با استفاده از pـمقدار انجام می‌شود. در این مقاله فرض می‌شود آزمون فرضیه، یک مسأله برآوردیابی در نظریه تصمیم است، که در آن برآورد کردن احتمال رخ دادن مجموعه $\Theta_۰$ یا برآورد تابع $I_{\Theta_۰}(\theta)$ مورد نظر است. در اینجا $I_A(\cdot)$ تابع نشانگر مجموعه A است. سؤالی که مطرح می‌شود آن است که آیا pـمقدار می‌تواند برآوردگر مناسبی برای احتمال رخ دادن مجموعه $\Theta_۰$ باشد؟ در واقع pـمقدار به‌عنوان برآوردی برای $I_{\Theta_۰}(\theta)$ در نظر گرفته می‌شود. آیا pـمقدار می‌تواند بهترین معیار برای رد یا تأیید فرضیه صفر باشد؟ آیا می‌توان معیاری بهتر از pـمقدار برای بهتر تصمیم گرفتن در مورد رد یا تأیید فرضیه صفر در اختیار داشت؟ انتقادات زیادی از pـمقدار توسط آماردانان بیز مطرح شده است. اخیراً دیدگاه استنباط شواهدی مطرح شده است که تنها مبتنی بر مشاهدات است و تحت تأثیر توزیع پیشینی و تابع زیان نیست و کاملاً متکی بر تابع درست‌نمایی است. این دیدگاه اولین بار توسط رویال (۱۹۹۷) مطرح شد. سپس فارستر و سابر (۲۰۰۱)، گودمن (۲۰۰۵) به بررسی و مطالعه این دیدگاه پرداختند. در ایران نیز ارقامی و همکاران (۱۳۸۷a و ۱۳۸۷b) به بررسی معیارها با استفاده از دیدگاه استنباط شواهدی پرداختند و معایب pـمقدار را از دیدگاه استنباط شواهدی مطرح کردند که در بخش ۲ مقاله به بعضی از این معایب اشاره خواهد شد. pـمقدار از دیدگاه استنباط شواهدی به‌عنوان یک معیار شواهدی خوب مطرح نیست و مورد انتقاد قرار گرفته است. اما در نظر گرفتن فضای پارامتری به‌صورت $(-\infty, +\infty)$ همواره مورد تردید محققان بوده است. آیا وقتی محققی میانگین قد دانش‌آموزان یک دبیرستان را بررسی می‌کند در نظر گرفتن اعداد بسیار بزرگ و نزدیک به $+\infty$ برای میانگین قد نوجوانان دبیرستانی هرچند با احتمال بسیار کم، امری عجیب نیست؟ در بسیاری از آزمون‌ها فرض می‌شود مشاهدات از توزیع نرمال با فضای پارامتری $(-\infty, +\infty)$ پیروی می‌کنند. ولی در اغلب مسائل کاربردی فضای پارامتری کراندار است. به‌عنوان مثال برای آزمون وزن نوزادان تازه متولد شده در یک بیمارستان می‌توان

باتوجه به اطلاعات قبل یا یک نمونه‌گیری اولیه یک کران بالا و پایین برای فضای پارامتری به‌دست آورد. مندلکرن (۲۰۰۲) مثال‌هایی ارائه کرد که وقتی فضای پارامتر کران‌دار باشد، روند کلاسیک نیمن برای بسیاری از علوم رضایت بخش نمی‌باشد. او به تناقضاتی در مورد طول بازهٔ اطمینان برای θ وقتی $X \sim N(\theta, \sigma^2)$ و θ کراندار (مثلاً $\theta \geq 0$) و σ^2 معلوم است، یا برای λ وقتی $X \sim \text{Poisson}(\lambda)$ و λ کراندار باشد اشاره کرد. اگر فضای پارامتری کراندار باشد آیا pـمقدار که فقط به مشاهده x و مقدار θ_0 بستگی دارد و از اطلاعاتی که ما در مورد فضای پارامتری داریم استفاده نمی‌کند معیار مناسبی برای رد یا تأیید فرضیه صفر هست؟ وقتی $X \sim N(\theta, \sigma^2)$، آیا برای آزمون فرضیه‌های یک طرفه $\left\{ \begin{array}{ll} H_0 & : \theta \leq \theta_0 \\ H_1 & : \theta > \theta_0 \end{array} \right.$ و فرضیه‌های دو طرفه $\left\{ \begin{array}{ll} H_0 & : \theta = \theta_0 \\ H_1 & : \theta \neq \theta_0 \end{array} \right.$ اگر فضای پارامتری $\Theta_0 \cup \Theta_1$ به‌جای $(-\infty, +\infty)$ به‌صورت بازهٔ (b, a) باشد، که در آن $-\infty \leq b \leq a \leq +\infty$ آیا pـمقدار که به کران‌های a و b بستگی ندارد، بهترین معیار برای رد یا قبول فرضیه صفر هست؟

در بخش ۲ معایب pـمقدار از دو دیدگاه آمار بیزی و آمار شواهدی بیان می‌شوند. در بخش ۳ تابع زیان مناسب معرفی خواهد شد. در بخش ۴ روایی تحت تابع زیان توان دوم خطا بررسی می‌شود. در بخش ۵ pـمقدار اصلاح شده معرفی می‌شود. در بخش ۶ pـمقدار اصلاح شده و pـمقدار معمولی از دیدگاه آمار بیزی مقایسه می‌گردد. در بخش ۷ بهتر بودن pـمقدار اصلاح شده نسبت به pـمقدار معمولی، با مینیمم کردن مجموع خطای نوع اول و خطای نوع دوم نشان داده خواهد شد.

۲ معایب pـمقدار

با اینکه pـمقدار در مواردی با احتمال بیز پسین برابر می‌شود، اما در آمار بیزی بیشترین انتقادها از آن انجام می‌شود. لیندلی (۱۹۵۷) نشان داد برای توزیع نرمال با میانگین θ، آزمون فرضیه $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta \neq \theta_0$، بین استفاده از pـمقدار و استفاده از احتمال بیز پسین تناقض وجود دارد. وی نشان داد ممکن است pـمقدار موجب رد فرضیه H_0 شود، در حالی که احتمال بیز پسین H_0 مقداری شود

که منجر به تأیید فرضیه H_0 گردد. بعدها نشان داده شد تفاوت بین نتایج آزمون معنی‌داری و آزمون بیزی در فرضیه‌های دو طرفه به توزیع پیشینی و تعداد نمونه بستگی ندارد، در واقع اختلاف ذکر شده بستگی به مجموعه‌ای دارد که تحت آن بزرگترین کران پایین احتمال پسین فرض صفر به‌دست آورده می‌شود (لهمن، ۲۰۰۵). برگر و دلامپدی (۱۹۸۷) نشان دادند بین pـمقدار و احتمال بیز پسین در آزمون دو طرفه برای خانواده توزیع دوجمله‌ای تناقض وجود دارد و ثابت کردند اگر توزیع پیشینی، متقارن حول θ_0 با مد θ_0 یا مزدوج با میانگین θ_0 یا هر توزیع با میانه θ_0 باشد، بین pـمقدار و احتمال بیز پسین H_0، اختلاف زیادی وجود دارد. همچنین برگر و سلکه (۱۹۸۷) نشان دادند در آزمون‌های دو طرفه توزیع نرمال، حداقل احتمال بیز پسین H_0 با توزیع پیشینی دلخواه به مقدار قابل ملاحظه‌ای از pـمقدار بیشتر است. کاسلا و برگر (۱۹۸۷) نشان دادند در آزمون‌های یک طرفه این تناقض وجود ندارد و pـمقدار به‌صورت حدی از برآوردگر بیز است.

در بسیاری از رشته‌های علوم انسانی تفسیرهای نادرستی از pـمقدار می‌شود و اغلب آن را احتمال خطای نوع اول می‌نامند، درحالی‌که احتمال خطای نوع اول را خود محقق تعیین می‌کند و به توزیع آماره آزمون تحت فرض H_0 بستگی دارد و تابعی از مشاهدات نیست.

pـمقدار از دیدگاه استنباط شواهدی نیز مورد انتقاد قرار گرفته است و به‌عنوان یک معیار شواهدی خوب درنظر گرفته نمی‌شود. از مهمترین انتقاداتی که از دیدگاه شواهدی به pـمقدار وارد می‌شود می‌توان به موارد زیر اشاره کرد:
الف- عدم وابستگی به فرض مقابل، به‌عنوان مثال اگر $X \sim N(\theta, \sigma^2)$، آنگاه pـمقدار برای هر دو آزمون فرضیه‌های $\left\{\begin{array}{ll} H_0 & : \theta = 0 \\ H_1 & : \theta = 1 \end{array}\right.$ و $\left\{\begin{array}{ll} H_0 & : \theta = 0 \\ H_1 & : \theta = 100 \end{array}\right.$، با مشاهده $x = 1/5$ برابر با $0/05$ خواهد شد، در حالی‌که اختلاف بین دو فرضیه زیاد است، کالیبره نبودن، عدم تقارن (بدین معنی که اگر جای دو فرضیه H_0 و H_1 را عوض کنیم انتظار داریم pـمقدار جدید، یک منهای pـمقدار قبلی باشد ولی این گونه نمی‌شود)، صعودی نبودن نسبت به H_0 (بدین معنی که اگر ناحیه بزرگتری از فضای پارامتری در فرضیه صفر قرار گیرد انتظار می‌رود pـمقدار با قاطعیت بیشتری

فرضیه صفر را تأیید کند ولی در بسیاری از موارد این گونه نمی‌شود.) همچنین اگر $H_{\circ} \to H_{1}$ (به‌طور مثال $H_{\circ} : \theta = 1 - \frac{1}{n}$ و $H_{1} : \theta = 1 + \frac{1}{n}$ وقتی $n \to \infty$) انتظار می‌رود pـمقدار به سمت $\frac{1}{2}$ میل کند، یعنی به یک میزان از دو فرضیه حمایت کند، که این گونه نیست.

در این مقاله به بررسی pـمقدار به‌عنوان یک معیار تصمیم‌گیری در مورد $H_{\circ}$ پرداخته خواهد شد. توجه شود که هدف رد یا تأیید فرضیه صفر با استفاده از pـمقدار نیست، بلکه هدف استفاده از آن به‌عنوان معیاری برای برآورد احتمال رخ دادن مجموعه $\Theta_{\circ}$ ($I_{\Theta_{\circ}}(\theta)$) است.

۳ تابع زیان

از آنجایی‌که پارامتر دو مقدار دارد، پس برای هر تابع زیان دلخواه داریم:

$$L(\theta, \phi(x)) = \begin{cases} L(1, \phi(x)) & \theta \in \Theta_{\circ} \\ L(\circ, \phi(x)) & \theta \notin \Theta_{\circ} \end{cases}$$

لیندلی (۱۹۸۵) تابع زیانی را مناسب دانست که برآورد $I_{\Theta_{\circ}}(\theta)$ حاصل از آن با احتمال بیز پسین $P(\theta \in \Theta_{\circ} \mid x)$ برابر باشد. یکی از این‌گونه توابع زیان تابع زیان توان دوم خطای

$$L_{2}(\theta, \phi) = (I_{\Theta_{\circ}}(\theta) - \phi(x))^{2} \quad (1)$$

است، که برآوردگر بیزی حاصل از آن عبارتست از

$$\phi(x) \quad = \quad E(I_{\Theta_{\circ}}(\theta) \mid x) = P(\theta \in \Theta_{\circ} \mid x)$$

یکی دیگر تابع زیان لگاریتمی است که به‌صورت

$$L(\theta, \phi) = \log |I_{\Theta_{\circ}}(\theta) + \phi(x) - 1|$$

تعریف می‌شود، اما تابع زیان قدر مطلق خطا مناسب نیست، زیرا تصمیم بیزی حاصل از آن به‌صورت زیر است:

$$\phi_{1}^{\pi}(x) = \begin{cases} \circ & P(\theta \in \Theta_{\circ} \mid x) < \frac{1}{2} \\ 1 & o.w \end{cases}$$

۴ روایی تحت تابع زیان توان دوم خطا

در آزمون فرضیه‌های یک طرفه pـمقدار در بسیاری از موارد تحت تابع زیان توان دوم خطا روا است که نشان می‌دهد معیار مناسبی است. اما در آزمون فرضیه‌های دو طرفه pـمقدار نارواست. در عین حال هر چند pـمقدار آزمون دو طرفه برای توزیع نرمال نارواست، اما هیچ برآوردگر بیزی بهتری از آن وجود ندارد.

قضیه ۱ : فرض کنید $X_1,\ldots,X_n$ نمونه‌ای تصادفی از توزیع $N(\theta,1)$ باشد. pـمقدار برای آزمون فرضیه‌های یک طرفه $\begin{cases} H_0 & :\theta\le\theta_0 \\ H_1 & :\theta>\theta_0 \end{cases}$ تحت تابع زیان توان دوم خطا روا است.

برهان بدون آنکه از کلیت مسأله کاسته شود فرض می‌شود $n=1$.

$$\begin{aligned} p-\text{مقدار} &= P_{\theta_0}(X\ge x) \\ &= 1-\Phi((x-\theta_0)) \end{aligned}$$

اگر برای θ توزیع پیشینی $N(\theta_0,r)$ در نظر گرفته شود، آنگاه:

$$\theta|x\sim N\left(\frac{rx+\theta_0}{r+1},\frac{r}{r+1}\right)$$

درنتیجه

$$\begin{aligned} \lim_{r\to\infty}P(\theta\in\Theta_0|X=x) &= \int_{-\infty}^{\theta_0}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(x-\theta_0)^2)d\theta \\ &= \Phi(\theta_0-x) \\ &= 1-\Phi(x-\theta_0) \\ &= p\text{ـمقدار} \end{aligned}$$

بنابراین برآورد بیزی تعمیم‌یافته تحت تابع زیان توان دوم خطا، وقتی $r\to\infty$، برابر با pـمقدار خواهد شد، و چون ریسک بیزی pـمقدار متناهی است $(r_{\pi,P}=E(R_P(\theta))=\frac{1}{3})$، pـمقدار روا می‌باشد (ص. ۷۶ روبرت، ۲۰۰۱)

قضیه ۲ : در آزمون یک طرفه $\begin{cases} H_0 & : \theta \le \theta_0 \\ H_1 & : \theta > \theta_0 \end{cases}$، با تابع زیان توان دوم خطا،

(a) اگر $f(x \mid \theta)$ تابع چگالی توزیع دو جمله‌ای $B(n,\theta)$ باشد، pـمقدار برابر با $P(x) = P_{\theta_0}(X \ge x) = \sum_{k=x}^{n} \frac{n!}{k!(n-k)!}\theta_0^k(1-\theta_0)^{n-k}$ و رواست.

(b) اگر $f(x \mid \theta)$ تابع چگالی توزیع پواسون $\mathrm{P}(\theta)$ باشد، pـمقدار برابر با $P(x) = P_{\theta_0}(X \ge x) = \sum_{k=x}^{\infty} \frac{e^{-\theta_0}}{k!}\theta_0^k$ و رواست.

قضیه ۳ : (هوانگ، ۱۹۹۲) برای آزمون فرض دو طرفه، با تابع زیان توان دوم خطا و $P(x)$ پیوسته، pـمقدار نارواست.

هرچند قضیه ۳ بیان می کند pـمقدار نارواست، اما در قضیهٔ زیر نشان داده می‌شود که ممکن است با وجود ناروا بودن pـمقدار در آزمون دو طرفه، هیچ برآوردگر بیزی برتر از آن وجود نداشته باشد.

قضیه ۴ : برای آزمون فرضیه‌های دو طرفه

$$\begin{cases} H_0 & : \theta = \theta_0 \\ H_1 & : \theta \neq \theta_0 \end{cases}$$

براساس تک مشاهده X از چگالی $N(\theta, 1)$ و استفاده از تابع زیان توان دوم خطا، هیچ برآوردگر بیزی مناسبی نمی‌تواند از pـمقدار بهتر باشد.

برهان بدون آنکه از کلیت مسأله کاسته شود، فرض کنید $\theta_0 = 0$. برآوردگر بیزی این آزمون عبارتست از

$$\phi^\pi(x) = \frac{\pi_0 f(x|0)}{\pi_0 f(x|0) + (1-\pi_0)\int f(x|\theta)g(\theta)\mu(d\theta)}$$

که در آن $f(x|\theta)$ تابع چگالی $N(\theta, 1)$ است و pـمقدار برابر است با

$$\begin{aligned} P(x) &= P(|X| \ge x) \\ &= 2(1-\Phi(x)) \end{aligned}$$

سه حالت زیر را در نظر بگیرید:

حالت اول. اگر $\pi_0 = 1$، آنگاه $\phi^\pi(x) = 1$. اگر $\theta \to \infty$ آنگاه $I_{\{\theta_0\}} \to 0$ و $\phi(x) \to 1$ درنتیجه $P(x) = 2(1 - \phi(x)) \to 0$. بنابراین

$$\begin{aligned} R(\theta, P(x)) &= E_\theta \left(I_{\Theta_0}(\theta) - P(x)\right)^2 \\ &= E_\theta \left(I_{\{\theta_0\}}(\theta) - P(x)\right)^2 \to 0 \end{aligned}$$

اما

$$\begin{aligned} R(\theta, \phi^\pi(x)) &= E_\theta \left(I_{\Theta_0}(\theta) - \phi^\pi(x)\right)^2 \\ &= E_\theta \left(I_{\{\theta_0\}}(\theta) - \phi^\pi(x)\right)^2 = 1 \end{aligned}$$

درنتیجه ریسک pـمقدار کمتر از ریسک برآوردگر بیزی $\phi^\pi(x)$ است، بنابراین برآوردگر $\phi^\pi(x)$ از pـمقدار بهتر نیست.

حالت دوم. اگر $\pi_0 = 0$، آنگاه $\phi^\pi(x) = 0$، $R(0, \phi^\pi(x)) = 1$ و

$$\begin{aligned} R(0, P(x)) &= E[(P(x) - 0)^2] \\ &= Var(P(x)) + E^2(P(x)^2) \end{aligned}$$

و از $P(x) \sim U(0, 1)$ نتیجه می‌شود

$$R(0, P(x)) = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$$
$$R(0, \phi^\pi(x)) = 1 > R(0, P(x)) = \frac{1}{3}$$

بنابراین برآوردگر $\phi^\pi(x)$ نمی‌تواند بهتر از pـمقدار باشد.

حالت سوم. اگر $1 > \pi_0 > 0$، وقتی $\theta \to \infty$، آنگاه $R(\theta, P(x))$ کوچکتر از $R\left(\theta, \phi^\pi(x)\right)$ است. برای xهای به میزان کافی بزرگ $|x| > a > 0$، $\phi^\pi(x) > P(x)$ است، زیرا

$$\phi^\pi(x) \geq \frac{\pi_0 f(x|0)}{\pi_0 f(x|0) + (1 - \pi_0) f(x|\hat{\theta})} > P(x)$$

که در آن $\hat{\theta} = x$، برآوردگر ماکسیمم درستنمایی (MLE) برای θ است. برای $\theta \neq \theta_0$، اختلاف ریسک دو برآوردگر برابر است با

$$E_\theta \left(I_{\{\theta_0\}}(\theta) - \phi^\pi(x)\right)^2 - E_\theta \left(I_{\{\theta_0\}}(\theta) - P(x)\right)^2 = E_\theta \left(\phi^\pi(x)^2 - P(x)^2\right)$$

باتوجه به پیوستگی، یک $\varepsilon > 0$ وجود دارد که برای هر $a < |x| < a + \varepsilon$، آنگاه $\phi^\pi(x)^2 - P(x)^2 > \varepsilon$. بنابراین

$$E_\theta[\phi^\pi(x)^2] \geq \varepsilon P_\theta\left(a < |X| < a + \varepsilon\right) - P_\theta\left(|X| < a\right)$$

این کران پایین برای θهای بزرگ مثبت است، زیرا اگر $\theta \to \infty$ آنگاه

$$\frac{P_\theta(a < |X| < a + \varepsilon)}{P_\theta(|X| < a)} \to \infty$$

بنابراین اختلاف بین دو تابع ریسک برای θهای بزرگ همواره مثبت است و $\phi^\pi(x)$ نمی‌تواند بهتر از pـمقدار باشد.

برآوردگر $I_{\Theta_0}(\theta)$ تحت تابع زیان‌های مناسب دیگر نیز برابر با برآوردگری است که با استفاده از تابع زیان توان دوم خطا به‌دست آمد. بنابراین نتایج به‌دست آمده برای هر تابع زیان مناسبی برقرار است.

۵ pـمقدار اصلاح شده

فرض کنید فضای پارامتر دارای کران پایین a و کران بالای b باشد و X دارای توزیع $N(\theta, \sigma)$ است، وقتی فضای پارامتری $S = \Theta_0 \cup \Theta_1$ به‌صورت (a, b) باشد، برای آزمون فرضیه یک طرفه $H_0 : \theta \in \Theta_0 = (a, \theta_0]$ در مقابل $H_1 : \theta \in \Theta_1 = (\theta_0, b)$، pـمقدار معمولی برای آزمون UMP براساس یک مشاهده x برابر با $P_{\theta_0}(X \geq x)$ و برای آزمون دو طرفه $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta \neq \theta_0$ در آزمون UMPU برابر $P(|X| \geq |x - \theta_0|)$ است. بنابراین pـمقدار معمولی تنها به مقدار مشاهده شده x و مقدار θ_0 بستگی دارد و نمی‌تواند از اطلاعاتی که در مورد فضای پارامتری وجود دارد استفاده کند. حال pـمقدار اصلاح شده به‌صورت زیر تعریف می‌شود:

آزمون یک طرفه:

$$r_{\theta_\circ}(x) = \frac{P_{\theta_\circ}(X \geq x) - \min_{\theta \in S} P_\theta(X \geq x)}{\max_{\theta \in S} P_\theta(X \geq x) - \min_{\theta \in S} P_\theta(X \geq x)}$$

آزمون دو طرفه:

$$r'_{\theta_\circ}(x) = \frac{P(|X| > |x - \theta_\circ|) - \min_{\theta \in (a,b)} P(|X| > |x - \theta|)}{\max_{\theta \in (a,b)} P(|X| > |x - \theta|) - \min_{\theta \in (a,b)} P(|X| > |x - \theta|)}$$

برای مشاهده x، دامنه p-مقدار معمولی به‌ازای $\theta_\circ \in S$ به‌صورت

$$(\min_{\theta \in S} P_\theta(X \geq x), \max_{\theta \in S} P_\theta(X \geq x))$$

است. برای تصمیم‌گیری در مورد فرضیه صفر، چنانچه p-مقدار از سطح اطمینان خاص α (۰/۰۱ یا ۰/۰۵) کمتر باشد، فرض صفر رد می‌شود. وقتی فضای پارامتری محدود شود استفاده از سطح آزمون مشابه با زمانی که فضای پارامتری نامحدود است، منطقی نیست. زیرا به ازای هر x، p-مقدار معمولی بزرگتر از $\min_{\theta \in S} P_\theta(X \geq x)$ است. اگر $\alpha = \min_{\theta \in S} P_\theta(X \geq x)$ آنگاه به‌ازای هر مشاهده x، فرضیه $H_\circ : \theta \leq \theta_\circ$ رد نخواهد شد، یعنی فرضیه صفر هر چقدر هم نادرست باشد، با هر نمونه‌ای همواره تأیید خواهد شد. بنابراین تبدیلی از p-مقدار معمولی ساخته می‌شود به گونه‌ای که $\min_{\theta \in S} P_\theta(X \geq x)$ در اندازه معیاری که بر علیه فرضیه صفر استفاده می‌شود، دخالت نداشته باشد. از طرفی باتوجه به اینکه این معیار با $\circ < \alpha < ۱$ مقایسه می‌شود، باید دامنه این تبدیل فاصله $[\circ, ۱]$ باشد. سپس از این قانون تصمیم‌گیری استفاده می‌شود، که فرضیه صفر رد می‌شود اگر این تبدیل از p-مقدار کمتر از سطح α باشد. بنابراین برای اینکه دامنه معیار بین $[\circ, ۱]$ باشد، $P_{\theta_\circ}(X \geq x) - \min_{\theta \in S} P_\theta(X \geq x)$ را بر $\min_{\theta \in S} P_\theta(X \geq x) - \min_{\theta \in S} P_\theta(X \geq x)$ تقسیم می‌شود. بنابراین p-مقدار اصلاح شده $(r_{\theta_\circ}(x))$ به‌صورت یک تبدیل معقول از p-مقدار معمولی به‌دست می‌آید. به عنوان مثال، اگر فضای پارامتر (b, ∞) و $\theta_\circ = b$ باشد، آزمون فرضیه‌های

$$\begin{cases} H_\circ & : \theta = b \\ H_۱ & : \theta \in (b, \infty) \end{cases}$$

را درنظر بگیرید، که کل فضای پارامتری در فرض مقابل قرار گرفته است. در این صورت فرضیه صفر برای هر x همواره رد می‌شود. اما pـمقدار معمولی برابر با $P_b(X \geq x)$ و بیشتر از صفر است، پس می‌توان α را طوری انتخاب کرد که از pـمقدار معمولی کمتر باشد و نتواند فرضیه صفر را رد کند. در حالی‌که pـمقدار اصلاح شده برابر

$$r_{\theta_\circ}(x) \;=\; \frac{P_{\theta_\circ}(X \geq x) - \min_{\theta \in S} P_\theta(X \geq x)}{\max_{\theta \in S} P_\theta(X \geq x) - \min_{\theta \in S} P_\theta(X \geq x)} = \circ$$

است و با استفاده از آن فرضیه صفر به ازای هر مشاهده رد می‌شود.

۶ تقریب بیزی

در این بخش به مقایسه pـمقدار اصلاح شده $r_{\theta_\circ}(x)$ و pـمقدار معمولی تحت تابع زیان توان دوم خطا پرداخته می‌شود. هووانگ و همکاران (۱۹۹۲) برای ارزیابی pـمقدار پیشنهاد دادند از تابع زیان توان دوم خطای

$$L\left(e(x), \theta\right) = \left(e(x) - I(\theta \in \Theta_\circ)\right)^{۲} \qquad (۲)$$

استفاده شود، که در آن $I(\theta \in \Theta_\circ)$ تابع نشانگر $I(\theta \in \Theta_\circ) = \begin{cases} ۱ & \theta \in \Theta_\circ \\ \circ & o.w \end{cases}$ و $\Theta_\circ$ فضای پارامتری تحت فرض صفر است.

برای آزمون فرضیه یک طرفه نشان داده می‌شود که pـمقدار اصلاح شده یک برآوردگر بیزی برای $I(\theta \in \Theta_\circ)$ است، که از آن روا بودن pـمقدار اصلاح شده نتیجه می‌شود و برای آزمون فرضیه دو طرفه نشان داده می‌شود که برای همه θهایی که عضو فضای پارامتری محدود شده هستند، میانگین توان دوم خطا (MSE) $\omega r_{\theta_\circ}(\overline{x})$ کوچکتر از میانگین توان دوم خطای pـمقدار معمولی است، که در آن ω مقدار ثابت متعلق به یک فاصله است.

۱.۶ مقایسه pـمقدارهای معمولی و اصلاح شده در آزمون فرضیه یک طرفه

برآوردگر بیزی $I(\theta \in \Theta_0)$ با توزیع پیشینی $\pi(\theta)$ و تابع زیان توان دوم خطا عبارتست از

$$\begin{aligned}\eta(x) =& E\left(I(\theta \in \Theta_0) \mid X\right) \\ =& \frac{\int_{\Theta_0} \exp\{-\frac{1}{2}(x-\theta)^2\}d\theta}{\int_{\Theta_0} \exp\{-\frac{1}{2}(x-\theta)^2\}\pi(\theta)d\theta + \int_{\Theta_1} \exp\{-\frac{1}{2}(x-\theta)^2\}\pi(\theta)d\theta}\end{aligned} \quad (۳)$$

قضیه ۵ : pـمقدار اصلاح شده‌ی $r_{\theta_0}(x)$ معادل برآوردگر بیزی $I(\theta \in \Theta_0)$ با توزیع پیشینی $\pi(\theta) = I_{\Theta_0 \cup \Theta_1}(\theta)$ است.

برهان برای فضای پارامتری $(b, +\infty)$ ناحیه Θ_0 به‌صورت (b, θ_0) است و برآوردگر بیزی $I(\theta \in \Theta_0)$ با توزیع پیشینی $\pi(\theta) = I_{\Theta_0 \cup \Theta_1}(\theta) = I_{(b,\infty)}(\theta)$ عبارت است از

$$\begin{aligned}\frac{\int_b^{\theta_0} \exp\{-\frac{1}{2}(x-\theta)^2\}d\theta}{\int_b^{\infty} \exp\{-\frac{1}{2}(x-\theta)^2\}d\theta} &= \frac{\int_{-\infty}^{x-\theta_0} f(t)dt - \int_{-\infty}^{x-b} f(t)dt}{-\int_{-\infty}^{x-b} f(t)dt} \\ &= \frac{\int_{b-x}^{\theta_0 - x} f(t)dt}{\int_{b-x}^{\infty} f(t)dt} \\ &= \frac{\int_{-\infty}^{x-b} f(t)dt - \int_{-\infty}^{x-\theta_0} f(t)dt}{\int_{-\infty}^{x-b} f(t)dt} \\ &= \frac{P_b(X \le x) - P_{\theta_0}(X \le x)}{P_b(X \le x)} \\ &= \frac{P_{\theta_0}(X \le x) - P_b(X \le x)}{1 - P_b(X \le x)} \\ &= r_{\theta_0}(x)\end{aligned}$$

حالت‌های دیگر نیز به‌طور مشابه ثابت می‌شوند.

گزاره ۱: برآوردگر بیزی $\eta(x)$ برای $I(\theta \in \Theta_0)$ تحت تابع زیان توان دوم خطا روا است.

با توجه به گزاره ۱، pـمقدار اصلاح شده برآوردگری روا برای $I(\theta \in \Theta_0)$ برای یک فضای پارامتری محدود شده است، اما در قضیه زیر نشان داده می‌شود pـمقدار معمولی برآوردگری ناروا برای $I(\theta \in \Theta_0)$ است.

قضیه ۶ : اگر $X \sim N(\theta, ۱)$ و فضای پارامتری دارای کران پایین b و $\Theta_\circ \in [b, \theta_\circ)$ باشد، آنگاه pـمقدار معمولی برآوردگری ناروا برای $I(\theta \in \Theta_\circ)$ تحت تابع زیان توان دوم خطا است.

برهان : شرط لازم و کافی برای روا بودن برآوردگر $\eta(x)$ این است که توزیع‌های پیشین $\theta_\circ(\theta)$ در $(b, \theta_\circ)$ و $\theta_۱(\theta)$ در $(\theta_\circ, \infty)$ وجود داشته باشند، به‌طوری که

$$\eta(x) = \frac{\int_b^{\theta_\circ} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta)}{\int_b^{\theta_\circ} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta) + \int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta)}$$

وقتی که $\pi = \pi_\circ + \pi_۱$ و $\infty > \int_{\theta_\circ \cup \theta_۱} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi(d\theta)$.
برای نشان دادن ناروایی pـمقدار معمولی با استفاده از برهان خلف، فرض کنید $P_{\theta_\circ}(X \geq x)$ روا باشد، بنابراین:

$$P_{\theta_\circ}(X \geq x) = \frac{\int_b^{\theta_\circ} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta)}{\int_b^{\theta_\circ} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta) + \int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta)}$$

در نتیجه

$$\begin{aligned} P_{\theta_\circ}(X \leq x) &= ۱ - P_{\theta_\circ}(X \geq x) \\ &= \frac{\int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta)}{\int_b^{\theta_\circ} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta) + \int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta)} \\ &= \frac{\int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta) + \int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta)}{\int_b^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}(\pi_۱ + \pi_\circ)(d\theta)} \end{aligned}$$

از آنجایی‌که دامنه توزیع پیشینی $\pi_\circ(\theta)$، $(b, \theta_\circ)$ است. بنابراین $\int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_\circ(d\theta) = \circ$ در نتیجه

$$P_{\theta_\circ}(X \leq x) = \frac{\int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta)}{\int_{b}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta)}$$

پس

$$P_{\theta_\circ}(X \leq x) \int_{b}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi(d\theta) = \int_{\theta_\circ}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^۲\}\pi_۱(d\theta) \quad (۴)$$

با توجه به اینکه $X \sim N(\theta, ۱)$، داریم

$$\begin{aligned} P_{\theta_\circ}(X \leq x) &= \int_{-\infty}^{x} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}(t-\theta_\circ)^۲\}dt \\ &= \int_{\theta_\circ}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}(t-x)^۲\}dt \end{aligned}$$

زیرا اگر $T \sim N(x, ۱)$، آنگاه

$$\begin{aligned} \int_{\theta_\circ}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}(t-x)^۲\}dt &= P(T > \theta_\circ) \\ &= P(T - x > \theta_\circ - x) \\ &= P(Z > \theta_\circ - x) \\ &= P(Z < x - \theta_\circ) \\ &= P_{\theta_\circ}(X - \theta_\circ \leq x - \theta_\circ) \\ &= P_{\theta_\circ}(X \leq x) \end{aligned}$$

در نتیجه

$$\begin{aligned} P_{\theta_\circ}(X \leq x) &= \int_{\theta_\circ}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}(t-x)^۲\}dt \\ &\propto \int_{\theta_\circ}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}(x^۲+t^۲)\}dt \end{aligned}$$

از طرفی داریم

$$\exp\{\frac{x^{۲}}{۲}\} = \int_{-\infty}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp(tx)\exp\{-\frac{۱}{۲}t^{۲}\}dt$$

فرض کنید U و Y و W متغیرهای تصادفی مستقلی باشند که Y مقادیر ۰ و ۱ را با احتمالات $\int_{-\infty}^{\theta_{\circ}} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}t^{۲}\}dt$ و $\int_{\theta_{\circ}}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}t^{۲}\}dt$ می‌گیرد. U دارای تابع چگالی $\frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{۱}{۲}u^{۲}\}$ و $W = YW_{\circ} + (۱-Y)W_{۱}$ است که $W_{\circ}$ و $W_{۱}$ متغیرهای تصادفی با توابع چگالی $\exp\{-\frac{۱}{۲}w^{۲}\}\pi_{\circ}(w)$ و $\exp\{-\frac{۱}{۲}w^{۲}\}\pi_{۱}(w)$ هستند. در این صورت

$$\begin{aligned} P_{\theta_{\circ}}(X \le x) &= \int_{\theta_{\circ}}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp(tx)\exp\{-\frac{۱}{۲}(t+x)^{۲}\}d\theta \\ &= \exp\{-\frac{x^{۲}}{۲}\} \int_{\theta_{\circ}}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp(tx)\exp\{-\frac{۱}{۲}t^{۲}\}dt \\ &= \frac{۱}{\sqrt{۲\pi}} \exp\{-\frac{x^{۲}}{۲}\} \int_{\theta_{\circ}}^{\infty} \exp(ux)\exp\{-\frac{۱}{۲}u^{۲}\}du \end{aligned}$$

از طرفی

$$\begin{aligned} \int_{b}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^{۲}\}\pi(d\theta) &= \int_{b}^{\theta_{\circ}\infty} \exp\{-\frac{۱}{۲}x^{۲}\}\exp(\theta x)\pi(d\theta) \\ &= \exp\{-\frac{x^{۲}}{۲}\} \int_{b}^{\infty} \exp\{-\frac{۱}{۲}\theta^{۲}\}\exp(\theta x)\pi(d\theta) \\ &= \exp\{-\frac{x^{۲}}{۲}\} \int_{b}^{\infty} \exp\{-\frac{۱}{۲}w^{۲}\}\exp(wx)\pi(dw) \\ &= \exp\{-\frac{x^{۲}}{۲}\} \int_{b}^{\infty} \exp\{-\frac{۱}{۲}w^{۲}\}\exp(wx)\pi(w)dw \end{aligned}$$

پس سمت چپ رابطه (۴) برابر است با

$$\begin{aligned} & P_{\theta_{\circ}}(X \le x) \int_{b}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^{۲}\}\pi(d\theta) \\ = & \exp(-x^{۲}) \int_{\theta_{\circ}}^{\infty} \frac{۱}{\sqrt{۲\pi}} \exp(ux)\exp\{-\frac{۱}{۲}u^{۲}\}du \int_{b}^{\infty} \exp\{-\frac{۱}{۲}w^{۲}\}\exp(wx)\pi(w)dw \end{aligned}$$

و سمت راست رابطه (۴) نیز برابر است با

$$\begin{aligned} \int_{\theta_{\circ}}^{\infty} \exp\{-\frac{۱}{۲}(x-\theta)^{۲}\}\pi(d\theta) &= \exp\{-\frac{x^{۲}}{۲}\} \int_{\theta_{\circ}}^{\infty} \exp\{-\frac{۱}{۲}\theta^{۲}\}\exp(\theta x)\pi(d\theta) \\ &= \exp\{-\frac{x^{۲}}{۲}\} \int_{\theta_{\circ}}^{\infty} \exp\{-\frac{۱}{۲}w^{۲}\}\exp(wx)\pi_{\circ}(w)dw \end{aligned}$$

پس ما می توانیم رابطه (۴) را به‌صورت زیر بنویسیم

$$\begin{aligned}\exp(-x^۲)\int_{\theta_۰}^{\infty}\frac{۱}{\sqrt{۲\pi}}\exp(ux)\exp\{-\frac{۱}{۲}u^۲\}du\int_{b}^{\infty}\exp\{-\frac{۱}{۲}w^۲\}\exp(wx)\pi(w)dw\\ = \exp\{-\frac{x^۲}{۲}\}\int_{\theta_۰}^{\infty}\exp\{-\frac{۱}{۲}w^۲\}\exp(wx)\pi_۰(w)dw\end{aligned}$$

درنتیجه

$$\begin{aligned}\int_{\theta_۰}^{\infty}\frac{۱}{\sqrt{۲\pi}}\exp(ux)\exp\{-\frac{۱}{۲}u^۲\}du\int_{b}^{\infty}\exp\{-\frac{۱}{۲}w^۲\}\exp(wx)\pi(w)dw\\ = \exp\{-\frac{x^۲}{۲}\}\int_{\theta_۰}^{\infty}\exp\{-\frac{۱}{۲}w^۲\}\exp(wx)\pi_۰(w)dw\end{aligned}$$

بنابراین

$$\begin{aligned}\int_{\theta_۰}^{\infty}\exp(ux)\exp\{-\frac{۱}{۲}u^۲\}du\int_{b}^{\infty}\exp\{-\frac{۱}{۲}w^۲\}\exp(wx)\pi(w)dw\\ = \int_{-\infty}^{\infty}\exp\{-\frac{۱}{۲}u^۲\}du\int_{\theta_۰}^{\infty}\exp\{-\frac{۱}{۲}w^۲\}\exp(wx)\pi_۰(w)dw\end{aligned}$$

با توجه به تساوی بالا و اینکه $P(Y=۱)=P(U\geq\theta_۰)$، برای هر x

$$E[e^{x(U+W)}\mid U\geq\theta_۰]=E[e^{x(U+W)}\mid Y=۱]$$

از آن‌جا که تابع مولد گشتاور توزیع‌های شرطی برای همه xها یکسان است، پس خود توزیع‌های شرطی یکسان هستند، یعنی برای همه مقادیر z

$$P(U+W\leq z\mid U\geq\theta_۰)=P(U+W\leq z\mid Y=۱)$$

اما از آنجا که همواره $U\geq\theta_۰$ و $W\geq b$، پس اگر $z<b+\theta_۰$ آنگاه $P(U+W\leq z\mid U\geq\theta_۰)=۰$، ولی مقدار $P(U+W\leq z\mid Y=۱)$ همیشه مثبت و مخالف صفر است، که این تناقض از فرض روا بودن برای pـمقدار معمولی ناشی می‌شود.

۲.۶ مقایسه pـمقدارهای معمولی و اصلاح شده در آزمون فرضیه دو طرفه

در قضیه ۷ فاصله در برگیرنده ω به‌گونه‌ای درنظر گرفته شده است که کارآیی $\omega r_{\theta_۰}(\overline{x})$ بهتر از کارآیی pـمقدار معمولی تحت تابع زیان توان دوم خطا باشد. در

تحلیل‌ها و روش‌های عددی، ω را می‌توان یک عدد ثابت نزدیک به یک درنظر گرفت. بنابراین pـمقدار اصلاح شده دارای کارآیی بهتری نسبت به pـمقدار معمولی تحت تابع زیان توان دوم خطا است.

قضیه ۷ : فرض کنید $\overline{X}$ دارای توزیع $N(\theta, ۱)$ و فضای پارامتری θ به‌صورت (a, b) باشد. تحت تابع زیان توان دوم خطا (۲)

(i) برای $\theta = \theta_{\circ}$، $\omega r_{\theta_{\circ}}(\overline{x})$ دارای میانگین توان دوم خطا کمتری نسبت به pـمقدار معمولی است وقتی که ω متعلق به بازه زیر باشد:

$$\frac{E_{\theta_{\circ}}\left(r_{\theta_{\circ}}(\overline{x})\right) \pm \sqrt{\left(E_{\theta_{\circ}}\left(r_{\theta_{\circ}}(\overline{x})\right)\right)^{۲} - E_{\theta_{\circ}}\left(r_{\theta_{\circ}}(\overline{x})\right)\left(۱ - E_{\theta_{\circ}}\left(P_{\theta_{\circ}}(\overline{x}) - ۱\right)^{۲}\right)}}{E_{\theta_{\circ}}\left(r_{\theta_{\circ}}(\overline{x})\right)}$$

(ii) برای $\theta \neq \theta_{\circ}$، $\omega r_{\theta_{\circ}}(\overline{x})$ دارای میانگین توان دوم خطای کمتر نسبت به pـمقدار معمولی است وقتی که ω متعلق به بازه زیر باشد:

$$\left(\circ, \min_{\theta \in (a,b), \theta \neq \theta_{\circ}} \left\{\frac{E_{\theta}[P_{\theta_{\circ}}^{۲}(\overline{X})]}{E_{\theta}[r_{\theta_{\circ}}^{۲}(\overline{X})]}\right\}^{\frac{۱}{۲}}\right) \qquad (۵)$$

برهان : (i) وقتی $\theta = \theta_{\circ}$، آنگاه $I(\theta \in \Theta_{\circ}) = ۱$.

$$MSE\left(P_{\theta_{\circ}}(\overline{x})\right) = E_{\theta_{\circ}}[P_{\theta_{\circ}}(\overline{X}) - ۱]^{۲}$$
$$MSE\left(\omega r_{\theta_{\circ}}(\overline{x})\right) = E_{\theta_{\circ}}[\left(\omega r_{\theta_{\circ}}(\overline{X}) - ۱\right)^{۲}]$$

حال از نامساوی $MSE\left(\omega r_{\theta_{\circ}}(\overline{x})\right) < MSE\left(P_{\theta_{\circ}}(\overline{x})\right)$ نتیجه می‌شود:

$$\omega^{۲} E_{\theta_{\circ}}[r_{\theta_{\circ}}^{۲}(\overline{x})] - ۲\omega E_{\theta_{\circ}}[r_{\theta_{\circ}}(\overline{x})] + ۱ - E_{\theta_{\circ}}(P_{\theta_{\circ}}(\overline{x}) - ۱)^{۲} < \circ$$

که برای برقراری آن، باید ω در فاصله (۱) قرار گیرد.

(ii) وقتی $\theta \neq \theta_{\circ}$، آنگاه $I(\theta \in \Theta_{\circ}) = \circ$. پس

$$MSE\left(P_{\theta_{\circ}}(\overline{x})\right) = E_{\theta}[(P_{\theta_{\circ}}(\overline{X}))^{۲}]$$

$$MSE\,(\omega r_{\theta_0}(\overline{x})) = E_{\theta_0}[(\omega r_{\theta_0}(\overline{X}))^{2}]$$

بنابراین

$$E_{\theta}[(P_{\theta_0}(\overline{X}))^{2}] - \omega^{2} E_{\theta}[(r_{\theta_0}(\overline{X}))^{2}] > 0$$

در نتیجه

$$\omega^{2} < \frac{E_{\theta}[(P_{\theta_0}(\overline{X}))^{2}]}{E_{\theta}[(r_{\theta_0}(\overline{X}))^{2}]}$$

یعنی ω متعلق به بازهٔ (۱) است.

فرع ۱ : با توجه به قضیه ۷ اگر یک ω وجود داشته باشد که شرایط (i) و (ii) برقرار باشند، آنگاه برای هر مقدار θ از فضای پارامتری (a,b)، $\omega r_{\theta_0}(\overline{x})$ بهتر از pـمقدار معمولی است.

۷ مینیمم کردن مجموع خطای نوع اول و دوم در آزمون فرض ساده

در این بخش دو pـمقدار از نقطه نظر مینیمم کردن مجموع خطای نوع اول و دوم در آزمون فرضیه ساده باهم مقایسه خواهند شد. برای $X \sim N(\theta, 1)$ و فرضیه‌های ساده

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta = \theta_1 \qquad (\theta_1 > \theta_0) \tag{۶}$$

اگر $x = \theta_0$ باشد، pـمقدار معمولی با استفاده از ناحیه بحرانی آزمون UMP برابر است با:

$$P_{\theta_0}(X \geq \theta_0) = 0/5$$

با توزیع پیشینی $\pi(\theta) = I_{\Theta_0 \cup \Theta_1}(\theta)$ برآوردگر بیزی برای این آزمون با توجه به رابطه (۳) به‌صورت

$$\eta(x) = \frac{f_{\theta_0}(x)}{f_{\theta_0}(x) + f_{\theta_1}(x)}$$

است، که فرضیه H_0 رد می‌شود اگر $\eta(x) < k$. همچنین

$$\eta(x) = \begin{cases} \frac{1}{2} & x = \frac{\theta_0+\theta_1}{2} \\ > \frac{1}{2} & x < \frac{\theta_0+\theta_1}{2} \\ < \frac{1}{2} & x > \frac{\theta_0+\theta_1}{2} \end{cases}$$

اگر $k = ۰/۵$ باشد، آزمون با ناحیه رد $\{x : x > \frac{\theta_۰+\theta_۱}{۲}\} = \{x : \eta(x) < k\}$ با توجه به گزاره زیر دارای ویژگی بهینه برای مینیمم کردن مجموع دو نوع خطا است.

گزاره ۲: اگر متغیر تصادفی X با تابع چگالی $f_\theta(x)$ دارای خاصیت MLR در X باشد، در آزمون (۵) یک ناحیه رد که مجموع خطای نوع اول و دوم را مینییم کند، به‌صورت $\{x : x > c^*\}$ است که c^* از رابطه $f_{\theta_۰}(c^*) = f_{\theta_۱}(c^*)$ به‌دست می‌آید.

برهان با توجه به لم نیمن ـپیرسون ناحیه ردی که مجموع خطای نوع اول و دوم را مینیمم کند باید به‌صورت $S_c = \{x : x > c\}$ باشد، که در آن c مقداری ثابت است. در این‌صورت

$$\begin{aligned} \alpha &= P_{\theta_۰}(X \geq c) \\ &= \int_c^\infty f_{\theta_۰}(x)d\,x \\ \beta &= P_{\theta_۱}(X < c) \\ &= \int_{-\infty}^c f_{\theta_۱}(x)d\,x \\ e(c) &= \alpha + \beta \\ &= \int_c^\infty f_{\theta_۰}(x)d\,x + \int_{-\infty}^c f_{\theta_۱}(x)d\,x \end{aligned}$$

چون

$$\frac{\partial}{\partial c}e(c) = f_{\theta_۱}(c) - f_{\theta_۰}(c) = ۰$$

اگر جواب معادله بالا c^* در نظر گرفته شود، آنگاه مقداری است که $e(c)$ را مینیمم می‌کند و $f_{\theta_۰}(c^*) = f_{\theta_۱}(c^*)$.

برای توزیع $N(\theta, ۱)$، $c^* = \frac{\theta_۰+\theta_۱}{۲}$ است. پس ناحیه ردی که مجموع خطای نوع اول و دوم را مینیمم کند به‌صورت $\{x : x > \frac{\theta_۰+\theta_۱}{۲}\}$ است، که همان ناحیه بحرانی در آزمون بیزی است. پس ناحیه رد به‌دست آمده از pـمقدار اصلاح شده، ناحیه ردی را ایجاد می‌کند که مجموع خطای نوع اول و دوم آزمون متناظر با آن مینیمم می‌شود. از طرفی ناحیه رد به‌دست آمده از pـمقدار معمولی برابر است با

$\{x : P_{\theta_\circ}(X \geq x) < ۰/۵\}$، که معادل با ناحیه $\{x : x > \theta_\circ\}$ است و با ناحیه ردی که مجموع خطای نوع اول و دوم را مینیمم می‌کند متفاوت است.

بحث و نتیجه‌گیری

در این مقاله ابتدا pـمقدار به‌عنوان برآوردگری برای احتمال رخ دادن مجموعه مشخص شده با $\Theta_\circ$، و نه به‌عنوان عددی برای رد یا تأیید فرضیه صفر، از دیدگاه نظریه تصمیم بررسی گردید. pـمقدار به‌طور کلی در آزمون‌های یک طرفه روا است و به‌صورت حدی از برآوردگر بیزی است. اما در حالت دو طرفه نارواست، مطابق با هیچ برآوردگر بیز تعمیم‌یافته‌ای نیست و توانایی هیچ برآوردگر بیزی از pـمقدار در آزمون‌های دو طرفه بیشتر نیست مگر آنکه بتوان یک برآوردگر خاص برتر پیدا نمود، که نشان دهنده جایگاه مناسب pـمقدار در میان برآوردگرهای دیگر است. بنابراین در آزمون‌های دو طرفه نیز pـمقدار می‌تواند یک اندازه معیار مناسب برای تأیید یا رد فرض صفر باشد. در حضور وجود اطلاعاتی دربارهٔ فضای پارامتری pـمقدار اصلاح شده، برای فضای پارامتری کراندار معرفی شد که نسبت به pـمقدار معمولی عملکرد بهتری داشته و از دیدگاه آمار بیزی و مینیمم کردن مجموع خطاهای نوع اول و دوم از pـمقدار معمولی بهتر عمل می‌کند.

تقدیر و تشکر

نویسندگان از پیشنهادهای ارزنده داوران و سردبیر محترم مجله کمال سپاس‌گزاری را دارند.

مراجع

ارقامی، ن.، دست برآورده، ع. و زمان‌زاده، الف. (۱۳۸۷)، شواهد آماری، قسمت دوم: قانون درستنمایی و معایب pـمقدار به‌عنوان پشتیبانی داده از فرضیه صفر. اندیشه آماری، ۱۲، **۲**، ۳ـ۱۶.

ارقامی، ن. ، زمان‌زاده، الف. و راحتی، س. (۱۳۸۷)، شواهد آماری، تفسیر نتایج آزمون‌های آماری کلاسیک. اندیشه آماری، ۱۲، ۱، ۳-۱۰.

کاظم‌نژاد، الف.، محبی، م. و ثناگو، م. (۱۳۸۱)، p-مقدار به‌عنوان معیاری برای پشتیبانی فرض‌ها، محاسن و معایب، مجله اندیشه آماری، ۷، **۲**، ۱۹-۲۴.

Berger, J. O. and Delmpady, M. (1987), Testing Precise Hypotheses (with Discussion), *Statistical Science*, **2**, 317-352.

Berger, J. O. and Sellke , T. (1987), Testing a Point Null Hypothesis: The Irreconcilability of p-value and Evidence (with discussion), *Journal of the American Statistical Association*, **82**, 112-139.

Casella, G. and Berger, R. L. (1987), Comment on Testing Precise Hypothesis by Berger, J. O. and Delmpady, Mathematical, *Statistical Science*, **2**, 344-347.

Forster, M. and Sobber, E. (2001), Why Likelihood? In The Nature of Scientific Evidence, Empirical, Statistical and Philosophical Consideration, University of Chcago Press.

Goodman, S. N. (2005), Introduction to Bayesian Methods I: Measuring The Strength of Evidence, Clinical Trials, **2**, 282-290.

Hwang, J. T., Casella, G., Robert, C., Wells, M. T., and Farrell, R. H. (1992), Estimation of Accuracy in Testing, *Annals of Statistics*, **20**, 490-509.

Lehmann, E. L. (2005), *Testing Statistical Hypotheses*, New York, John Wiley.

Lindley, D. V. (1957), A Statistical Paradox, *Biometrika*, **44**, 187-192.

Lindley, D. V. (1985), *Making Decisions*, 2$^{\text{nd}}$ ed, New York, John Wiley.

Mandelkern, M. (2002), Setting Confidence Intervals for Bounded Parameters, *Statistical Science*, **17**, 149-172.

Robert, P. C. (2001), *The Baysian Choice*, 2$^{\text{nd}}$ ed. Springer-Verlag, New York.

Royall, R. (1997), Statistical Evidence: A Likelihood Paradigm, Chapman & Hall /CRC.